The Journal of Molecular Diagnostics, Vol. 22, No. 12, December 2020

ELSEVIER

the Journal of Molecular Diagnostics
jmdjournal.org

Detection of Copy Number Variants by Short Multiply Aggregated Sequence Homologies

Vaidehi Jobanputra,*† Peter Andrews,‡ Vanessa Felice,* Avinash Abhyankar,* Lukasz Kozon,* Dino Robinson,* Ferrah London,* Inessa Hakker,‡ Kazimierz Wrzeszczynski,* and Michael Ronemus‡

From the New York Genome Center, New York; the Department of Pathology and Cell Biology,† Columbia University Medical Center, New York; and the Cold Spring Harbor Laboratory,‡ Cold Spring Harbor, New York*

Accepted for publication September 24, 2020.

Address correspondence to Vaidehi Jobanputra, Ph.D., Pathology and Cell Biology, Columbia University Medical Center, Molecular Diagnostics New York Genome Center, 101 Ave. of the Americas, New York, NY 10013; or Michael Ronemus, Ph.D., Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724. E-mail: vjobanputra@nygenome.org or ronemus@cshl.edu.

Chromosomal microarray testing is indicated for patients with diagnoses including unexplained developmental delay or intellectual disability, autism spectrum disorders, and multiple congenital anomalies. The short multiply aggregated sequence homologies (SMASH) genomic assay is a novel next-generation sequencing technology that performs copy number analysis at resolution similar to high-coverage whole genome sequencing but requires far less capacity. We benchmarked the performance of SMASH on a panel of genomic DNAs containing known copy number variants (CNVs). SMASH was able to detect pathogenic copy number variants of $\geq$10 kb in 77 of 77 samples. No pathogenic events were seen in 32 of 32 controls, indicating 100% sensitivity and specificity for detecting pathogenic CNVs $>$10 kb. Repeatability (interassay precision) and reproducibility (intra-assay precision) were assessed with 13 samples and showed perfect concordance. We also established that SMASH had a limit of detection of 20% for detection of large mosaic CNVs. Finally, we analyzed seven blinded specimens by SMASH analysis and successfully identified all pathogenic events. These results establish the efficacy of the SMASH genomic assay as a clinical test for the detection of pathogenic copy number variants at a resolution comparable to chromosomal microarray analysis. *(J Mol Diagn 2020, 22: 1476–1481; https://doi.org/10.1016/j.jmoldx.2020.09.009)*

Certain types of clinical genetic testing, including chromosome analysis and chromosomal microarray analysis (CMA), are standard practice for patients with suspected genetic diseases, including developmental delay/intellectual disability, autism spectrum disorders, and multiple congenital anomalies. These categories of disorders account for the largest proportion of cytogenetic testing because of their high prevalence in the population.[1] Most patients lack sufficient specific histories or features from physical examination to suggest specific genetic (or nongenetic) causes. Published guidelines for testing such patients have emphasized screening for chromosomal abnormalities by karyotyping and CMA as well as testing for single-gene disorders and panel or whole exome sequencing.

The utility of detecting constitutional genomic copy number gains and losses in patients with intellectual disabilities, autism, and other congenital anomalies has been well documented. CMA is now recommended as a first-tier test for neuropsychiatric and neurodegenerative indications. The short multiply aggregated sequence homologies (SMASH) genomic assay is a powerful new approach for detecting copy number variation in the human genome by next-generation sequencing. SMASH performs a similar function to CMA but at a higher resolution for genomic imbalances, comparable to that obtained by read-depth analysis of high-coverage whole genome sequencing (WGS).[2]

SMASH utilizes a combination of low-coverage WGS and a novel mapping algorithm to find copy number variants (CNVs) in the human genome. CMA testing yields a diagnosis in approximately 20% of sporadic disease cases,

Supported by the National Heart, Lung, and Blood Institute of the NIH award U0 1 HL 127522 (V.J. and M.R.);and a Simons Foundation Autism Research Initiative award SFARI 235988 (awarded to Michael Wigler, M.R.).

Disclosures: M.R. is founder of and holds an equity stake in Marvel Genomics PBC, which has licensed the short multiply aggregated sequence homologies technology from Cold Spring Harbor Laboratory.

https://doi.org/10.1016/j.jmoldx.2020.09.009

which supports its use as a diagnostic test for patients with suspected copy number disorders or simplex disease presentations of possible genetic cause and for patients with clinical diagnoses of heterogeneous genetic conditions.[1,3,4] Compared with CMA (the current American College of Medical Genetics and Genomics first-tier diagnostic test for identifying CNVs), SMASH can directly interrogate a far more diverse range of nucleotides within a human haplotype than is possible with oligonucleotide-based probes. Because it has a much higher signal/noise ratio than hybridization-based technologies, SMASH has superior sensitivity for calling CNVs relative to clinical CMA and can effectively call CNVs as small as 4 kb.[5] This is a substantial increase in resolution over what is typically possible with CMA. Although current microarray platforms can detect smaller CNVs, which range from 25 to 50 kb in size, this detection is dependent on the probe distribution in the genomic region and a requirement of certain number of probes in the genomic region. Because of this limitation, clinically significant CNVs may be missed in genomic regions with inadequate number and distribution of the probes.

To benchmark the performance of SMASH relative to CMA, we analyzed a set of cell lines and anonymous patient samples with known pathogenic CNVs. The CNVs were first detected by CMA or Sanger sequencing, and effectively represent the spectrum of events detectable by clinical CMA. We present herein our assay validation data for detection of constitutional copy number variation by SMASH.

Materials and Methods

Sample Selection

For this assay validation, a sample set of 115 DNA specimens with ($n = 83$) and without ($n = 32$) known pathogenic copy number variants was used. Most specimens ($n = 76$) were purchased from the Coriell Institute for Medical Research (Camden, NJ). A small number ($n = 7$) of de-identified DNA specimens were obtained from Columbia University Medical Center's Laboratory of Personalized Genomic Medicine (New York, NY), and the remaining ($n = 32$) were from the New York Genome Center (New York, NY).

Library Preparation and Sequencing

Random fragmentation of genomic DNA was performed as described previously using double-stranded DNA fragmentase.[5] DNA is then repaired and ligated into novel chimeric DNA molecules. Following size selection, the chimeric DNA molecules are ligated to standard Illumina paired-end sequencing adapters (Illumina, San Diego, CA), multiplexed using unique barcodes, and then amplified in clusters on a flow cell by means of short-cycle PCR. Paired-end sequencing (2×150) was performed on the Illumina HiSeq 2500 platform to provide a mean read depth of 2.5 million read pairs with $\geq$7.5 million maps per individual genome.

Informatics Analysis

SMASH utilizes a proprietary algorithm to identify multiple mappable tags from within each paired-end sequencing read. Each SMASH tag, or map, represents a unique match to the human genome. The maps are then divided into discrete genomic bins and tabulated to generate a digital count.[5] This output is segmented to generating a profile of genomic copy number using the circular-binary segmentation algorithm.[6] CNVs were visualized in the Nexus Copy Number software package version 10.0 (BioDiscovery Inc., El Segundo, CA).

Results

SMASH is a generalized WGS-based approach, and it does not target specific diseases or genetic variants. Correspondingly, the validation efforts were not focused on specific sequence contexts, but rather on developing and evaluating end-to-end metrics for high-quality sequencing. On the basis of Next-Generation Sequencing: Standardization of Clinical Testing workgroup recommendations,[7] a validation strategy was devised to measure and establish the performance characteristics of the SMASH assay. This strategy is described below.

Reference Range

To establish the reportable sequence variations the assay can detect that are expected to occur in an unaffected population, we selected the GRCh37/hg19 human reference genome build against which all sequence reads were evaluated. Raw sequence data in FASTQ format are mapped to human genome build GRCh37/hg19. All variant positions are reported in relation to this build.

Reportable Range

SMASH does not target any specific genomic regions. To determine which regions in the genome can be accessed, we computed map density across the genome and identified regions/bases that fail to meet minimum coverage thresholds. The human genome reference sequence has assembly gaps, and such gaps are represented using the ambiguity code N. We computed the size of the mappable genome, which represents the portion of reference genome that has no N sequence. The mappable genome is equal to the total length of contigs minus the gaps in reference. We only report high-quality variants identified by the SMASH assay within this mappable genome (2,861,959,612 bases).

Determination of Optimal Bin and Map Density

To estimate the optimal depth of overall genomic coverage required for SMASH, we sequenced multiplexed samples. On the basis of earlier estimates of the optimal depth of overall genomic coverage required for SMASH,[5] 40 samples were multiplexed per lane on the HiSeq 2500 for this analysis and processed through the SMASH Analytics version 1.0.1 pipeline (Figure 1); unique maps to the GRCh37 build of the human reference genome were tabulated. For sufficient genomic coverage to make copy number calls, a minimum of 2.5 million 150-bp paired-end reads per genome are required. When all additional sequencing quality metrics are met, ≥6 million high-quality maps are obtained. These mapped reads were divided into varying bin counts ranging from 20,000 to 500,000 and segmented as described earlier. Adjacent bins (≥3) showing deviation of identical polarity from the normal copy number state were required to make a CNV call.

Analytical Sensitivity and Specificity

To estimate the likelihood that the assay will detect known copy number variations, we assessed previously reported pathogenic CNVs from CMA from Coriell cell lines derived from patients as well as primary samples from a clinical laboratory. Concordance between the identified CNVs and previously reported pathologic CNVs was evaluated. A total of 115 DNA samples were assessed: 83 samples with previously reported pathogenic CNVs (76 cell lines from the Coriell Institute for Medical Research and 7 patient samples from Columbia University Medical Center) and 32 samples without any known pathogenic CNVs (from the New York Genome Center). We detected all 77 known pathogenic CNVs >7 kb in size, and found no pathogenic CNVs in the 32 negative samples (Supplemental Table S1). The SMASH assay was unable to detect six of six pathogenic CNVs in this experimental set that were <7 kb, ranging from 1.2 to 6.5 kb in size. We believe that this mainly reflects the relatively low depth of SMASH maps (and high degree of sample multiplexing) used in this validation. Greater read depth would likely allow identification of smaller events. On the basis of these data, we conclude that our assay has 100% sensitivity and specificity to detect pathogenic CNVs >10 kb when ≥6 million high-quality maps are obtained from an individual genome.

Precision

Repeatability (interassay precision) and reproducibility (intra-assay precision) measurements were performed on 15 samples, and agreement between these replicates was computed using a variety of metrics. The 15 samples included a range of CNVs, including aneuploidies as well as smaller gains and losses, and reflected the optimal number to load on the Illumina HiSeq 2500 Rapid Run flow cell used in these studies. The samples were sequenced in several replicates. For the interassay validation, the same

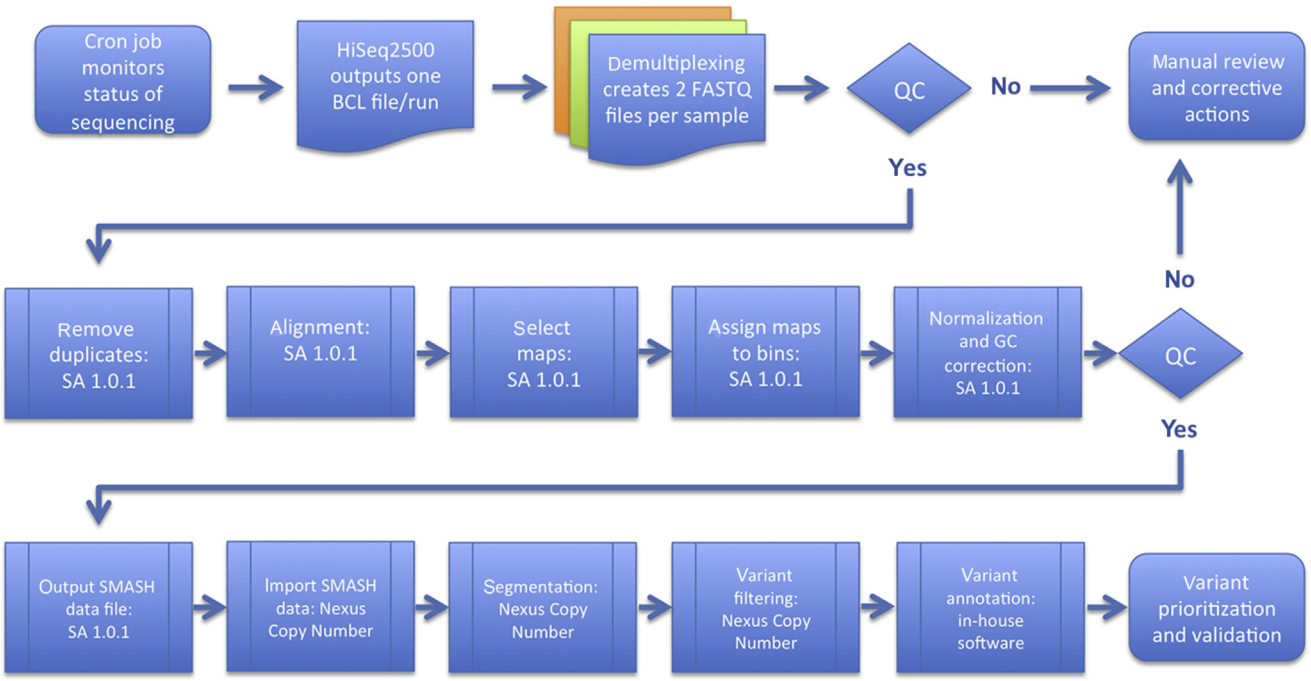

Figure 1 Schematic of the short multiply aggregated sequence homologies (SMASH) Analytics version 1.0.1 pipeline (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). The informatics processing steps following SMASH sequencing are detailed. SMASH utilizes a novel algorithm[5] to parse minimal unique matches from chimeric sequence reads. These maps are then normalized and tabulated to generate a digital output of copy number across the genome. This workflow was developed using the Illumina HiSeq 2500 instrument, but can be readily adapted to any high-throughput sequencing technology. BCL, binary base call; GC correction, guanine and cytosine base composition normalization; QC, quality control; SA, SMASH analytics.

pool of 15 samples was run on three different sequencing instruments on three separate days. For the intra-assay validation, 15 samples prepared in triplicate were run together on the same instrument on the same day. These sequencing runs are shown schematically in Supplemental Figure S1. Reproducibility was evaluated by comparing the pathogenic CNV calls between replicates (Table 1). All samples were concordant, illustrating the robustness of the SMASH technique.

Limit of Detection

To determine the limit of detection of mosaic CNVs, we performed a dilution experiment and determined the lower detection limit of the assay. A series of dilutions using NA12878 as the diluent, at 10%, 20%, 40%, and 80% for abnormal DNA, were performed in triplicate in three samples containing three types of CNVs (monosomy X, 22q11.2 deletion, and trisomy 21). As the dilution factor increased, the relative copy number (shown as linear ratios) for deletions decreases from approximately 1.8 in the 20% dilution to approximately 1.6 in the 40% dilution to approximately 1.2 in the 80% dilution (Supplemental Figure S2, A and B). Conversely, the dilution series of a trisomy 21 sample (Supplemental Figure S2C) showed only a slight increase from copy number 2 for chromosome 21 in a 10% dilution. As the dilution factor increases, higher copy number (linear ratio) became readily discernable, ranging from approximately 2.2 in the 20% dilution to approximately 2.6 in the 40% dilution and approximately 2.8 in the 80% dilution. For all samples, dilutions of 20% to 80% were detected accurately, but the dilution of 10% was difficult to detect. We therefore determined the limit of detection to be 20% for mosaic CNVs using SMASH. Although the limit of detection studies were performed only for aneuploidies and one representative recurrent deletion (22q11.2), we expect a similar detection limit for smaller CNVs using SMASH.

Parallel Testing

To independently assess performance of the SMASH assay, we sequenced seven blinded specimens (V.F., D.R., and F.L.) that were used for CNV analysis in a WGS assay in our laboratory. Following SMASH analysis, the results were compared and showed 100% concordance with the pathogenic CNV calls from this sample set. These samples are indicated (samples 30 to 36) in Supplemental Table S1.

Discussion

A marked evolution of the resolution of the conventional diagnostic cytogenetic testing has been observed since the introduction of the genomic microarrays. We first demonstrated the accuracy and sensitivity of oligonucleotide arrays to describe copy number changes in patients with chromosomal abnormalities using representational oligonucleotide microarray analysis.[8] The increased resolution of microarray technology over conventional cytogenetic analysis is well documented, and this is the current standard of care for both pediatric and prenatal cases suspected to have an underlying

Table 1 Comparison of Overlap between Expected Copy Number Event Genomic Loci

Sample ID	Expected genomic event	Size, bp	Copy number event	Inter-R1, %	Intra-R1/ inter-R2, %	Intra-R2, %	Intra-R3, %	Inter-R3, %	Concordant
SMASH-30	X:32807178-32954074	146,890	Loss	94.82	94.82	94.82	94.82	96.46	Yes
SMASH-32	1:6759458-10107452	3,347,990	Loss	99.97	100.00	99.97	100.00	99.97	Yes
SMASH-34	17:9474-6718180	6,708,710	Loss	100.00	100.00	100.00	100.00	100.00	Yes
SMASH-35	16:2119094-2232781	113,690	Loss	100.00	97.87	100.00	100.00	88.99	Yes
SMASH-36	22:19004735-21465659	2,460,930	Loss	99.87	99.32	99.10	99.32	99.87	Yes
NA06870-D	18:11542-15401751	15,390,210	Gain	91.54	91.54	91.54	91.54	91.54	Yes
NA20027-D	X:228465-155260560	155,032,100	Loss	100.00	100.00	100.00	100.00	100.00	Yes
NA08585-D	5:162075815-180719790	18,643,980	Gain	100.00	100.00	100.00	100.00	100.00	Yes
NA08585-D	6:165165530-170982522	5,816,990	Loss	100.00	100.00	100.00	100.00	100.00	Yes
NA23099-D	X:32536390-32716704	180,320	Gain	96.48	88.23	85.19	88.07	88.09	Yes
NA01059-D	7:100527710-125856601	25,328,890	Loss	100.00	100.00	100.00	100.00	100.00	Yes
NA10074-D	14:19002111-97463788	78,461,680	Gain	100.00	98.03	98.03	98.03	100.00	Yes
NA10074-D	X:72521611-155233846	82,712,230	Gain	99.85	99.85	99.85	99.85	99.85	Yes
NA04519-D	16:2032042-2101749	69,707	Loss	100.00	100.00	100.00	100.00	100.00	Yes
NA05117-D	X:31953078-32123855	170,780	Loss	95.15	95.15	95.15	95.15	95.15	Yes

SMASH-XX indicates blinded cohort (Columbia University Medical Center); NA-XXXXX-D indicates Coriell cell lines. Expected genomic event indicates the genomic coordinates of known copy number calls, and Size displays the copy number variant (CNV) sizes in base pairs (rounded to the nearest multiple of 10); these values were determined using Affymetrix Cytoscan HD microarrays (Affymetrix, Santa Clara, CA). Percentage values under replicate identity (Inter-R*/ Intra-R*) denote the overlap of the identified CNV (by SMASH) with expected genomic locations for the same calls. There was no substantive difference in the genomic content of the CNV events detected by the two methods.

ID, identification; SMASH, short multiply aggregated sequence homologies.

chromosomal abnormality or a pathogenic copy number event.[1]

Clinical CMA platforms can detect CNVs with a lower limit of resolution of approximately 200 kb throughout the genome.[9] With improvements in design and greater probe density, some microarrays are able to detect a limited number of imbalances as small as 20 to 50 kb in regions of the genome that contain comparatively little repetitive sequence. At this level of resolution, we are able to detect known recurrent microdeletion and microduplication syndromes mediated by segmental duplication as well as other nonrecurrent large pathogenic genomic imbalances.[10–12] CMA can detect CNVs affecting single genes and single exons in targeted regions using custom arrays and exon-level arrays. However, current genotyping-based microarrays lack the ability to identify smaller events across the genome that allow diagnosis of diseases caused by one haploinsufficient or triplosensitive gene, which limits the diagnostic capability of this testing. In addition, the resolution and yield of clinical CMA are limited by i) the genomic coverage (the length of and variable spacing between probes) on the microarray[13–16]; ii) the specific statistical algorithms used to set the criteria for gains and losses[6]; and iii) the high levels of system noise endemic to hybridization-based technology.[17] We demonstrate that the SMASH test represents an incremental improvement in terms of an expanded size range of CNVs detected that are not typically found by most CMA platforms.

At the moment, WGS is being used to interrogate all forms of variation, including single-nucleotide variants, small insertion/deletion variants, and structural variants, including CNVs. Although WGS is the gold standard for detecting most forms of genomic variation, it is not yet widely utilized for clinical testing because of the cost and complexity. This is not so much a consequence of the wet bench costs to generate data as it is due to the computational infrastructure required for analysis, as well as the large volume and spectrum of genetic variants identified. Each class of variants requires specialized algorithms to detect, and for many cases a simpler and more direct test for a specific class of genetic variants is sufficient. SMASH produces genome-wide copy number data that are nearly equivalent to read depth–based analysis of high-coverage WGS (29X to 32X) at a fraction of the cost, with an effective lower limit of detection of CNVs $\geq$4 kb in size.[5] In our data set, we show that the SMASH assay can reliably detect known pathogenic CNVs $\geq$10 kb (samples 71 to 73) (Supplemental Table S1).

SMASH has some of the same limitations as CMA in that it cannot detect balanced rearrangements, such as translocations or inversions. Unlike single-nucleotide polymorphism–based CMA, SMASH also lacks the ability to detect copy neutral regions of allelic homozygosity[5] that may include imprinted regions suggestive of uniparental disomy. When such regions of homozygosity (ROHs) are detected, clinicians typically order targeted sequencing of select candidate genes within the regions or uniparental disomy testing on the basis of a patient's phenotype.

Although analytical tools have been developed to facilitate the selection of genes that may explain the phenotype,[18] it is generally challenging to limit the list of candidate disease genes to be considered within extensive ROHs. This can only be used in the small number of cases in which the patient's clinical presentation strongly suggests a specific genetic disorder and when the number of genes consistent with the phenotype is limited; for most cases, this approach does not provide a diagnosis. Additional caveats are that CMA may not detect regions of homozygosity nor all types of uniparental disomies, and that most clinical laboratories limit the search for ROH to regions >10 Mb in which many candidate genes may reside.[19] Independently of the size of the ROH, a separate gene panel of potential candidate genes is required for each patient on the basis of his/her clinical features and the ROH identified. From a current clinical perspective, whole exome sequencing (WES) is preferable in such cases as it allows identification of genetic causes under an autosomal recessive model as well as other modes of inheritance, including heterozygous variants for X-linked or autosomal dominant inheritance patterns. WES has been proposed as a more effective approach following detection of ROH from single-nucleotide polymorphism–based CMA, because limiting the analysis to ROH would fail to find actual pathogenic variants outside the ROH.[20] WES or WGS is a standard approach for diagnosis of mendelian disorders following a negative CMA.[21–24] Similarly, because CMA can only detect isodisomy but not heterodisomy, it is not considered a standard test for the detection of uniparental disomy.[25] Furthermore, CMA cannot detect whole chromosome heterodisomy but is used to raise suspicion for uniparental disomy with crossover based on large blocks of ROH on a single chromosome, which then needs confirmation testing.

On the basis of this work, SMASH is superior to and can effectively replace standard CMA for detecting CNVs $\geq$10 kb. SMASH provides an effective means of assessing genome-wide copy number without the inherent bias and sequence limitations of hybridization-based technologies, and it is potentially applicable to a wide variety of indications in both oncological and constitutional genomic testing. Because the efficiency of SMASH scales directly with increasing read depth and length in short-read WGS, SMASH is inherently more cost-effective than even low-coverage WGS as it provides multiple maps per read pair and can therefore be multiplexed to a higher degree.[5] The entire SMASH procedure, from DNA extraction and library construction to Illumina sequencing and data analysis, can be performed in 2 to 3 days. SMASH resolution is highest in exonic regions that are less repetitive than the overall human genome, which are the regions most likely to produce pathogenic copy number variants. The random fragmentation of genomic DNA that produces SMASH sequencing reads addresses an inherent weakness in targeted sequencing

approaches, such as WES, that do not provide balanced genomic representation. The combination of SMASH and WES in genomic analysis provides a powerful and inexpensive means for detecting many pathogenic genetic variants compared with high-coverage WGS.

Supplemental Data

Supplemental material for this article can be found at *http://doi.org/10.1016/j.jmoldx.2020.09.009*.

References

1. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, Church DM, Crolla JA, Eichler EE, Epstein CJ, Faucett WA, Feuk L, Friedman JM, Hamosh A, Jackson L, Kaminsky EB, Kok K, Krantz ID, Kuhn RM, Lee C, Ostell JM, Rosenberg C, Scherer SW, Spinner NB, Stavropoulos DJ, Tepperberg JH, Thorland EC, Vermeesch JR, Waggoner DJ, Watson MS, Martin CL, Ledbetter DH: Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet 2010, 86:749−764
2. Trost B, Walker S, Wang Z, Thiruvahindrapuram B, MacDonald JR, Sung WWL, Pereira SL, Whitney J, Chan AJS, Pellecchia G, Reuter MS, Lok S, Yuen RKC, Marshall CR, Merico D, Scherer SW: A comprehensive workflow for read depth-based identification of copy-number variation from whole-genome sequence data. Am J Hum Genet 2018, 102:142−155
3. Manning M, Hudgins L; Professional Practice and Guidelines Committee: Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. Genet Med 2010, 12:742−745
4. South ST, Lee C, Lamb AN, Higgins AW, Kearney HM; Working Group for the American College of Medical Genetics; Genomics Laboratory Quality Assurance Committee: ACMG standards and guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. Genet Med 2013, 15:901−909
5. Wang Z, Andrews P, Kendall J, Ma B, Hakker I, Rodgers L, Ronemus M, Wigler M, Levy D: SMASH, a fragmentation and sequencing method for genomic copy number analysis. Genome Res 2016, 26:844−851
6. Olshen AB, Venkatraman ES, Lucito R, Wigler M: Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 2004, 5:557−572
7. Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, et al: Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotechnol 2012, 30:1033−1036
8. Jobanputra V, Sebat J, Troge J, Chung W, Anyane-Yeboa K, Wigler M, Warburton D: Application of ROMA (representational oligonucleotide microarray analysis) to patients with cytogenetic rearrangements. Genet Med 2005, 7:111−118
9. Papenhausen P, Schwartz S, Risheg H, Keitges E, Gadi I, Burnside RD, Jaswaney V, Pappas J, Pasion R, Friedman K, Tepperberg J: UPD detection using homozygosity profiling with a SNP genotyping microarray. Am J Med Genet A 2011, 155A:757−768
10. Rosenfeld JA, Patel A: Chromosomal microarrays: understanding genetics of neurodevelopmental disorders and congenital anomalies. J Pediatr Genet 2017, 6:42−50
11. Fan Y, Wu Y, Wang L, Wang Y, Gong Z, Qiu W, Wang J, Zhang H, Ji X, Ye J, Han L, Jin X, Shen Y, Li F, Xiao B, Liang L, Zhang X, Liu X, Gu X, Yu Y: Chromosomal microarray analysis in developmental delay and intellectual disability with comorbid conditions. BMC Med Genomics 2018, 11:49
12. Scionti F, Di Martino MT, Pensabene L, Bruni V, Concolino D: The Cytoscan HD Array in the diagnosis of neurodevelopmental disorders. High Throughput 2018, 7:28
13. Friedman J, Adam S, Arbour L, Armstrong L, Baross A, Birch P, Boerkoel C, Chan S, Chai D, Delaney AD, Flibotte S, Gibson WT, Langlois S, Lemyre E, Li HI, MacLeod P, Mathers J, Michaud JL, McGillivray BC, Patel MS, Qian H, Rouleau GA, Van Allen MI, Yong SL, Zahir FR, Eydoux P, Marra MA: Detection of pathogenic copy number variants in children with idiopathic intellectual disability using 500 K SNP array genomic hybridization. BMC Genomics 2009, 10:526
14. Moeschler JB, Shevell M; American Academy of Pediatrics Committee on Genetics: Clinical genetic evaluation of the child with mental retardation or developmental delays. Pediatrics 2006, 117:2304−2316
15. Shaffer LG, Beaudet AL, Brothman AR, Hirsch B, Levy B, Martin CL, Mascarello JT, Rao KW; Working Group of the Laboratory Quality Assurance Committee of the American College of Medical Genetics: Microarray analysis for constitutional cytogenetic abnormalities. Genet Med 2007, 9:654−662
16. Warburton D, Ronemus M, Kline J, Jobanputra V, Williams I, Anyane-Yeboa K, Chung W, Yu L, Wong N, Awad D, Yu CY, Leotta A, Kendall J, Yamrom B, Lee YH, Wigler M, Levy D: The contribution of de novo and rare inherited copy number changes to congenital heart disease in an unselected sample of children with conotruncal defects or hypoplastic left heart disease. Hum Genet 2014, 133:11−27
17. Lee YH, Ronemus M, Kendall J, Lakshmi B, Leotta A, Levy D, Esposito D, Grubor V, Ye K, Wigler M, Yamrom B: Reducing system noise in copy number data using principal components of self-self hybridizations. Proc Natl Acad Sci U S A 2012, 109:E103−E110
18. Wierenga KJ, Jiang Z, Yang AC, Mulvihill JJ, Tsinoremas NF: A clinical evaluation tool for SNP arrays, especially for autosomal recessive conditions in offspring of consanguineous parents. Genet Med 2013, 15:354−360
19. Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF: Genomic runs of homozygosity record population history and consanguinity. PLoS One 2010, 5:e13996
20. Prasad A, Sdano MA, Vanzo RJ, Mowery-Rushton PA, Serrano MA, Hensel CH, Wassman ER: Clinical utility of exome sequencing in individuals with large homozygous regions detected by chromosomal microarray analysis. BMC Med Genet 2018, 19:46
21. Stavropoulos DJ, Merico D, Jobling R, Bowdin S, Monfared N, Thiruvahindrapuram B, et al: Whole genome sequencing expands diagnostic utility and improves clinical management in pediatric medicine. NPJ Genom Med 2016, 1:15012
22. Tan TY, Dillon OJ, Stark Z, Schofield D, Alam K, Shrestha R, Chong B, Phelan D, Brett GR, Creed E, Jarmolowicz A, Yap P, Walsh M, Downie L, Amor DJ, Savarirayan R, McGillivray G, Yeung A, Peters H, Robertson SJ, Robinson AJ, Macciocca I, Sadedin S, Bell K, Oshlack A, Georgeson P, Thorne N, Gaff C, White SM: Diagnostic impact and cost-effectiveness of whole-exome sequencing for ambulant children with suspected monogenic conditions. JAMA Pediatr 2017, 171:855−862
23. Taylor JC, Martin HC, Lise S, Broxholme J, Cazier JB, Rimmer A, et al: Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nat Genet 2015, 47:717−726
24. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, Hardison M, Person R, Bekheirnia MR, Leduc MS, Kirby A, Pham P, Scull J, Wang M, Ding Y, Plon SE, Lupski JR, Beaudet AL, Gibbs RA, Eng CM: Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med 2013, 369:1502−1511
25. Hoppman N, Rumilla K, Lauer E, Kearney H, Thorland E: Patterns of homozygosity in patients with uniparental disomy: detection rate and suggested reporting thresholds for SNP microarrays. Genet Med 2018, 20:1522−1527