

## SNPs ANALYSIS AS A TOOL IN MOLECULAR GENETICS DIAGNOSTICS

Dewi Rusnita

### Abstrak

*Single Nucleotide Polymorphism* (SNP) merupakan variasi genetik yang ditemukan pada lebih dari 1% populasi. Haplotipe, yang merupakan sekelompok SNP atau alel dalam satu kromosom, dapat di turunkan ke generasi selanjutnya dan dapat digunakan untuk menelusuri gen penyebab penyakit (marker genetik). Artikel ini bertujuan menjelaskan aplikasi analisis SNP dalam diagnosis beberapa sindrom yang disebabkan gangguan genetik. Berdasarkan laporan studi terdahulu, sindrom yang disebabkan oleh UPD (*uniparental disomy*) maupun penyakit autosomal resesif yang muncul sebagai akibat perkawinan sedarah dapat dideteksi dengan SNP array melalui analisis *block of homozygosity* dalam kromosom. Kelebihan lain SNP array adalah kemampuannya dalam mendeteksi *mosaicism* level rendah yang tidak terdeteksi dengan pemeriksaan sitogenetik konvensional. Bahkan saat ini, SNP array sedang diujicobakan dalam IVF untuk mendapatkan bayi yang sehat. Hal ini dapat dilakukan dengan mendeteksi ada atau tidaknya gen tunggal penyebab penyakit pada embrio hasil bayi tabung sebelum embrio ditanamkan ke uterus. Analisis SNP dengan SNP array mempunyai banyak kelebihan dibanding metode pemeriksaan SNP lainnya dan diharapkan dapat digunakan secara luas dalam bidang diagnostik molekuler genetik di masa mendatang.

**Kata kunci:** SNP, SNP array, block of homozygosity, diagnostik, penyakit autosomal resesif

### Abstract

*Single Nucleotide Polymorphism* (SNP) is a genetic variant with a frequency of >1% of a large population. Haplotypes, a combination of a set of SNPs/alleles that appear as "associated blocks" on one chromosome, tend to be inherited together to the next offspring and can be used as genetic markers to trace particular diseases. This article aimed at explaining of SNP analysis application in diagnosis of genetic-disorder related syndrome. Previous studies showed that syndromes caused by UPD or autosomal recessive disorder as a result of consanguineous marriage can be identified by SNP array through analysing *block of homozygosity* region in a chromosome. Another advantage of SNP arrays is its ability in detecting low level *mosaicism* which was unidentified by conventional cytogenetic examination. Nowadays, SNP arrays are included in IVF process to obtain a healthy baby. It can be done by detecting the absence or the presence of disease-causing single gene in an embryo before it implanted to the womb. SNP analysis with SNP array has many advantages over other SNP analysis methods and is therefore expected can be widely used in the future in the field of molecular diagnostic.

**Keywords:** SNP, SNP array, block of homozygosity, diagnostic, autosomal recessive disorder

**Afiliasi Penulis :** Anatomy Department, Medical Faculty of Andalas University, Padang, West Sumatera, Indonesia. Correspondence: [dewirusnita@gmail.com](mailto:dewirusnita@gmail.com), HP: 081374160017

## INTRODUCTION

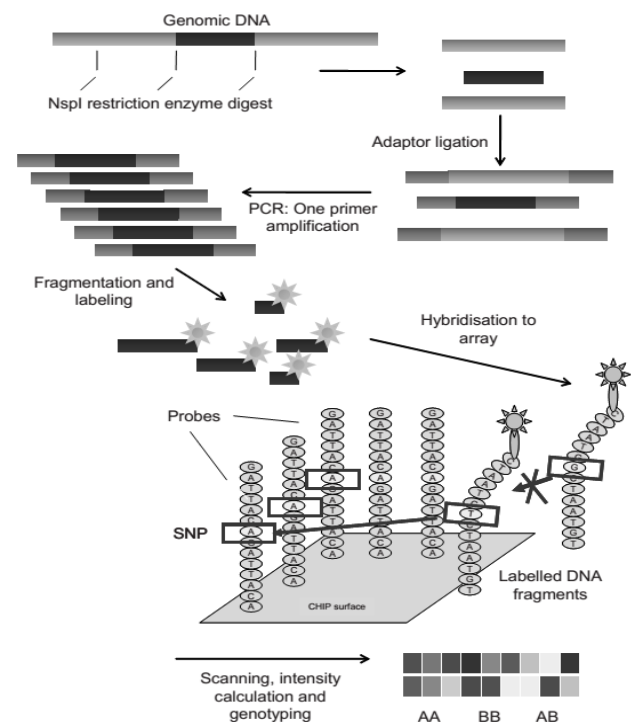
Single nucleotide polymorphism (SNP) is a single base pair change in a DNA sequence that is a genetic variant with a frequency of >1% of a large population. In human 99,9% of DNA sequence is identical and there is only 0.1% of polymorphisms which give unique characteristic of each individual. There are approximately 10 million SNPs spread throughout the human genome, which means that SNPs present in every 300 bp of DNA sequence<sup>1</sup>.

Most SNPs (2/3) are located in non-coding regions (introns), and the remaining are in the coding regions. In general one SNP has two possible genotype/allele (A/T or C/G or called 'A' or 'B'), but although tri- or four-allelic are rare, they do exist. Alleles are assigned to a chromosome forming what are called haplotypes, and each individual has two haplotypes. Haplotypes are defined as a combination of a set of SNPs/alleles that appear as "associated blocks" on one chromosome that tend to be inherited together<sup>2</sup>.

This SNPs haplotype blocks can be used as genetic markers to trace particular diseases (means indirectly tracking the causative genes). As SNP can be used as a genetic marker, an advanced technology such as SNP array platform (figure 1) was designed to identify the SNP across the whole genome. This was done by employing hundreds to thousands of probes (made of short oligonucleotides) which correspond to the SNPs across the whole genome in order to facilitate identification of disease-gene loci near the SNPs marker. The technique allows for mapping blocks of homozygosity in

order to select the candidate genes in a family with unknown mutation as well as discovering novel genes that are the cause of a disease. In addition, it can also lessen the need for sequencing numerous genes in a complex disease<sup>3</sup>.

Other technique which is also used for genotyping is TaqMan. It is a simple PCR-based assay, high-throughput but expensive<sup>4</sup>. Meanwhile, SNP array is more cost-effective and it can detect loss of heterozygosity and copy number variation as well. Currently, SNP array is commonly used for diagnostic purposes (prenatal and postnatal setting) and research.



**Figure1. Workflow of Affimetrix SNP array.** The genomic DNA that contains the SNP is digested by a restriction enzyme resulting in various sizes of DNA fragments that are then followed by adaptor ligation, then this ligated contains SNP-fragments are amplified by using one primer, the amplicon (200-1100bp in size) is then fragmented and labelled with fluorochrome. The samples were then hybridized to the arrays that con-

tained thousands of oligonucleotide probes that correspond to two allele of SNPs. The DNA fragment will only bind to its allele specific perfect match probes. Subsequently, the array is washed, scanned and the computer software generates the data based on the intensity of fluorescent signal of the SNP to determine the copy number variation (deletion or duplication) and the genotype of the SNP based on allele ratio (*ie*, AA, AB, BB)<sup>5</sup>.

## METHODS

A literature-based was used to obtain any information related to SNP analysis methods and its application in molecular genetics diagnosis.

## RESULTS AND DISCUSSION

In post natal settings, array CGH and SNP array are employed in the diagnostic field as first line investigation to identify the cause of unexplained diseases/phenotypes observed in the patients, such as developmental delay, intellectual disability, congenital abnormalities or whenever conventional cytogenetic analysis gave a normal result. Both techniques can detect subtle copy number variations (CNV) such as microdeletion or microduplication which may be responsible for the phenotype seen in patient. Other advantages of SNP arrays are that SNP array provides genotype information (since it contains high density of SNP probes) makes it enable to detect low-level mosaicisms, and mapping region of homozygosity (that was marked by the software based on the loss of heterozygosity alleles markers) in chromosomes which could represent uni parental disomy (UPD), consanguinity or identity by descent (means the alleles inherited/shared is identical from the common ancestor)<sup>6</sup>.

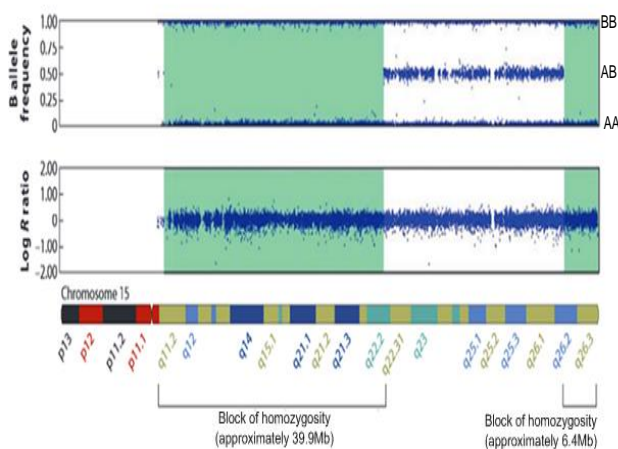
Neither technique can detect balanced rearrangement such as balanced translocations and inver-sions. In addition, whenever SNP arrays gives a normal result that may be due to small genome changes involving one exon only, whole-exome sequencing can be employed<sup>7</sup>.

### SNP array in detecting UPD

Uni parental disomy (UPD) is a term used to define the inheritance of two copies of chromosome pairs from a single parent (either mother or father) without the presence of the other parent's chromosome due to mitotic or meiotic errors. UPD can be result of trisomy rescue (loss of one of the three chromosomes of a fertilized egg and remaining both copies from one parent), monosomy rescue (duplication of monosomic chromosome due to fertilization of a nullisomic gamete with a normal gamete) and gamete complementation. There are types of UPD; isodisomy (inheriting both copies of identical homologue chromosome of the parent as a result of monosomy rescue) and heterodisomy (inheriting two copies of both parental homologue chromosomes as a result of trisomy rescue and gamete complementation)<sup>8</sup>.

In the SNP array result, isodisomy can be identified by the presence of block of homozygosity in one chromosome only, meanwhile heterodisomy cannot be detected unless it presents with isodisomy interspersed with heterozygous alleles (hetero-isodisomy) indicated by recombination (figure 2). UPD does not necessarily mean pathogenic, unless it involves the imprinting region which is present in chromosomes 6, 7, 11, 14, and 15 (the

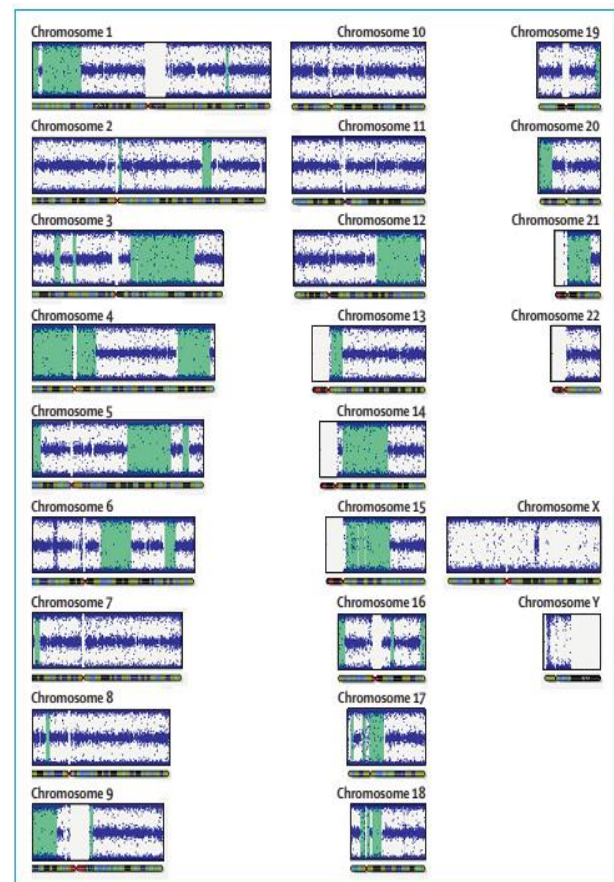
symptoms or the syndrome observed will depend on the parental origin) or when it unmasks an autosomal recessive disorders (if the child inherit two copies of identical chromosome from a single parent that harbour a recessive mutation) <sup>6,8,9</sup>. However, when SNP arrays gives a normal result, it does not exclude the possibility of UPD since complete heterodisomy cannot be detected by SNP arrays. If there is a strong suspicion to UPD, a methylation test can be performed<sup>10</sup>.



**Figure 2. Example of SNP array result in chromosome 15 with UPD.** Log R ratio represents copy number variation. There is no deletion or duplication seen in this figure since the blue line is in zero region. If the blue line shifted to above zero (indicating duplication), if it is shifted to below zero (means deletion). B allele frequency represents the proportion of genotype calls for the homozygous alleles (AA=0, BB=1) and the heterozygous allele (AB=0.5 in the middle). The region of hetero-isodisomy was shown within the homozygosity blocks (green block) that interspersed with heterozygosity line. If there is only long contiguous stretches of homozygosity in one chromosome, isodisomy can be assumed <sup>7</sup>.

## SNP array reveal identity by descent/ consanguinity status

SNP arrays have never been intentionally used to uncover parental relationships. During analysis the arrays result of a child with congenital abnormalities and developmental disabilities, the genetic relationship (close relatives) between parents could be incidentally revealed (figure 3), and can even uncover non-paternity <sup>11</sup>.



**Figure 3. Example of SNP arrays result which indicates consanguinity.** There were found several block of homozygosity (represented by green block) in several chromosomes with the total of block of homozygosity is 668MB (indicate the individual tested, a child with multiple abnormality in this case, is derived from consanguineous family with first relative degree). The degree of relatives can be predicted by sum up the size of block of

homozygosity present in each chromosome. The total size of block of homozygosity expected for first degree relative is 716Mb, or for second degree relatives is 358Mb<sup>11</sup>.

As yet, laboratory staff and clinician have no guide in the reporting of this incidental finding. As this finding can raise ethical and legal concerns, it is essential to establish a guideline for laboratories and clinician in reporting and disclosing this unexpected result. On one hand, clinicians can legally report this result to the child protection services or even to the police whenever there is suspicion of an incestuous relationship (due to perhaps sexual abuse by the relatives) of a minor girl, the other hand, clinicians' duty is less clear for reporting this when the case is an adult mother<sup>11</sup>.

Consanguinity and identity by descent can increase the risk and recurrence risk of autosomal recessive disorders. By taking the benefit of blocks of homozygosity (which indicate the presence of the same SNP on both alleles or loss of heterozygosity) shared by two affected people in a consanguineous family, can aid the clinician to narrow the differential diagnosis (as sometimes the phenotypes are similar even though the disease-causing genes are different) and narrowing the genes to be tested to make a correct diagnosis and to give the correct treatment to the patient particularly in a rare disease which is commonly present in a homozygous state instead of heterozygous.

As reported by Sund (2013), two affected children who suffer from mental disability, seizures and hypotonia, and who were thought to have ACOX deficiency (the gene encode this enzyme is found on 17q25) by clinician, in fact, both of the children shared a block of

homozygosity at 5q23.1 which included *HSD17B4* (mutation in this gene has similar symptoms with ACOX deficiency), and they had consanguineous parents. Another example were sisters who had recurrent seizures, developmental delay and were resistant to various antiepileptic drugs, both of them were found to share a block of homozygosity on 17q21.32 which include *PNPO* gene (encode enzyme of pyridoxine phosphate oxidase). This finding was followed by sequencing for this gene and the non synonymous mutation was found. Subsequently, the therapy of pyridoxine was administered to them, and this therapy can control the seizures<sup>12</sup>.

### SNP analysis in prenatal diagnosis

In the prenatal setting, SNP array with high density SNP markers (Illumina) allow the identification of inherited genetic disorders (particularly single gene disorders) and chromosomal abnormalities of an embryo of IVF (*in vitro* fertilization) technique by amplifying the DNA with multiple displacement amplification (MDA) from one or two cells obtained from the blastocyst (stage of embryo cleavage) which were then hybridized to a SNP array and the raw data are generated by the software based on linkage approach. This process is performed prior to implantation. As it is known that the major problems with single cell level testing is allele drop out (ADO). Karyomapping has eliminated this problem by only taking into account the informative heterozygote marker calls of the sample<sup>13</sup>.

By comparing and analysing the linkage of key SNP markers (informative heterozygote marker) of four parental haplotypes, a reference child (usually the

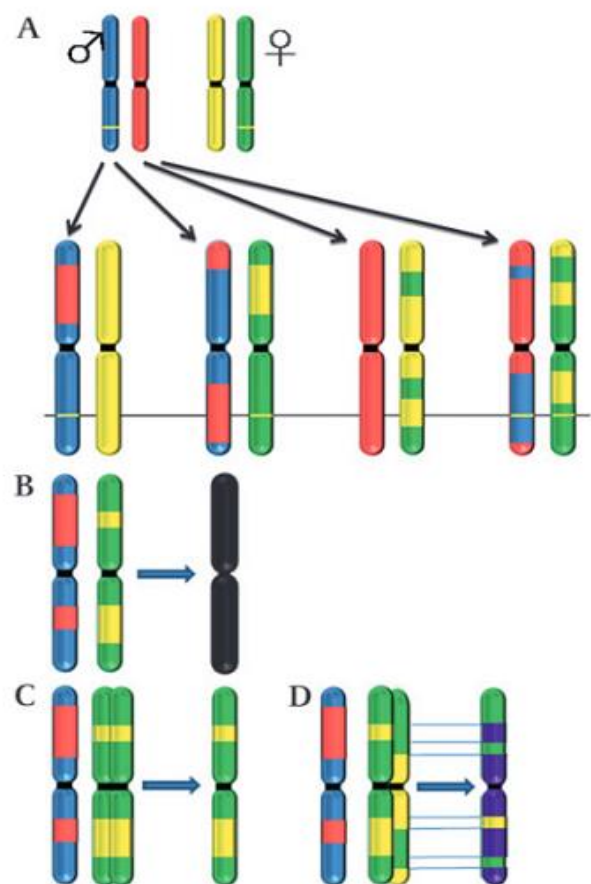
affected one to establish phase (give information which alleles are together on the same chromosome)), and the embryo, the chromosomal aneuploidy and single gene defect can be identified through karyomapping (figure 4). Only a healthy embryo, which has not inherited parental mutant allele that segregates in the affected child, will be implanted. In fact, this method has increased the success rate of PGD up to 69%<sup>13</sup>.

Karyomapping provides a wide range of information regarding the parental origin of haplotypes, the site of crossing over, chromosomal aneuploidy and paternity and maternity. Moreover, karyomapping can facilitate the locating of the possible linkage of families with singleton case (not certain whether it is *de novo* or inherited case). Karyomapping is intentionally designed to aid couples who had experienced fertility problems (in cases of being a carrier of balance translocation or Robertsonian translocation) or had family history of a particular inherited single gene disorder (ie, cystic fibrosis) and desire to have a normal baby. Conventional test prior to pre-implantation genetic diagnosis (PGD) such as FISH is laborious and time consuming compared to SNP array which can be automated and does not requiring cell culture, and with the advantage proposed by SNP array, it may increase the clinical pregnancy rate and decrease the miscarriage rate after preimplantation<sup>14</sup>.

### SNP array analysis in low-level mosaicism

Mosaicism refer to genetic abnormality that arises during mitosis, post fertilisation that resulting an individual possess both normal and

genetically abnormal cell line<sup>15</sup>. The effect of mosaicism to human would depend on the number of abnormal cells (the lower number of abnormal cells, the milder symptoms will be) and depend on the location of abnormal cells (brain, heart, skin, etc). Mosaicism identification can be labor intensive, since it requires a number of cells to be identified (at least 20 cells or even more) to find abnormal cells in order to confirm mosaicism diagnosis<sup>16</sup>.



**Figure 4. Example of karyomapping result in inherited disease and chromosomal aneuploidy<sup>13</sup>.**

(A). Inheritance pattern of single gene disorder. Parental homolog chromosome from father (coded red and blue color always at the left side) and from mother (coded yellow and green colour always at the right side), both of them harbour *CFTR* gene mutation (yellow line). By comparing the pattern of parental haplotype that flanking the gene in each child with those present in the affected child, carrier status of each child can be identified. The result from left to right

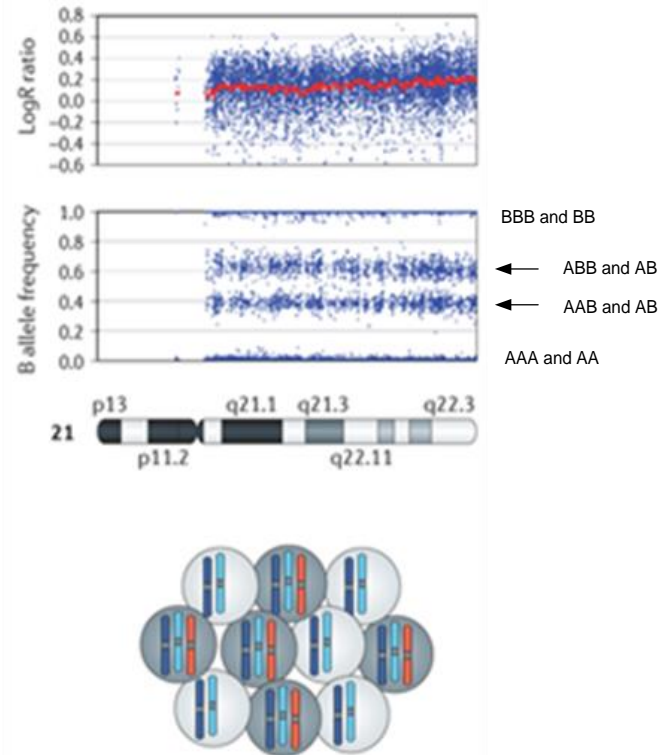
shows carrier from paternal allele, carrier from maternal allele, healthy (unaffected), affected (had both parental mutant allele). B, C, and D represent chromosomal aneuploidy; (B) the absence of maternal/paternal haplotype (black) indicated monosomy; (C). Trisomy was identified due to the presence of duplication of maternal homolog chromosome, but karyomapping cannot detect duplication of single chromosome because both alleles are identical; (D). Trisomy due to non-disjunction in meiosis I depicted since both maternal haplotype around the centromere region are present and nearby informative marker (purple), if non-disjunction in meiosis II would depicted both haplotype in more distal of chromosome.

SNP array with its high density of SNPs and probes to detect copy number variation, allows it to detect low level mosaicism which is missed / undetected by conventional cytogenetic analysis. By analysing the genotype calls pattern reflected on BAF, low-level mosaicism can be identified (figure 5). As reported by Leon et al (2010) that SNP array can detect low mosaicism of chromosome 21 (8-13% in lymphocytes) of Down syndrome patients<sup>17</sup>.

## CONCLUSION

In summary, SNP analysis gives great impact in molecular diagnosis by indirectly tracking the causative genes, which at the end will reveal the candidate gene to be tested. SNP array is more cost-effective than TaqMan technique for genotyping. SNP arrays result can help diagnosis UPD, low level mosaicism, and discover disease-causing gene in a family. However, a guideline for reporting incidental finding of parental blood relationship with SNP array is necessary to establish. Moreover, by its ability in detecting chromosomal abnormality and single gene defect in a single cell of embryo that employ linkage-based approach, make it a powerful tool to

increase the successful rate of *invitro* fertilisation which gives such a new hope to couples who had fertility problems and familial inherited disease to have normal baby.



**Figure 5. Mosaic trisomy 21 identified by SNP arrays.** LogR ratio slightly above zero (approximately 0.1 (red line)), in normal condition it should be zero. B allele frequency (BAF) is proportion between genotype calls (for example AAA and AA=0, BBB and BB=1, ABB and AB= 0.6 and AAB and AB=0.4); in normal condition, AB/heterozygous calls is in the midline (0.5). Thus from the result above, it indicates mosaicism. The smaller range of split between ABB and AB; and AAB and AB, the lower the mosaicism level represented. The normal cell line (disomy) depicted in white cell, abnormal cell line (trisomy) depicted in grey cell<sup>16</sup>.

## REFERENCES

1. LaFramboise T. Single nucleotide polymorphism assay: a decade of biological, computational and technological advances. *Nucleic Acids Res.* 2009; 37(13): 4181-93
2. Crawford DC, Nickerson DA. Definition and clinical importance of haplotypes. *Annu Rev Med.* 2005;56:302-20.
3. Alkuraya FS. Homozygosity mapping: one more tool in the clinical geneticist's toolbox. *Genet Med.* 2010; 12(4):236-39
4. Ragoussis J. Genotyping technologies for all. *Drug Discovery Today: Technologies.* 2006; 3(2):115-122
5. Nowak D, Hoffman WK, Koeffler HP. Genome wide mapping of copy number variation using SNP arrays. *Transfus Med and Hemother.* 2009; 36(4):246-51
6. Kearney HM, Kearney JB, Conlin KL. Diagnostic implication of excessive homozygosity detected by SNP-based microarray: consanguinity, uniparental disomy, and recessive single gene mutation. *Clin Lab Med.* 2011; 31(4):596-613
7. Schaaf CP, Wiszniewska J, Baudet AL. Copy number and SNP array in clinical diagnostics. *Ann Rev Genomics Hum Genet.* 2011b; 12:25-51.
8. Liehr T. Cytogenetic contribution to uniparental disomy (UPD). *Mol cytogenet.* 2010; 3:1-8
9. Kozd D. Prenatal testing for uniparental disomy: indications and clinical relevance. *Ultrasound Obstet and Gynaecol.* 2008; 31(1):100-5
10. Tucker T, Schlade-Bartusiak K, Eydoux P, Nelson T, Brown L. Uniparental disomy: can SNP array data be used for diagnosis?. *Genet Med.* 2012; 14(8); 753-6
11. Schaaf CP, Scott DA, Wiszniewska J, Baudet AL. Identification of incestuous parental relationship by SNP-based microarray. *Lancet.* 2011a; 337(9765); 555-6
12. Sund KL, Zimmerman SL, Thomas C, Mitchell AL, Prada CE, Grote L, Bao L, Martin LJ, Smolarek TA. Regions of homozygosity identified by SNP microarray analysis aid in the diagnosis of autosomal recessive disorder and incidentally detect parental blood relationship. *Genet Med.* 2013;15(1):70-78
13. Handyside AH, Harton GL, Mariani B, Thornhill AR, Affara N, Shaw MA, Griffin DK. Karyomapping: a universal method for genome wide analysis of genetic disease based on mapping crossover between parental haplotype. *J Med Genet.* 2010; 47(10); 651-8
14. Tan YQ, Tan K, Zhang SP, Gong F, Cheng DH, Xiong B, Lu CF, Tang XC, Luo KL, Lin G et al. Single – nucleotide polymorphism microarray-based preimplantation genetic diagnosis is likely to improve clinical outcome for translocation carrier. *Hum Reprod.* 2013; 28(9):2581-92
15. Iourov IY, Vorsanova SG, Yurov YB. Chromosomal mosaicism goes global. *Mol Cytogenet.* 2008; 1;1-26
16. Biasecker LG, Spinner NB. A genomic view of mosaicism and human disease. *Nature Rev Genet.* 2013; 14(5):307-20
17. Leon E, Zou YS, Milunsky JM. Mosaic Down syndrome in a patient with low-level mosaicism detected by microarray. *Am J Med Genet A.* 2010; 152A(12);3154-6