The Journal of Molecular Diagnostics, Vol. 16, No. 5, September 2014

the Journal of Molecular Diagnostics
jmd.amjpathol.org

Copy Number Variation Sequencing for Comprehensive Diagnosis of Chromosome Disease Syndromes

Desheng Liang,* Ying Peng,* Weigang Lv,* Linbei Deng,* Yanghui Zhang,* Haoxian Li,* Pu Yang,* Jianguang Zhang,† Zhuo Song,† Genming Xu,† David S. Cram,† and Lingqian Wu*

From the State Key Laboratory of Medical Genetics, Central South University, Changsha, Hunan; and the Berry Genomics Co., Ltd.,† Beijing, China*

Accepted for publication May 16, 2014.

Address correspondence to Lingqian Wu, Ph.D., State Key Laboratory of Medical Genetics, Central South University, 110 Xiangya Rd., Changsha, Hunan 410078, China; or David Cram, Ph.D., Berry Genomics, Beijing 100015, China. E-mail: wulingqian@sklmg.edu.cn or david.cram@berrygenomics.com.

Detection of chromosome copy number variation (CNV) plays an important role in the diagnosis of patients with unexplained clinical symptoms and for the identification of chromosome disease syndromes in the established fetus. In current clinical practice, karyotyping, in conjunction with array-based methods, is the gold standard for detection of CNV. To increase accessibility and reduce patient costs for diagnostic CNV tests, we speculated that next-generation sequencing methods could provide a similar degree of sensitivity and specificity as commercial arrays. CNV in patient samples was assessed on a medium-density single nucleotide polymorphism array and by low-coverage massively parallel CNV sequencing (CNV-seq), with mate pair sequencing used to confirm selected CNV deletion breakpoints. A total of 10 ng of input DNA was sufficient for accurate CNV-seq diagnosis, although 50 ng was optimal. Validation studies of samples with small CNVs showed that CNV-seq was specific and reproducible, suggesting that CNV-seq may have a potential genome resolution of approximately 0.1 Mb. In a blinded study of 72 samples with known gross and submicroscopic CNVs originally detected by single nucleotide polymorphism array, there was high diagnostic concordance with CNV-seq. We conclude that CNV-seq is a viable alternative to arrays for the diagnosis of chromosome disease syndromes. *(J Mol Diagn 2014, 16: 519—526; http://dx.doi.org/10.1016/j.jmoldx.2014.05.002)*

There have been >200 chromosome disease syndromes recorded and studied in the human population.[1] Whole or partial chromosome numerical changes account for most chromosome syndromes, with Down syndrome being the most prevalent. The remaining chromosome diseases are classified as submicroscopic deletion and duplication syndromes caused by the loss or gain of variably sized chromosome segments [copy number variations (CNVs)] that can affect the normal expression pattern of one or more genes.

There is a wide spectrum of clinical phenotypes associated with chromosome disease syndromes, which can be categorized into multiple congenital abnormalities, physical disabilities, dysmorphic features, developmental delay, intellectual disability, seizure disorders, autistic behaviors, and learning disabilities.[1—3]

Recent molecular analyses of human gametes and preimplantation human embryos[4—6] have revealed that inherent chromosome instability is a major source of these chromosome abnormalities. Driven by well-known mechanisms of nonhomologous end joining, as well as some recently proposed mechanisms involving perturbation of DNA replication and replication of noncontiguous DNA segments, *de novo* CNVs form at rates far exceeding other kinds of mutagenic events.[7] In addition, a small but significant proportion of chromosome disease continues to be perpetuated by inadvertent familial inheritance.[1] Most of these chromosome abnormalities of *de novo* or parental origin are generally compatible with fetal development to term, leading to approximately 0.3% of children born with a chromosome disease.[8]

For >30 years, prenatal diagnosis has played a key role in detecting these chromosome abnormalities in the late first

Supported by National Key Basic Research Program of China grant 2012CB944601 (L.W.) and National Key Technology R&D Program of China grant 2012BAI09B05 (D.L.).

Disclosures: J.Z., Z.S., G.X., and D.S.C. are employees of Berry Genomics. None of these authors have any stock or bond holdings.

http://dx.doi.org/10.1016/j.jmoldx.2014.05.002

trimester or early second trimester of pregnancy, with the aim of reducing the incidence of chromosome disease in the newborn by elective termination of pregnancy.[1] The application of prenatal diagnosis involves a range of fetus analyses, including screening tests, such as maternal serum screening and ultrasound, and if chromosome disease is suspected, follow-up diagnosis is usually performed by either chorionic villous sampling or amniocentesis and then karyotyping. Karyotyping, with a resolution of approximately five megabases (Mb), is the gold standard enabling the detection of fetal aneuploidies, polyploidies, balanced and unbalanced structural rearrangements, large microdeletions and duplications, and mosaicism when >20 metaphase cells are analyzed.[9–11] Other methods, such as fluorescence *in situ* hybridization,[12] quantitative fluorescent PCR,[13] and multiplex ligation-dependent probe amplification,[14] have also been used on uncultured fetal cells to provide more rapid diagnosis of fetal aneuploidies. More recently, the introduction of high-resolution oligonucleotide and single nucleotide polymorphism (SNP) arrays has revolutionized prenatal diagnosis,[15] enabling more comprehensive and accurate analysis of a wider spectrum of chromosome abnormalities, including clinically significant chromosome deletions and duplications <5 Mb in size.[16]

Current arrays in clinical application are generally custom designed using a high density of oligonucleotide or SNP probes uniformly covering the backbone of each chromosome as well as exonic regions of disease genes.[3,16,17] Dual oligonucleotide and SNP combination arrays have been designed for very high-resolution chromosome analysis and can detect additional clinically significant abnormalities missed by oligonucleotide arrays.[18,19] Furthermore, with the availability of more detailed and comprehensive public databases of chromosome abnormalities and their associated clinical phenotypes, CNVs called by custom arrays can be readily translated into an accurate clinical diagnosis[20] and, in addition, has led to further knowledge on the genetic basis of more complex syndromes, such as autism.[21] However, occasionally, some rare CNVs not represented in databases still remain problematic to give a firm diagnosis.[22] Current data suggest that arrays can detect an additional 5% to 15% of chromosome abnormalities missed by karyotyping.[16,23] On the other hand, array-based diagnosis of polyploidies and balanced translocations still remains problematic. Based on a study of low- and high-risk pregnancies, it has been suggested that arrays could be used as a primary test to detect fetal chromosome abnormalities.[24]

The combined application of ultrasound, karyotyping, and array technologies, particularly in Western countries, has had a substantial effect on our ability to identify fetal abnormalities in pregnancy, to provide an explanation for recurrent miscarriage, and to give a diagnosis to children and adults who experience unknown physical and mental problems, enabling better treatment regimens to improve their quality of life.[16,25,26] Nonetheless, in developing countries and some developed countries, lack of accessibility, inadequate expertise, and high cost have precluded the widespread use of arrays, and, therefore, the incidence of children born with chromosome disease remains unacceptably high.[27–29] Therefore, there is an unmet clinical need for an alternative technology that is equally comprehensive and accurate as arrays for most chromosome diseases but more affordable to all patients. We, therefore, speculated that a next-generation sequencing–based technology would fulfill this need. We previously showed that a low-coverage shotgun sequencing method of approximately five million mapped sequencing reads allocated to sequential 20-kb sequencing bins across each chromosome can detect levels of mosaicism of the X chromosome down to 5%.[30] We, therefore, hypothesized that this method could be equally applied to detect CNV at a relatively high resolution across the 22 autosomal chromosome pairs plus the sex chromosomes X and Y. Herein, we applied CNV sequencing (CNV-seq) to blinded DNA samples with known abnormalities defined by a medium-density SNP array and showed that this method was highly concordant, reproducible, and sensitive, with a potential resolution of approximately 0.1 Mb.

Materials and Methods

Patient Samples

Karyotyping and SNP array analyses of patient samples were performed at the State Key Laboratory of Medical Genetics, Central South University (Hunan, China), and Hunan Jiahui Genetics Hospital (Hunan, China). Genomic DNA samples selected for the study were from 62 patients with either developmental delay/intellectual disability or congenital abnormalities. In addition, genomic DNA from 10 miscarriage samples were also included in the study. Details of these 72 samples are shown in Supplemental Table S1. Additional research samples for further assessment of the sensitivity and specificity of CNV-seq included three samples with ring chromosomes and four samples with small CNVs <0.25 Mb. RNA-free high-molecular-weight DNA was prepared from patient blood and miscarriage tissue using the DNeasy blood and tissue kit (Qiagen GmbH, Hilden, Germany). The quality and concentration of genomic DNA samples was assessed by agarose gel electrophoresis using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE).

SNP Array Analysis

Chromosome CNV analysis was performed using the HumanCytoSNP-12 BeadChip array (Illumina Inc., San Diego, CA), with an SNP probe density of 298,563 and average genome spacing of 19 kb. The log R ratio and the *A* and *B* allele frequency values were calculated using GenomeStudio software version 2011.1 (Illumina Inc.), and then detailed CNV analysis was performed using cnvPartition plug-in v3.1.6 software (Illumina Inc.). Duplications