Molecular Human Reproduction, Vol.15, No.11 pp. 739–747, 2009
Advanced Access publication on August 11, 2009 doi:10.1093/molehr/gap066

MHR Basic science of reproductive medicine

ORIGINAL RESEARCH

Evaluation of genome coverage and fidelity of multiple displacement amplification from single cells by SNP array

Jiawei Ling[1,2], Guanglun Zhuang[2], Barbara Tazon-Vega[1], Chenhui Zhang[1], Baoqiang Cao[3], Zev Rosenwaks[1], and Kangpu Xu[1,4]

[1]Ronald O. Perelman and Claudia Cohen Center for Reproductive Medicine and Infertility, Weill Cornell Medical College of Cornell University, 1305 York Avenue, New York, NY 10021, USA [2]Center for Reproductive Medicine, First Affiliated Hospital, Sun Yat-sen University, 58 Zhongshan Road II, Guangzhou 510080, People's Republic of China [3]The Institute for Computational Engineering and Sciences, the University of Texas at Austin, Austin TX 78712, USA

[4]Correspondence address. Laboratory of Preimplantation Genetics, Ronald O. Perelman and Claudia Cohen Center for Reproductive Medicine and Infertility, Weill Cornell Medical College of Cornell University, Box 30, 1300 York Avenue, New York, NY 10065, USA. E-mail: kpxu@med.cornell.edu

ABSTRACT: The scarce amount of DNA contained in a singe cell is a limiting factor for clinical application of preimplantation genetic diagnosis mainly due to the risk of misdiagnosis caused by allele dropout and the difficulty in obtaining copy number variations in all 23 pairs of chromosomes. Multiple displacement amplification (MDA) has been reported to generate large quantity of products from small amount of templates. Here, we evaluated the fidelity of whole-genome amplification MDA from single or a few cells and determined the accuracy of chromosome copy number assessment on these MDA products using an Affymetrix 10K 2.0 SNP Mapping Array. An average coverage rate (86.2%) from single cells was obtained and the rates increased significantly when five or more cells were used as templates. Higher concordance for chromosome copy number from single cells could be achieved when the MDA amplified product was used as reference (93.1%) than when gDNA used as reference (82.8%). The present study indicates that satisfactory genome coverage can be obtained from single-cell MDA which may be used for studies where only a minute amount of genetic materials is available. Clinically, MDA coupled with SNP mapping array may provide a reliable and accurate method for chromosome copy number analysis and most likely for the detection of single-gene disorders as well.

Key words: chromosomal abnormality / genome coverage / multiple displacement amplification / PGD / SNP array

Introduction

Accurate analysis from a minute amount of DNA has been a challenge for geneticists. Several PCR-based protocols for whole-genome amplification (WGA) have been established and their applications evaluated, including primer extension preamplification (PEP) (Zhang *et al.*, 1992; Xu *et al.*, 1993) and degenerate oligonucleotide-primed PCR (DOP-PCR) (Telenius *et al.*, 1992; Wilton *et al.*, 2001). However, these protocols all have their limitations, including limited yield, strong biases and/or low genome coverage defined as gene representation (Cheung *et al.*, 1996; Paunio *et al.*, 1996; Kittler *et al.*, 2002; Coskun *et al.*, 2007). In the past few years, a non-PCR-based WGA called multiple displacement amplification (MDA) has been introduced (Dean *et al.*, 2002). Taking advantage of the phi29 DNA polymerase, an enzyme with higher processivity and better proofreading activity compared with *Taq* DNA polymerase, MDA shows some unique advantages over PCR-based WGA, including better fidelity as it has: an error rate of less than 3×10^{-6} (Nelson *et al.*, 2002), compared with 3×10^{-5} for *Taq* DNA polymerase; a higher average yield from a single cell varying from 1.7 to over 35 μg (Handyside *et al.*, 2004; Jiang *et al.*, 2005; Spits *et al.*, 2006); larger amplified DNA fragments (>10 kb); and more uniform representation of sequences (Paez *et al.*, 2004). For these reasons, MDA is now increasingly used in many studies including preimplantation genetic diagnosis (PGD) where molecular and cytogenetic diagnosis is based on the analysis of one cell only, in majority of cases, or a few cells from human embryos. Nevertheless, the scarce amount of DNA contained in a single cell as templates still poses a challenge for wide use of PGD

because of a phenomenon called 'allele dropout' or ADO, defined as the loss of heterozygosity due to amplification failure of one of the two alleles. ADO rates of MDA products from single cells are reported to be from 0 to 60% (Handyside *et al*., 2004; Hellani *et al*., 2004; Spits *et al*., 2006; Ren *et al*., 2007; Renwick *et al*., 2007) and appear to be affected by multiple factors when PCR is used to amplify single-cell DNA (Piyamongkol *et al*., 2003). However, these studies are based on the analysis of small number of markers (up to 64 polymorphic microsatellite markers) and on a limited number of chromosomes (seven different chromosomes). A comprehensive profile of MDA genome coverage is not available.

Although different types of chromosomal abnormalities in all 23 pairs of chromosomes have been successfully detected by array-based comparative genomic hybridization (aCGH) using MDA products from single cells (Hellani *et al*., 2004; Le Caignec *et al*., 2006), several critical issues need to be addressed before it can be safely used for clinical PGD. First, the extent of over- and/or under-representations in the final MDA products is unclear and the accuracy of aCGH from MDA products requires further evaluation (Paez *et al*., 2004; Le Caignec *et al*., 2006). Second, aCGH has yet to be tested for detecting copy number aberrations and single-gene disorders simultaneously. And finally, the current aCGH protocol is a time-consuming process that does not seem to fit easily into all clinical PGD schedule, particularly if specimens are required to be shipped to a reference laboratory, though a promising report on this has been published recently (Hellani *et al*., 2008). Looking for alternative platforms and potentials for the detection of both chromosome copy number abnormalities and single-gene disorders, the present study focuses on the analysis of fidelity of MDA in terms of whole-genome analysis of ADO and the genome coverage of MDA from single cells. This study, however, does not focus on the detection of any single-gene disorders though potentially feasible. Initially used for linkage analysis (Middleton *et al*., 2004), an SNP mapping array can not only provide a reliable method to evaluate DNA products on genotype, but is also a potential tool for the detection of chromosome copy number aberrations with high accuracy and reproducibility. Our objective is, therefore, to evaluate single-cell DNA amplified by MDA using SNP array, which may overcome above-mentioned limitations of aCGH. In this study, the Affymetrix 10 K 2.0 SNP mapping array was chosen because this platform contained over 10 000 SNPs and was reported to perform accurately for the analysis of small quantities of DNA (Paez *et al*., 2004; Tzvetkov *et al*., 2005; Corneveaux *et al*., 2007).

Materials and Methods

MDA sample preparation

Two cell lines, a trisomy 18 (Tri-18; GM02 732, 47, XY,+18) and a chromosome 4 segment deletion [sDel-4; GM00 343, 46,XY,4(del)(qter>p14)] from Coriell Cell Repositories/NIGMS (Camden, NJ), were chosen for the study. Cells were collected as described before (Spits *et al*., 2006) with minor modifications. Briefly, prior to cell collection, cell cultures were treated with 0.25% trypsin–EDTA and twice washed in PBS. Genomic DNA was extracted from cells using QIAamp DNA Blood Mini Kit (QIAGEN Sciences, MD). Different number of cells (1, 2, 5 and 10) were picked up under a dissection microscope by a glass pipette and placed into PCR tubes containing 5 μl alkaline lysis buffer (200 mM KOH, 50 mM dithiothreitol) and the tubes were then stored at −80°C before use. Aliquotes of 5 μl lysis buffer without cells were used as negative controls. Lysis was performed at 65°C for 10 min followed by addition of 5 μl neutralization buffer (400 mM Tricine). All necessary steps were taken to avoid contamination during the whole procedure.

MDA and experimental groups

Whole-genome amplification by phi29 polymerase-based MDA was performed using REPLI-g Midi Kit (QIAGEN Sciences, MD) with a final volume of 50 μl. MDA reactions were carried out at 30°C for 8 h and terminated by incubation at 65°C for 10 min. The products were purified using QIAamp DNA Micro Kit (QIAGEN Sciences, MD), quantified using ND-1000 spectrophotometer (NanoDrop Technologies, Denmark) and then stored at −20°C before next step. Considering the amount of specimens potentially available from clinical PGD, a total of seven experimental groups, using cells or extracted DNA from Tri-18 cell line, were established for genotype analysis. These included MDA products from: (1) a negative control with no templates (Neg); (2) a single cell (1C); (3) pooled from two separate single-cell MDA reactions (1 + 1C); (4) two cells (2C); (5) five cells (5C); (6) ten cells (10C); and (7) unamplified genomic DNA (gDNA). In a separate experiment, single cells were obtained from the sDel-4 cell line and MDA was performed in the same manner as in group 2 (1C-sDel-4). Three replicates were included in each group; except group 2 (1C) and group 3 (1 + 1C), both of which had six replicates, respectively. Detailed numbers of array used for each group are listed in Table I. For aneuploidy detection, a mean and standard deviation (SD) of all $\log_2$ signal intensity ratios of each array (10 M smoothing treatment by CNAT4.0) was calculated, and the array-specific threshold to define euploidy was determined as the mean plus or minus three times of the SD. The average $\log_2$ ratio of each chromosome on a certain array was compared with this threshold and ratios exceeding the threshold indicated aneuploidy or euploidy otherwise.

The MDA products were treated as previously described (Tzvetkov *et al*., 2005). Briefly, 250 ng of MDA products or unamplified genomic DNA was digested with XbaI, ligated to a universal adaptor and amplified using primers complementary to the adaptor sequence. After purification, the products, which were less than 1 kb in sizes, were digested to the sizes of about 50 bp with DNase I, end-labeled with biotin, and hybridized to the mapping array for 16 h at 48°C. The arrays were then washed and stained with R-Phycoerythrin Streptavidin using the Fluidics Station 450 and scanned using the GeneChip Scanner 3000 from Affymetrix.

Genotyping and copy number analysis

The 10 K 2.0 SNP mapping arrays, containing over 10 000 SNPs, and assay kits were purchased from Affymetrix along with 45 normal reference DNA data sets from blood samples (Affymetrix, Santa Clara, CA). The probes on this array were 25-mer each and the mean intermarker distance was 210 kb on 22 autosomes and the X chromosome. Data were analyzed using GeneChip® Genotyping Analysis Software (GCOS 1.2), GeneChip® Genotyping Analysis Software (GTYPE 4.1) and Copy Number Analysis Tool (CNAT 4.0) from Affymetrix. The GC content of each SNP segment was obtained from NetAffx™ Analysis Center of Affymetrix. A window of 250 000 bases to each side of that SNP was used and the fraction of bases that were G and C was calculated for each SNP in a chromosome. For coverage and fidelity analysis, the call rate on the array was defined as the SNP that could be correctly read and the call rate was used as the index of genome coverage in the study.

An algorithm called relative allele signal (RAS) was used for calculation of genotype from the signals measured (Kennedy *et al*., 2003). Briefly, the genotype of a locus is determined by the ratio of allele A signal to the sum of signals of A and B [A/(A + B)], and RAS should be 0, 0.5 and 1 for BB, AB and AA, respectively. However, artifacts such as ADO, PA

Table I An overview of each array on the expected aneuploidy identification

Group		1C						1 + 1C						2C						5C						10C			
Array no.		1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4
G	Mean $\log_2$	0.0006	0.0012	−0.002	−0.0017	−0.0002	−0.001	−0.0021	−0.0023	−0.0022	−0.0012	−0.0016	−0.0008	0.0004	−0.0025	−0.0023	0.0005	−0.0012	−0.0008	0.0002	0.0001	−0.0008	0.0012	0.0011	0.0013	−0.0003	−0.0003	−0.0002	−0.0002
	SD	0.21	0.18	0.18	0.17	0.21	0.19	0.19	0.19	0.19	0.17	0.17	0.19	0.16	0.19	0.19	0.15	0.16	0.18	0.13	0.10	0.10	0.11	0.11	0.11	0.16	0.15	0.15	0.16
	Averaged ratio of expected abnormal chromosome	0.59	0.34	0.25	0.27	0.46	0.23	0.24	0.23	0.23	0.33	0.23	0.32	0.37	0.18	0.26	0.32	0.28	0.26	0.45	0.41	0.39	0.42	0.43	0.44	0.29	0.30	0.30	0.30
M	Mean $\log_2$	0.0021	0.0028	−0.0007	−0.0004	0.0013	0.0004	−0.0008	−0.001	−0.0009	−0.0003	0.0004	0.0007	−0.0003	0.0019	−0.0012	0.0002	0.0001	−0.0008	0.0012	0.0011	0.0013							
	SD	0.18	0.15	0.11	0.13	0.12	0.12	0.10	0.10	0.10	0.11	0.10	0.13	0.11	0.12	0.10	0.13	0.10	0.10	0.11	0.11	0.11							
	Averaged ratio of expected abnormal chromosome	**0.75**	**0.51**	**0.38**	**0.48**	**0.61**	**0.39**	**0.37**	**0.36**	**0.35**	**0.47**	**0.36**	**0.46**	**0.52**	**0.38**	**0.39**	**0.45**	**0.41**	**0.39**	**0.42**	**0.43**	**0.44**							

G: gDNA data set as reference. M: single-cell MDA–DNA as reference.
Mean and SD: mean and standard deviation of $\log_2$ intensity ratio of all probes on the array (results from the Affymetrix CNAT4.0 with 10 M smoothing treatment). When single-cell MDA product was used as reference, the results from the abnormal chromosomes were excluded to avoid the fake trisomy. The threshold to define euploidy or aneuploidy was determined as the overall average ratio plus or minus three times of the standard deviation of the average $\log_2$ ratio. The average ratios of the expected abnormal chromosome (trisomy 18) were compared with the threshold. Values exceeding the threshold would indicate the trisomy and are shown in bold-italic. Otherwise regular values mean that the expected trisomy could not be identified from the results.

(preferential amplification, where one allele is over-amplified with respect to the other, but both are above the threshold of detection) or LOH (loss of heterozygosity) may occur. For MDA fidelity analysis, it is critical to distinguish ADO and PA from LOH. Here we developed an analysis method as followings. First, using the GCOS1.2 and GTYPE4.1, the SNP types (AA, AB or BB) on each array were obtained. For each replicate, those SNPs showing AA type in all unamplified samples (Group gDNA) and this replicate were picked up and the allele B signal intensity of these SNPs in the replicate could be then calculated. The 95th percentile of the allele B signal intensities was set as the threshold. The signal intensity of each LOH SNP (AB→AA) in this replicate was compared with the threshold. A locus was assigned as PA, or otherwise ADO, if the allele B intensity of a locus with LOH was above the threshold. With the same method we could differentiate ADO from PA of other LOH SNPs (AB→BB). The assignment of PA and ADO was done on an individual chip with the threshold for the chip and the false discovery rate (an SNP was true ADO but misdiagnosed as PA) was restricted to below 5%. We applied this method to all chips to assign PA and ADO of the locus to differentiate from LOH.

The copy number analysis tool from Affymetrix divides the results into five categories: 0, 1, 2, 3, 4 or more. To evaluate how well the copy number estimates were preserved after MDA from different number of cells, we defined the copy number zero and one as under-representation, two as correct representation and three or more as over-representation. For copy number analysis, two types of references were compared. First, we used combined data from 25 unamplified male samples provided by an Affymetrix database. Second, we used the data from the MDA products from single cells, either from Tri-18 or from sDel-4. When Tri-18 cells were used as reference, copy number 1 for chromosome 18 was regarded as correct representation. Likewise, a copy number 3 for the deleted segment of chromosome 4 was regarded as correct representation when sDel-4 cells were used as reference. Thus correct representation could be regarded as copy number concordance after MDA.

Results

Yield and genome coverage of MDA

Following the MDA, the yield of double-stranded DNA in each group was 19.3 ± 1.0, 19.1 ± 1.2, 19.6 ± 0.6 and 20.9 ± 1.4 μg, respectively from 1C, 2C, 5C and 10C Tri-18 fibroblasts; and MDA from negative controls generated a similar amount of products (19.4 ± 1.5 μg). No significant difference was found between any groups ($P > 0.05$). The genome coverage following whole-genome amplification was estimated based on the call rate of each sample. The SNP call rates from 1C, 1 + 1C and 2C groups showed no significant differences (86.2, 85.9 and 86.7%, respectively, $P > 0.05$). When the cell numbers increased to 5–10 cells, however, the call rate presented a significant difference (90.4 and 96.3%, respectively $P = 0.02$). An extremely low call rate (5.3 ± 1.2%) in the negative control group was observed compared with the experimental groups indicating that these products were non-specific and their impact on the SNP signals was likely insignificant. The increase in the mean call rate reflected random gain of genome coverage when cell numbers increased in the template as most uncalled SNPs were not consistent in different experimental groups (Fig. 1).

Compared with the mean number of uncalled SNPs in each group, ranging from 375 to 1420, the number of consistent uncalled SNPs was only about 10% (from 35 to 185) (Fig. 1a and b). To further investigate whether GC content influenced the occurrence of these uncalled

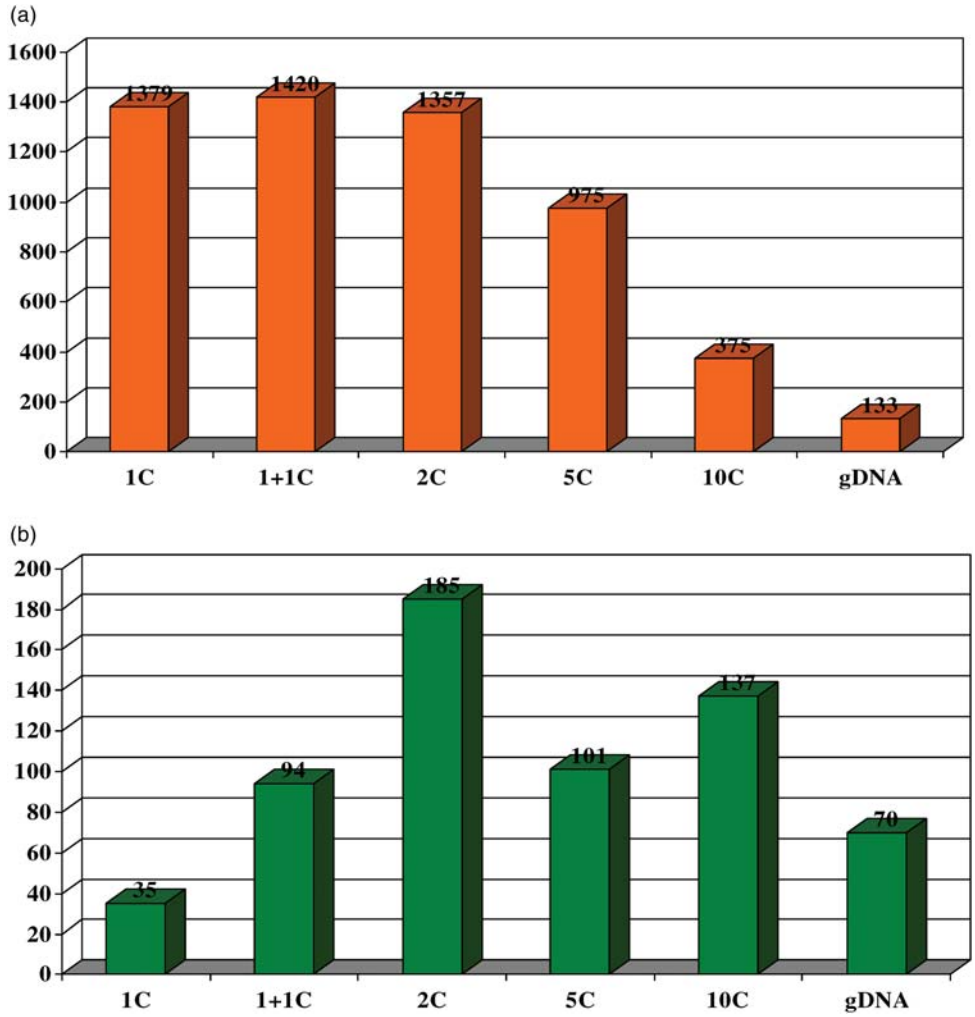

Figure 1 Distribution analysis of uncalled SNPs along the genome. The comparison between each multiple displacement amplification (MDA) experimental groups and the gDNA group showed that more uncalled SNPs existed when less number of cells was used as templates. The mean number of uncalled SNPs decreased when number of cells used as templates increased. Most of uncalled SNPs appeared to be not consistent, indicating a random loss of genome coverage. (**a**) Mean number of uncalled SNPs. The group 1C, 1 + 1C and 2C showed similar mean number of uncalled SNPs. A very small number of uncalled SNPs could be seen in the gDNA group. (**b**) Number of consistent uncalled SNPs. The total number of consistent uncalled SNPs was small in all groups, ranging from 35 to 185. Compared with that in group 1C (six replicates), the proportion of the number of consistent uncalled SNPs to mean number of uncalled SNPs were higher in group 2C, 5C and 10C (three replicates). The smaller sample size in these three groups might have contributed to this variation.

SNPs, we compared the GC content of the segments from which these consistent SNPs stood in each group to that of the overall SNPs on the array. Only those consistent uncalled SNPs in a single-cell group (group 1C) showed a significantly higher GC content (42.15% versus 39.86%, $P < 0.05$), but no correlation of GC content related to uncalled SNPs was seen in other groups ($P > 0.05$).

Fidelity of MDA for SNP analysis

Allele dropout was assessed in each group. Only those SNPs giving 'calls' in both unamplified and replicate pairs were included for further analysis. Three SNP genotypes among MDA amplified and unamplified groups were defined in this study: loss of heterozygosity (LOH, AB→AA or BB), gain of heterozygosity (GOH, AA or BB→AB) and change of homozygosity (COH, AA→BB or BB→AA). All these changes (LOH, GOH and COH) decreased as the cell number increased. The most dramatic difference was observed in the LOH type which decreased significantly with the increasing number of cells in the MDA reaction ($P < 0.05$, respectively), whereas GOH was observed in much smaller number of SNPs and only four SNPs were COH in four replicates (three in group 1C and one in group 2C) (Fig. 2a). Similar to LOH, the rates of ADO and PA decreased dramatically with an increasing number of starting cells (Fig. 2b). ADO rates varied greatly in Group 1C (19.46 ± 9.13%, with a range from 4.71 to 29.5%) and only a very small number of consistent ADO and PA SNPs existed (from 0 to 2) in each group. Further comparison between the GC content of the segments for which these consistent SNPs were located and those of the overall SNPs on the array showed no correlation ($P > 0.05$).

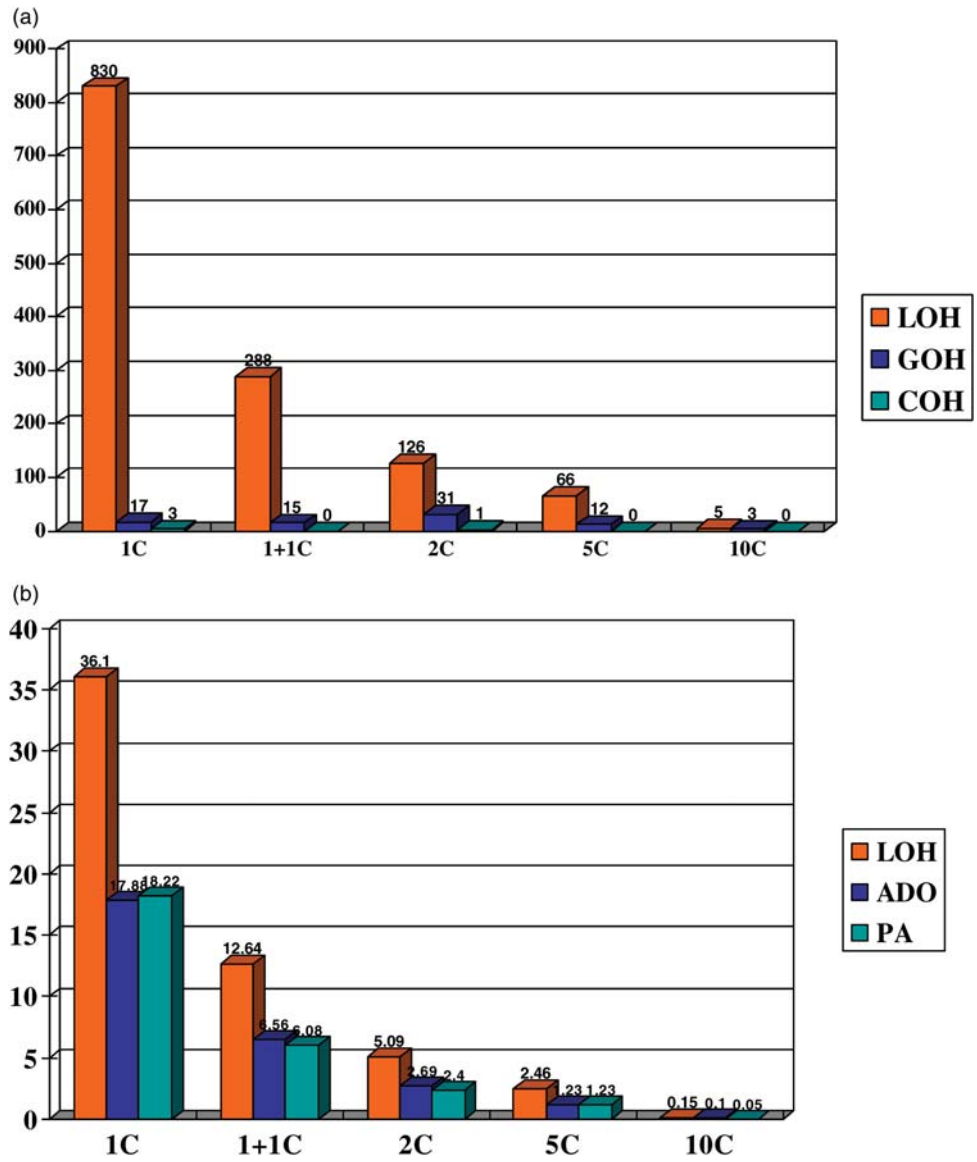

Figure 2 SNP genotype accuracy of MDA from different number of cells. (**a**) Impact of cell number on SNP genotype accuracy. Loss of heterozygosity (LOH), but not gain of heterozygosity (GOH) nor change of homozygosity (COH), appeared to be affected significantly by the number of cells used as templates. Only those SNPs giving call in both MDA groups and gDNA group were included in the analysis and the mean number of SNPs of each group was showed. LOH (AB→AA or BB); GOH (AA or BB→AB); COH (AA→BB or BB→AA). (**b**) Percentage of allele dropout (ADO) and preferential amplification (PA) of MDA products from different number of cells. Unlike the call rate, the ADO and PA rates decreased significantly even if the cell number increased from one to two cells or if two single-cell MDA products were pooled together ($P < 0.05$).

Detection of chromosomal abnormalities from different number of cells

A total of 24 arrays were used for the chromosome copy number detection (Table I). Concordance in chromosome copy number analysis was significantly higher when data from the single-cell MDA product was used as reference than that when the gDNA data were used as reference in group 1C and 1 + 1C ($P < 0.05$), but not in other groups ($P = 0.05$) (Fig. 3). Although both the trisomy 18 and chromosome 4 segment deletion could be detected by this 10K platform from a single cell, a few specific chromosomes showed a higher tendency of PA than others, including chromosome 16, 17, 19 and 22, when gDNA data were used as reference (Fig. 4a). PA disappeared, however, when data from the single-cell MDA product were used as reference (Fig. 4b).

Discussion

MDA genome coverage

Compared with other PCR-based WGA methods, MDA is reported to be a robust and less biased procedure for genome coverage (Nelson *et al*., 2002; Handyside *et al*., 2004; Paez *et al*., 2004; Jiang

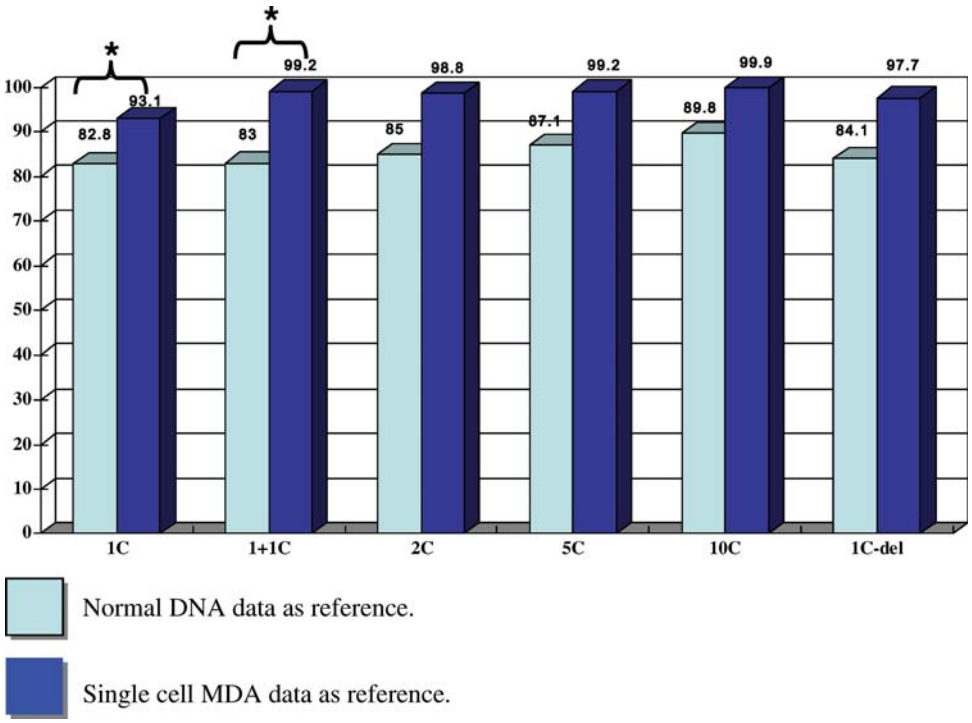

Figure 3 Comparison between SNP copy number concordance rates using two different references. The percentages are the average proportion of SNPs with concordant copy number of each group (see Materials and Methods). To make the data comparable between different groups, the SNPs locating on the chromosome 18 and chromosome 4 (between pter and p14) were excluded from the analysis. The copy number concordance was significantly higher in groups (*) 1C and 1 + 1C groups when the single-cell MDA product was used as reference than when the normal DNA was used as reference ($P < 0.05$), but no statistical difference could be observed in other groups ($P = 0.05$). 1C-del: mDA product from a single cell of chromosome 4 segment deletion cell line.

et al., 2005; Spits *et al.*, 2006). No accurate and comprehensive assessment of MDA products from single or a small number of cells is yet available on its fidelity, particularly, by high-resolution genetic analysis. Though a previous study has shown that MDA could generate a large amount of DNA product (Jiang *et al.*, 2005), the nature of the products has not been tested. Recently using SNP arrays a profile of MDA bias in single cells was reported (Iwamoto *et al.*, 2007), however, only one SNP for each chromosome was chosen to define the allele threshold of amplification in the study. Moreover, the knowledge of unspecific products of MDA on the genotyping accuracy (Handyside *et al.*, 2004) is still limited. The present study confirmed the previous finding that a large quantity of DNA, close to 20 μg, can be obtained from single cells after 8 h of MDA. The quantity of DNA did not increase as more cells (up to 10) were included as templates. When high-density SNP arrays were used to assess the performance of MDA from single cells and a small number of cells, we found that the call rate, as an index of genome coverage, was very low in the negative control group (5.3%), suggesting that these were indeed non-specific products and would have a limited impact on SNP signals. The call rate showed no differences between MDA amplification using 1C, 1 + 1C or 2C (86.2, 85.9 and 86.7%, respectively, $P > 0.05$). The genome coverage from a single cell in the present study is lower than that reported by previous studies, which ranged from 92 to 95% (Handyside *et al.*, 2004; Hellani *et al.*, 2004; Jiang *et al.*, 2005; Renwick *et al.*, 2007). The difference could be due to the fact that a limited number of markers (up to 64 SNPs on seven chromosomes) were involved in estimating the genome coverage in those studies whereas more than 10 000 SNPs were assessed in the present study. Compared with the call rate of the one-cell group, significant differences were observed when 5 (90.4%) or 10 (96.3%) cells were used ($P = 0.03$ and 0.02). With the benefit from the progression of blastocyst cultivation in recent years, more cells can be removed from an embryo (blastocyst biopsy). Our study showed that when 10 cells were used as the template, the call rate could be similar to that of unamplified DNA, indicating that a much more reliable result could be achieved by blastocyst biopsy.

In our study, only a small number of SNPs ($n = 35$) were not called consistently in the one-cell group (Fig. 1b), suggesting a random and not locus-specific failure of amplification after MDA reaction from single cells. Further analysis showed that the genomic regions where these consistent uncalled SNPs located had a higher GC content than the overall level through the array. This may be due to the higher tendency for these GC rich regions to form secondary structures, which in turn may hamper the DNA synthesis from the template during MDA. Although such a correlation could not be observed in other groups (group 2C, 5C and 10C), the small sample numbers in these groups ($n = 3$ for each) might contribute to the result. Further research with more samples is warranted.

SNP array for MDA fidelity

If MDA is applied to PGD or other critical genetic analysis, the extent of ADO must be evaluated. Handyside *et al.* (2004) reported an ADO

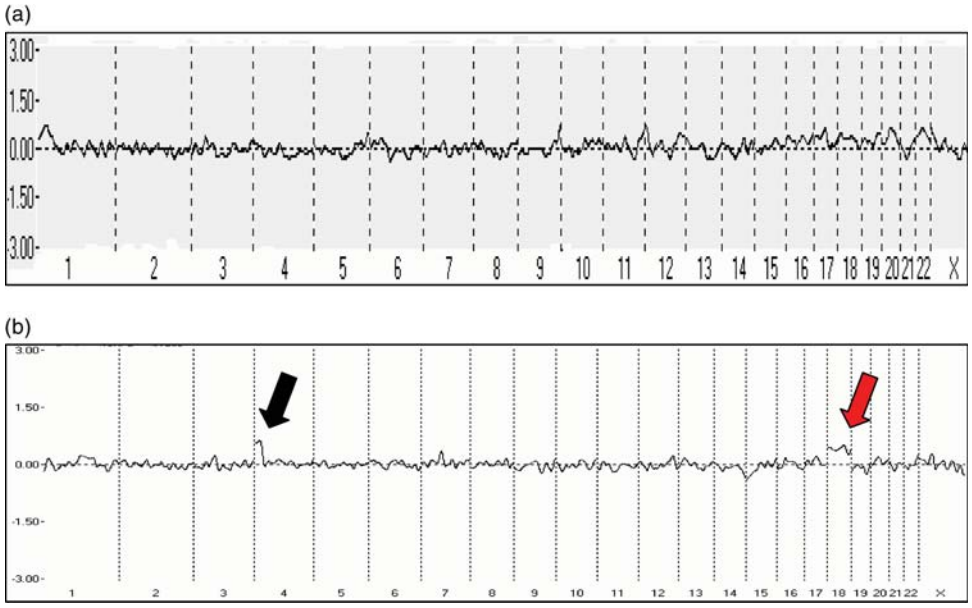

Figure 4 Comparison between copy number profiles using two different references. The single-cell MDA products of trisomy 18 and chromosome 4 segment deletion were used as reference to each other, so the copy number profile showed a 'fake' trisomy 4 segment (black arrow in b) in trisomy 18 samples. (**a**) When the gDNA data were used as reference, some chromosomes (16, 17, 19, 20 and 22) presented obvious preferential amplification and made it difficult to draw a conclusion. (**b**) Most preferential amplification disappeared and trisomy 18 could be diagnosed (red arrow in b) when data from a single-cell MDA product were used as reference.

rate of 31% on the analysis of 20 STRs with single-cell MDA. A similar study suggested the ADO rate to be 10.25% for the β-globin gene and 5% for a total 16 STRs (Hellani *et al*., 2004). The variable ADO rates observed may be due to the limited number of loci involved. A recent study reported an ADO rate of 38.9% for MDA in single cells using an SNP array (Iwamoto *et al*., 2007). To define whether or not the alleles A and B were amplified, 23 SNPs (one from each autosome and X chromosomes) were genotyped using a Taqman assay to set up the threshold. Again, only one SNP was chosen for each chromosome. The present study is the first in which ADO from single-cell MDA products is estimated throughout the whole genome and thus represents a more comprehensive assessment of ADO. In our study, a much higher rate of LOH than that of GOH and COH was observed in each group (Fig. 2a), suggesting that MDA rarely introduces new alleles. Nevertheless, the possibilities that some calls in the experimental groups may be due to the non-specific products cannot be ruled out since the negative group also showed 5.3% call signals.

The ADO rates varied greatly from 4.71 to 29.5% among group the one-cell group, which may be due to several reasons. Because only one copy of template was involved, the state of the DNA might have a great impact on the final result. For example, the ADO rate of replicate no. 3 in 1C group was much higher than those of other replicates in the same group but close to the results of 2C group. It is possible that the status of cell cycle may have contributed to the variation. For example, if a cell that had just finished the DNA replication but not yet divided was picked up so that actually two copies of DNA were present as templates. Furthermore, as fibroblast cells lines were cultured for number of passages, apoptosis or DNA fragmentation might have contributed to this variability. The overall ADO and PA rates decreased dramatically from 17.9 and 18.2% with a single cell to 0.1 and 0.05% with 10 cells (Fig. 2b), in agreement with a previous report (Handyside *et al*., 2004). It is important to emphasize that the present study obtained a more comprehensive profile for the impact of cell number on the genotyping compared with a previous study (Handyside *et al*., 2004) in which the fidelity of MDA from small numbers of cells was evaluated based on one locus as index only.

Interestingly, unlike the results of coverage analysis, in which the pooled DNA from two single-cell (1 + 1C) MDA reactions showed a similar performance as that from two cells (Fig. 1b), the LOH rate of the 2C group was much less than that of 1 + 1C (Fig. 2a). This might indicate that unspecific products were able to hybridize to the SNP array, but they were not able to have the same impact on signal capture or program analysis as the LOH rate is higher. In other words, slightly higher LOH from the 1C group may have limited impact on accuracy. The implication for clinical PGD is that biopsy of two cells may not provide a huge advantage. Separate MDA reactions should be performed when two single blastomeres are available, whereas pooled two cells would be recommended for MDA if LOH was to be minimized.

Finally, when the SNP array is used for the detection of the fidelity of any amplification methods, we should not regard LOH as ADO because some allelic PA will be easily mistaken as LOH by the RAS algorithm. As our method has successfully been used to distinguish the ADO from LOH in this 10K SNP mapping array, it can likely be extended to the analysis of WGA products using other SNP mapping array platforms.

SNP array for copy number analysis

As an alternative to FISH, array-CGH has been successfully used for chromosomal imbalances detection in whole-genome amplified DNA

products from single cells (Le Caignec *et al*., 2006; Fiegler *et al*., 2007). It has been known, and was demonstrated in the present study, that MDA will introduce background noise in the absence of template DNA. Whether these non-specifically amplified DNA will array randomly or specifically hybridize to the array remains unclear (Paez *et al*., 2004; Le Caignec *et al*., 2006). Controversy still exists regarding what sources of reference should be used. Paez *et al*. (2004) suggested that an amplified product using the same method should be used as reference because the errors caused by MDA seemed to be systematic. Le Caignec *et al*. (2006), however, held a different view based on their findings on a BAC array to detect the chromosomal imbalance from single cells. After comparing the copy number concordance between the results using these two different references, we found that in each group the mean copy number concordance was much higher when the data from a single-cell MDA product were used as reference (93.1% to 99.9%) than that from when normal gDNA was used as a reference (82.8–89.8%) (Fig. 3). A statistical difference was observed in the 1C and 1 + 1C groups ($P < 0.05$) but not in others ($P = 0.05$), most likely due to the small number of samples involved in the present study.

Two explanations could be provided for this difference. First, unlike what we found in the genotype analysis, most of the copy number variations caused by MDA seem to be not randomly but systematically distributed. From the copy number profiling (Fig. 4a and b), it appeared that the PA was consistent in some specific chromosomes (i.e. chromosome 16, 17, 19 and 22). This supports previous finding that these chromosomes had poorer representation compared with others (Paez *et al*., 2004). Although we could not find any correlation between GC content and PA on specific chromosomes as suggested by others (Teo *et al*., 2008), the different gene density may explain this phenomenon. We speculate that there are two types of chromatin locating inside the interphase nuclei: a highly condensed form called heterochromatin and a much less condensed form called euchromatin. Most DNA that is folded into heterochromatin does not contain genes (Alberts, 2002). Those chromosomes with higher gene density (Renwick *et al*., 2007), such as chromosome 16, 17, 19 and 22, are believed to have a higher proportion of euchromatin, whose loose structure would allow more efficient amplification which might produce some 'artifacts'. Amplified controls from a similar low amount of starting DNA may help to level up these errors and increase the concordance rate. Second, as other oligonucleotide-CGH platforms, an Affymetrix SNP mapping array cannot make a definite call for a loss or gain for a specific SNP, signals of a few adjacent oligonucleotides are required to be analyzed as a whole for a reliable call (Ylstra *et al*., 2006). For this purpose, the Affymetrix copy number analysis tool integrates the loci within a certain smoothing size as one by the hidden Markov model, and one identical copy number was given to all these loci in the final result to reduce random errors. All these may contribute to the high copy number concordance when MDA amplified products were used as reference. More rigorous comparison between the two types of references is required before it can be used in clinical applications.

Here, in combination with MDA and an Affymetrix 10K 2.0 SNP mapping array, we were able to successfully detect chromosomal abnormalities such as trisomy and monosomy. Though this relatively early version of an array can produce reproducible results, it has some obvious disadvantages, such as incomplete genome coverage (as it does not include Y chromosomes). SNP arrays with higher density such as a 500K array may overcome this limitation. It is also possible when an array with a much higher density of probes is used, the copy number variation (CNV) along the genome may further complicate the analysis. The optimized density of probes of array for PGD aneuploidy detection needs to be further studied and some analytic software must be improved.

Another disadvantage of this as a CGH system is the period required for MDA and hybridization, which may not be a problem for those facilities with an IVF laboratory and PGD laboratory, but could limit the clinical application for transport PGD. One strategy to overcome this problem is to cryopreserve the embryos after biopsy, and transfer the embryos in a subsequent cycle, as has been done with metaphase CGH (Wilton *et al*., 2003). The main drawback is that the embryo implantation potential would be reduced by the freezing and thawing procedure. An alternative approach is the investigation of the first polar body, which will provide sufficient time to perform transport PGD using aCGH system, but the presence of meiosis II or a paternally or post-zygotically derived error increases the misdiagnosis risk significantly. After all, further research on how to shorten the time required for MDA and hybridization may provide the best hope for resolving the problem.

In conclusion, the present study indicates that satisfactory genome coverage can be obtained from single-cell MDA. The assessment of the fidelity of MDA for single cells or a small number of cells and the comprehensive profile using the SNP mapping array revealed a reasonable accurate representation of MDA from a minute amount of DNA template. Furthermore, a high concordance of copy number estimation could be obtained when the MDA amplified product was used as reference. With a high-density SNP array, MDA products from single cells can accurately and reliably be used for chromosome copy number analysis and most likely for genotyping as well.

Acknowledgement

We are grateful to Dr. Jenny Z. Xiang, Director of Microarray Core Facility, Weill Cornell Medical College of Cornell University for processing the SNP array.

Funding

This study was partly supported by Guangdong Natural Science Fund (No. 06021276) and Guangdong Research Fund for Medical Science and Technology (No. B2006027).

References

Alberts B. *Molecular Biology of the Cell*. New York: Garland Science, 2002.

Cheung VG, Nelson SF. Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA. *Proc Natl Acad Sci USA* 1996;**93**:14676–14679.

Corneveaux JJ, Kruer MC, Hu-Lince D, Ramsey KE, Zismann VL, Stephan DA, Craig DW, Huentelman MJ. SNP-based chromosomal copy number ascertainment following multiple displacement whole-genome amplification. *Biotechniques* 2007;**42**:77–83.

Coskun S, Alsmadi O. Whole genome amplification from a single cell: a new era for preimplantation genetic diagnosis. *Prenat Diagn* 2007; **27**:297–302.

Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J *et al.* Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci USA* 2002;**99**:5261–5266.

Fiegler H, Geigl JB, Langer S, Rigler D, Porter K, Unger K, Carter NP, Speicher MR. High resolution array-CGH analysis of single cells. *Nucleic Acids Res* 2007;**35**:e15.

Handyside AH, Robinson MD, Simpson RJ, Omar MB, Shaw MA, Grudzinskas JG, Rutherford A. Isothermal whole genome amplification from single and small numbers of cells: a new era for preimplantation genetic diagnosis of inherited disease. *Mol Hum Reprod* 2004; **10**:767–772.

Hellani A, Coskun S, Benkhalifa M, Tbakhi A, Sakati N, Al-Odaib A, Ozand P. Multiple displacement amplification on single cell and possible PGD applications. *Mol Hum Reprod* 2004;**10**:847–852.

Hellani A, Abu-Amero K, Azouri J, El-Akoum S. Successful pregnancies after application of array-comparative genomic hybridization in PGS-aneuploidy screening. *Reprod Biomed Online* 2008;**17**:841–847.

Iwamoto K, Bundo M, Ueda J, Nakano Y, Ukai W, Hashimoto E, Saito T, Kato T. Detection of chromosomal structural alterations in single cells by SNP arrays: a systematic survey of amplification bias and optimized workflow. *PLoS ONE* 2007;**2**:e1306.

Jiang Z, Zhang X, Deka R, Jin L. Genome amplification of single sperm using multiple displacement amplification. *Nucleic Acids Res* 2005;**33**:e91.

Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, Liu G, Su X, Cao M, Chen W, Zhang J *et al.* Large-scale genotyping of complex DNA. *Nat Biotechnol* 2003;**21**:1233–1237.

Kittler R, Stoneking M, Kayser M. A whole genome amplification method to generate long fragments from low quantities of genomic DNA. *Anal Biochem* 2002;**300**:237–244.

Le Caignec C, Spits C, Sermon K, De Rycke M, Thienpont B, Debrock S, Staessen C, Moreau Y, Fryns JP, Van Steirteghem A *et al.* Single-cell chromosomal imbalances detection by array CGH. *Nucleic Acids Res* 2006;**34**:e68.

Middleton FA, Pato MT, Gentile KL, Morley CP, Zhao X, Eisener AF, Brown A, Petryshen TL, Kirby AN, Medeiros H *et al.* Genomewide linkage analysis of bipolar disorder by use of a high-density single-nucleotide-polymorphism (SNP) genotyping assay: a comparison with microsatellite marker assays and finding of significant linkage to chromosome 6q22. *Am J Hum Genet* 2004;**74**:886–897.

Nelson JR, Cai YC, Giesler TL, Farchaus JW, Sundaram ST, Ortiz-Rivera M, Hosta LP, Hewitt PL, Mamone JA, Palaniappan C *et al.* TempliPhi, phi29 DNA polymerase based rolling circle amplification of templates for DNA sequencing. *Biotechniques* 2002; **Suppl**:44–47.

Paez JG, Lin M, Beroukhim R, Lee JC, Zhao X, Richter DJ, Gabriel S, Herman P, Sasaki H, Altshuler D *et al.* Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Res* 2004; **32**:e71.

Paunio T, Reima I, Syvanen AC. Preimplantation diagnosis by whole-genome amplification, PCR amplification, and solid-phase minisequencing of blastomere DNA. *Clin Chem* 1996;**42**:1382–1390.

Piyamongkol W, Bermudez MG, Harper JC, Wells D. Detailed investigation of factors influencing amplification efficiency and allele drop-out in single cell PCR: implications for preimplantation genetic diagnosis. *Mol Hum Reprod* 2003;**9**:411–420.

Ren Z, Zhou C, Xu Y, Deng J, Zeng H, Zeng Y. Mutation and haplotype analysis for Duchenne muscular dystrophy by single cell multiple displacement amplification. *Mol Hum Reprod* 2007;**13**:431–436.

Renwick PJ, Lewis CM, Abbs S, Ogilvie CM. Determination of the genetic status of cleavage-stage human embryos by microsatellite marker analysis following multiple displacement amplification. *Prenat Diagn* 2007;**27**:206–215.

Spits C, Le Caignec C, De Rycke M, Van Haute L, Van Steirteghem A, Liebaers I, Sermon K. Optimization and evaluation of single-cell whole-genome multiple displacement amplification. *Hum Mutat* 2006;**27**:496–503.

Telenius H, Carter NP, Bebb CE, Nordenskjold M, Ponder BA, Tunnacliffe A. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 1992;**13**:718–725.

Teo YY, Inouye M, Small KS, Fry AE, Potter SC, Dunstan SJ, Seielstad M, Barroso I, Wareham NJ, Rockett KA *et al.* Whole genome-amplified DNA: insights and imputation. *Nat Methods* 2008;**5**:279–280.

Tzvetkov MV, Becker C, Kulle B, Nurnberg P, Brockmoller J, Wojnowski L. Genome-wide single-nucleotide polymorphism arrays demonstrate high fidelity of multiple displacement-based whole-genome amplification. *Electrophoresis* 2005;**26**:710–715.

Wilton L, Williamson R, McBain J, Edgar D, Voullaire L. Birth of a healthy infant after preimplantation confirmation of euploidy by comparative genomic hybridization. *N Engl J Med* 2001;**345**:1537–1541.

Wilton L, Voullaire L, Sargeant P, Williamson R, McBain J. Preimplantation aneuploidy screening using comparative genomic hybridization or fluorescence in situ hybridization of embryos from patients with recurrent implantation failure. *Fertil Steril* 2003;**80**:860–868.

Xu K, Tang Y, Grifo JA, Rosenwaks Z, Cohen J. Primer extension preamplification for detection of multiple genetic loci from single human blastomeres. *Hum Reprod* 1993;**8**:2206–2210.

Ylstra B, van den Ijssel P, Carvalho B, Brakenhoff RH, Meijer GA. BAC to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array CGH). *Nucleic Acids Res* 2006;**34**:445–450.

Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N. Whole genome amplification from a single cell: implications for genetic analysis. *Proc Natl Acad Sci U S A* 1992;**89**:5847–5851.

Submitted on April 3, 2009; resubmitted on July 30, 2009; accepted on August 5, 2009