TITLE: Whole Genome Analysis Using Microarrays

PRESENTER: Elena A. Repnikova, PhD, FACMG

Slide 1:
Hello, my name is Elena Repnikova. I am an Assistant Director of Cytogenetics and Molecular Genetics Laboratories at Children's Mercy Hospital in Kansas City and an Assistant Professor of Pathology and Pediatrics in the University of Missouri Kansas City Medical School. Welcome to this Pearl of Laboratory Medicine on "Whole Genome Analysis Using Microarrays".

Slide 2:
Copy number variation (CNV) is a form of structural genome variation which leads to changes in copy number of a particular genomic region from one individual to another. According to recent study, between 4.8 to 9.5% of human genome is estimated to contribute to CNV. Phenotypic consequences due to CNV can vary from no obvious effect to disease causing or even embryonic lethality. CNV may also contribute to the development of cancer. Characterization of such DNA-based changes is important for both the basic understanding of genetic disease and diagnosis. However, it is often difficult to interpret the effect of CNVs on human health due to unknown functional significance of these regions.

Slide 3:
A lot of human genetic disorders result from gains or losses of genetic material. Deletions or duplications of chromosomal material of approximately 5Mb or more in size can be detected by traditional chromosome banding analysis. Smaller (80-200 kb) genetic alterations can be detected by fluorescent in situ hybridization (FISH). In recent years, a newer technique called microarray has been extensively developed. Cytogenomic microarrays have been universally adopted in clinical diagnostics owing to the comprehensive nature of genome-wide analysis. Microarray analysis is accepted as an appropriate first-tier test for evaluation of imbalances associated with intellectual disability, autism, and congenital anomalies. The technology is also used in the clinical prenatal setting and may be useful in evaluation of fetal demise or stillbirth. In the last years, use of microarray in the evaluation of hematological malignancies and solid tumors have added a significant value to the diagnosis and prognosis of both benign and malignant neoplasms.

© 2015 *Clinical Chemistry*

Slide 4:
There are several commercially available platforms to study CNVs. Companies such as Agilent, Affymetrix, and Illumina offer platforms with various coverage of genome. Microarray platforms designs can be variable too, with probes 1) targeted to specific genomic regions, 2) distributed genome-wide with a determined spacing and density, or 3) both targeted to a specific genomic region as well as distributed across the genome. Depending on the overall probe density and probes for a specific genomic region, functional resolution across different regions of the genome for a given platform may be different. Probe density for a given platform also determines accuracy of the CNV call. Extensive validation to assess performance and reproducibility of a chosen array platform should be performed by the clinical diagnostic laboratory that decides to incorporate microarray analysis into their testing practice.

Slide 5:
Two types of microarrays are typically used for detection of CNV: array-based comparative genomic hybridization (array CGH) and SNP-based. Array CGH uses a two-color system, in which a patient DNA is labeled with one fluorescent dye and a control, normal DNA is labeled using a second fluorescent dye. Both patient and control DNA are simultaneously hybridized onto the probes spotted on a chip. Genetic alterations (losses or gains) in the patient DNA can be detected based on the differences in fluorescence intensity of the two dyes at various genomic locations. SNP-based array is the type of DNA microarray that detects single-nucleotide polymorphisms (SNPs) in the population. SNP arrays do not directly compare a patient and a control specimen. Instead, in the SNP array patient DNA is labeled with fluorochrome, and the intensity is compared to a reference DNA in silico (a database of control individuals) to derive the log2 ratio, an intensity ratio of SNP and copy number probes in the patient DNA, which helps to determine if there is a gain or loss of genomic material.

The choice to use array CGH or SNP-based array depends on several important factors, such as resolution of the platform and ability to customize probe content.

Slide 6:
Both array CGH and SNP array can detect gains and losses of the genome, as well as mosaic copy number changes that are present in greater than about 10-20% abnormal cells. In comparison to array CGH, the combination of multiple SNPs and relevant bioinformatics analysis allows the SNP array to detect regions with absence of heterozygosity as well as ploidy changes. Balanced chromosomal rearrangements cannot be routinely identified by array CGH or SNP array.

Slide 7:
The clear advantage of microarray analysis is the high resolution and comprehensive evaluation of genetic imbalances in a single analysis. In comparison to classical chromosome analysis, which for example, only detects chromosome abnormality in ~3-4% of children with congenital anomalies and intellectual and developmental disabilities, microarray abnormality detection rate is ~15-20%. DNA copy number variations are mapped to the specific chromosome location, which allows for examination of gene content within the region of CNV. Knowledge of chromosomal abnormality is not required prior to microarray testing as opposed to FISH analysis, which requires a physician to suspect the diagnosis in order to choose the right probe. Small amount of DNA is necessary to perform microarray analysis. DNA

© 2015 *Clinical Chemistry*

can be extracted directly from the tissue without the need to set up culture, which is especially important for products of conception or tumors in which culture failure can be anticipated.

Slide 8:
Despite microarray testing being advantageous in comparison to chromosomal or FISH analysis, it has several limitations. For example, balanced chromosomal rearrangements, such as translocations or inversions, cannot be routinely identified by microarrays. Several companies offer fairly large coverage of the genome. Despite that, there are always regions which lack probes (generally due to difficulties in their designing for those regions). Therefore, any imbalances in these regions will not be detected. Microarray may not be sensitive enough to detect low-level mosaic abnormalities in constitutional samples or low-level clonal changes in tumors. In addition, copy number variations detected by microarray may be difficult to interpret clinically due to their unknown significance on human development and health.

Slide 9:
In clinical diagnostics, the specific nomenclature has been refined to encompass changes detected by microarray analysis. If no CNVs were detected using any type of array platform, the results are expressed as follows. The sex chromosomes are separated with comma from autosomes, which are listed first. If the results are abnormal, only the aberrations are listed. Sex chromosome abnormalities are listed first, followed by the lowest chromosome number, regardless of whether it is a gain or loss of genetic material. Aberrant nucleotides should be listed from the p terminus (*pter*) of the chromosome to the q terminus (*qter*) of the chromosome. The ***arr*** precedes the chromosome band designation involved in the abnormality, followed by genomic coordinates of the boundary probes involved in the deleted or duplicated segment. Outside the parenthesis, the copy number of the abnormal segment present in genome is listed. The loss of 1 copy is designated as x1, the loss of two copies is designated as x0, the gain of an extra third copy is designated as x3, and so on.

The descriptive interpretation of array findings should also include the platform used, its resolution, and the genome build used for analysis and interpretation of data.

Slide 10:
SNP arrays, in addition to copy number variations, can identify regions of homozygosity. The symbols **hmz** and **htz** can be used to define zygosity of the chromosomal region.

Slide 11:
A majority of clinical laboratories are using a 5-tiered system for the classification of array CNVs: pathogenic, likely pathogenic, variant of unknown significance, likely benign, and benign. CNV detected by array is assessed for its size, gene content, associations with benign CNV of this region, and mode of inheritance, if known. In general, gene-rich abnormalities are thought to be more significant than gene-poor. The duplications are thought to be better tolerated than deletions. However, the significance of the finding is also dependent on the presence of CNV in normal populations, such as those collected within the Database of Genomic Variants. Determination of pathogenic nature of a CNV can be done through examination of online databases such as DECIPHER, OMIM, ISCA, etc, which are listed on the

© 2015 *Clinical Chemistry*

next slide. Variants of unknown significance copy number changes represent a large portion of microarray findings. They typically include genomic alterations with no established connection to abnormal phenotypes. These changes usually require a significant amount of time to investigate and describe appropriately in the report.

Slide 12:

The links below are helpful to examine the region involved in the abnormality (through UCSC browser), determine if the abnormality has ever been reported in the control, healthy individuals (through Database of Genomic Variants), determine if abnormality has ever been reported in association with abnormal phenotypes (through DECIPHER and ISCA databases, and Clinical Genome Resource).

Slide 13:

In the next few slides, we will review examples of CNV detection by Affymetrix CytoScan HD platform and how these abnormalities are displayed in the analysis software.

Slide 14:

This example shows an interstitial ~5Mb deletion on chromosome 18 detected by Affymetrix Cytoscan HD SNP array and displayed by ChAS software. Within the highlighted area, the log2 ratio is equal to ~-0.7, which translates into 1 copy of this region being deleted, and there are only two allele peak tracks (A and B). Deletions of this region are pathogenic and have been associated with several developmental phenotypes.

Slide 15:

This example shows an interstitial ~4Mb duplication on chromosome 8 detected by Affymetrix Cytoscan HD SNP array and displayed by ChAS software. Within the highlighted area, the log2 ratio is equal to ~0.45, which translates into 3 copies of this region instead of two per diploid genome, and there are four allele peak tracks (AAA, AAB, ABB, and BBB). Duplications of this region are pathogenic and have been associated with delays and learning problems.

Slide 16:

This example shows triploidy detected by Affymetrix Cytoscan HD SNP array platform. The copy number state is equal to two, the log2 ratio is equal to zero; however, there are four different allele peak tracks (AAA, AAB, ABB, and BBB). The example demonstrates how triploidy can be detected by SNP probes. In comparison, triploidy cannot be detected by array CGH platform.

Slide 17:

This example demonstrates the region with absence of heterozygosity (AOH) detected by the SNP array platform. The Log2 ratio is equal to zero, which indicates that there are two copies of chromosome 15. However, only two allele peak tracks are detected in the 15q11.2q12 region, indicating this region has lost heterozygosity. The region with absence of heterozygosity may indicate a region identical by descent (IBD), which may raise the likelihood of a recessive disorder in the region. Or this finding may

indicate full/segmental uniparental disomy (UPD) of chromosome 15 (both chromosomes inherited from the same parent). This region contains two important genes, *SNRPN* and *UBE3A*. UPD of chromosome 15 is known to be associated with Prader-Willi or Angelman syndrome. Confirmation of UPD would raise a suspicion for Prader-Willi or Angelman syndrome. Further molecular analysis is necessary to distinguish between IBD and UPD and determine if both chromosomes 15 are maternal or paternal to differentiate between Prader-Willi and Angelman syndrome diagnosis.

Slide 18:

This example demonstrates an ~15% mosaic gain for a large segment in the long arm of chromosome 7. The log2 ratio is slightly above zero, and the allele tracks start to split, but these two things are easy to miss. Smooth signal representing the copy number state clearly rises above the 2 line providing an indication for a mosaic gain of this region. The gain of this region was found to be attached to the telomeric end of chromosome 1, which couldn't be detected by microarray due to absence of probes in the telomeric regions of chromosomes. In this case, chromosome or FISH analysis are an important adjunct to microarray testing to determine the location of the duplicated segment, mechanism, and potential recurrence risks in the family.

Slide 19: References

Slide 20: References

Slide 21: Disclosures

Slide 22: Thank You from www.TraineeCouncil.org

Thank you for joining me on this Pearl of Laboratory Medicine on "Whole Genome Analysis Using Microarrays." My name is Elena Repnikova.