COMMENTARY

Genome-wide Testing: Genomic Medicine

Commentary on the article by Bar-Shira *et al.* on page 353

SEAN A. MCGHEE, AND EDWARD R.B. McCABE

Departments of Pediatrics [S.A.M., E.R.B.M.] and Human Genetics [E.R.B.M.], David Geffen School of Medicine, Mattel Children's Hospital at UCLA, UCLA Center for Society and Genetics [E.R.B.M.], University of California, Los Angeles, Los Angeles, CA, 90095

Karyotyping has been in use in genetics clinics for decades as the standard screening and diagnostic test for major cytogenetic abnormalities. Many of these abnormalities involve changes in gene copy number, resulting in insertions and/or deletions (indels). The limitations of karyotyping in detecting some copy number changes is well known. For example, in del22q11.2 (diGeorge) syndrome, the deletion is usually too small to be detected even by high-resolution banding techniques (1). Techniques have been developed to deal with this lack of resolution, such as fluorescence *in situ* hybridization (FISH) and spectral karyotyping, but frequently, these require a high degree of suspicion, as the techniques are targeted to specific regions and not genome-wide interrogations.

A relatively new method for copy number assessment by comparative genomic hybridization (CGH), uses the signal intensity from microarrays to determine copy number changes and is known as array-CGH. Like karyotyping, array-CGH is a whole-genome technique, but unlike the karyotype, the high density of probes in a microarray means copy number can be assessed simultaneously at many thousands of sites throughout the genome. This promises to greatly improve the identification of indels in patients with genetic diseases, since prior suspicion of a particular indel is not required, as it is in FISH.

The article by Bar-Shira *et al.* (2) in this month's Pediatric Research demonstrates the clinical implementation of array-CGH, the GenoSensor Array 300. Their results clearly highlight both the potential of this method, as well as the problems that remain to be solved, and builds on the work of many others (3–6). Many of the problems with this technology derive from its strength, and this is the enormously increased information content of the data returned from the microarray compared with FISH or karyotyping. What remains to be accomplished to fully realize the potential of array-CGH is to learn how to manage this flood of information and how to derive meaning from it.

The advantages of array-CGH are clear. Genome-wide, high-resolution copy number information enables a single assessment to provide information in any given patient without requiring the degree of suspicion necessary to request a specific FISH test. This ability to interrogate the entire genome for indels would be expected to improve the identification of genetic disease in patients, particularly those with unusual phenotypes or syndromes that do not have constant features.

Case five in the study, the patient discovered to have del22q11.2 by array-CGH, demonstrates this potential. The del22q11.2 syndrome is known to have an extremely wide range of phenotypic features and penetrance, and many patients with identical deletions have widely differing phenotypes (1). The phenotype of the patient described in the article is clearly unusual for the del22q11.2 syndrome, making FISH unlikely to be ordered, but the deletion could be discovered using array-CGH as a screening tool. Improved diagnosis may lead to better management of patients, as in this case, where speech delay could be attributed to velopalatal insufficiency. It is not always possible to have a high degree of clinical suspicion and would be extremely expensive to carry out multiple FISH analyses.

Case three demonstrates the importance of improving resolution with array-CGH beyond current karyotyping. All of the unknown patients in this study were missed by standard karyotyping, demonstrating the problems raised by lack of resolution in the karyotype technique. Case three shows that the problem of low resolution may extend to different implementations of the array-CGH principle.

Array-CGH may be implemented using bacterial artificial chromosomes, cosmids, or oligonucleotides as probes, all varying in the size of the probe and density of the array that can be constructed from them. With increasing density of the array there is a corresponding increase in the resolution of indel detection (7). In case three, detection of the deletion in WT1 appeared to be dependent on the size of the probe used in the array, as was subsequently confirmed by FISH. Furthermore, large probes may promote cross-hybridization leading to aberrant results (8). Improving resolution may improve the ability to detect smaller deletions, suggesting that very high density arrays with small probes, such as can be obtained with oligonucleotide arrays, may ultimately prove to be the most accurate diagnostically (7). In fact, though FISH is currently the gold

standard for interstitial deletions, it has clear weaknesses, and array-CGH may ultimately prove to be the better test. FISH is less capable of detecting insertions and duplications than array-CGH and the nature of the FISH probe may affect results, as in case three.

Bar-Shira *et al.* (2) demonstrate that a number of challenges remain to reliable implementation of array-CGH technology. Microarrays are inherently "noisy" tools, and many factors in the array methodology contribute to this deterioration of the signal-to-noise ratio. Random variation across the chip surface, manufacturing variance, PCR and hybridization efficiencies, and DNA contamination can all introduce noise and reduce the ability to detect signal. To respond to this, a number of different analytical algorithms have been proposed, including clustering algorithms, Bayesian hidden Markov models, change-point models, and spatially structured mixture models (9–12). Evaluation of the output from these different algorithms shows that there are clear improvements when they are used to determine copy number, indicating that the signal-to-noise issue in microarray copy number analysis may yield to mathematical manipulation of the raw data.

A larger issue results from the genome-wide nature of the results. Particularly as resolution increases, the ability to detect normal large-scale indel variation will also increase. The extent and nature of large-scale indel polymorphisms in the human genome is just now becoming known (13). Traditional karyotyping is less able to detect these structural variations because alterations large enough to cause microscopically visible changes in the karyotype frequently (though not always) are associated with phenotypic changes and many of these polymorphic variants are submicroscopic. The problem of polymorphism detection is similar to the problem faced by the infectious disease field during the discovery of microorganisms. How can normal commensals be reliably distinguished from pathogens? This was addressed by Koch's postulates, creating a standard by which physicians could agree on what constitutes a pathogen. As experience with microorganisms grew, the need to demonstrate all of the postulates became less necessary as some organisms clearly were pathogenic in certain settings and others clearly were not.

This sort of information will emerge as microarray assessments of copy number become more common, and eventually associations will be developed, at least for common indels, that will help distinguish polymorphism from pathology. For patients with uncommon indels, however this will remain a problem. A clear set of standards is needed that can be used to help determine whether a deletion is likely to be pathogenic or not. This will require investment in gene-association and mapping studies, improved cataloging of the extent of normal indel variation, and careful assessment of population structure. Furthermore, the nature of these studies may need to develop and change from traditional methods to adequately deal with the demands imposed by this type of analysis. Data mining, machine learning and bayesian approaches may be more useful in these studies, particularly in settings where a specific indel is rare and a frequentist approach may simply not be possible. Large population studies may be required (14) to identify associations between these variations, representing heterogeneity millions of base pairs in length within each individual genome, and predispositions to common complex disorders like cancer, diabetes mellitus and cardiovascular diseases (13).

Regulatory issues are also potential pitfalls for array-CGH copy number analysis. Approval for diagnostic testing typically focuses on highly accurate tests for specific diseases. Yet many of the most valuable tests available to medicine do not necessarily focus on specific diseases, such as radiographic tests, morphologic assessment of peripheral blood smears, karyotyping, and even history and physical examination. These tests are screening tests, casting a broad, descriptive net, but playing an important role in narrowing diagnostic possibilities.

The work by Bar-Shira *et al.* (2) is an example of the first steps that are now being taken in the clinical application of genome-wide microarray technology as tools of genomic medicine. Like all tests, and screening tests in particular, it will need to be interpreted within the clinical context, which will include not only the phenotype of the patient, but also our knowledge of the genome, and the structure and diversity of populations. Although some technical issues with array-CGH clearly remain, it is ultimately this broader knowledge of genomic diversity and disease predisposition that will define the limits and usefulness of the technique.

REFERENCES

1. Lindsay EA 2001 Chromosomal microdeletions: Dissecting DEL22Q11 syndrome. Nat Rev Genet 2:858–868
2. Bar-Shira A, Rosner G, Rosner S, Goldstein M, Orr-Urtreger A 2006 Array-Based Comparative Genome Hybridization in Clinical Genetics. Pediatr Res 60:353–358
3. Shaw CJ, Shaw CA, Yu W, Stankiewicz P, White LD, Beaudet AL, Lupski JR 2004 Comparative genomic hybridisation using a proximal 17p BAC/PAC array detects rearrangements reponsible for four genomic disorders. J Med Genet 41:113–119
4. Cheung SW, Shaw CA, Yu W, Li J, Ou Z, Patel A, Yatsenko SA, Cooper ML, Furman P, Stankiewicz P, Lupski JR, Chinault AC, Beaudet AL 2005 Development and validation of a CGH microarray for clinical cytogenetic diagnosis. Genet Med 7:422–432
5. Sahoo T, Peters SU, Madduri NS, Glaze DG, German JR, Bird LM, Barbieri-Welge R, Bichell TJ, Beaudet AL, Bacino CA 2005 Microarray based comparative genomic hybridization testing in deletion-bearing Angelman Syndrome patients: genotype-phenotype correlations. J Med Genet 43:512–516
6. Yu W, Ballif BC, Kashork CD, Heilstedt HA, Howard LA, Cai WW, White LD, Liu W, Beaudet AL, Bejjani BA, Shaw CA, Shaffer LG 2003 Development of a comparative genomic hybridization microarray and demonstration of its utility with 25 well-characterized 1p36 deletions. Hum Mol Genet 12:2145–2152
7. Ylstra B, van den Ijssel P, Carvalho B, Brakenhoff RH, Meijer GA 2006 BAC to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array CGH). Nucleic Acids Res 34:445–450
8. Bejjani BA, Saleki R, Ballif BC, Rorem EA, Sundin K, Theisen A, Kashork CD, Shaffer LG 2005 Use of targeted array-based CGH for the clinical diagnosis of chromosomal imbalance. Am J Med Genet A 134:259–267
9. Broet P, Richardson S 2006 Detection of gene copy number changes in CGH microarrays using a spatially correlated mixture model. Bioinformatics 22:911–918
10. Guha S, Li Y, Neuberg D 2006 Bayesian hidden Markov modeling of array CGH data. http://www.bepress.com/harvardbiostat/paper24.
11. Lai WR, Johnson MD, Kucherlapati R, Park PJ 2005 Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. Bioinformatics 21:3763–3770
12. Minin VN, Dorman KS, Fang F, Suchard MA 2005 Dual multiple change-point model leads to more accurate recombination detection. Bioinformatics 21:3034–3042
13. Feuk L, Carson AR, Scherer SW 2006 Structural variation in the human genome. Nat Rev Genet 7:85–97
14. McCabe ER 2002 Translational genomics in medical genetics. Genet Med 4:468–471