Studying copy number variations using a nanofluidic platform

Jian Qin, Robert C. Jones and Ramesh Ramakrishnan*

Fluidigm Corporation, South San Francisco, CA 94080, USA

ABSTRACT

Copy number variations (CNVs) in the human genome are conventionally detected using high-throughput scanning technologies, such as comparative genomic hybridization and high-density single nucleotide polymorphism (SNP) microarrays, or relatively low-throughput techniques, such as quantitative polymerase chain reaction (PCR). All these approaches are limited in resolution and can at best distinguish a twofold (or 50%) difference in copy number. We have developed a new technology to study copy numbers using a platform known as the digital array, a nanofluidic biochip capable of accurately quantitating genes of interest in DNA samples. We have evaluated the digital array's performance using a model system, to show that this technology is exquisitely sensitive, capable of differentiating as little as a 15% difference in gene copy number (or between 6 and 7 copies of a target gene). We have also analyzed commercial DNA samples for their CYP2D6 copy numbers and confirmed that our results were consistent with those obtained independently using conventional techniques. In a screening experiment with breast cancer and normal DNA samples, the ERBB2 gene was found to be amplified in about 35% of breast cancer samples. The use of the digital array enables accurate measurement of gene copy numbers and is of significant value in CNV studies.

INTRODUCTION

Variation in the human genome occurs on multiple levels, from single nucleotide polymorphisms (SNPs) to duplications or deletions of contiguous blocks of DNA sequences (1–5). Copy number variation (CNV) is an important polymorphism of DNA segments across a wide range of sizes and one of the primary sources of variation in the human genome (6). Recently, CNV has been studied extensively because of its close association with large numbers of human disorders (7,8). An understanding of this variation is important not only to understand the full spectrum of human genetic variation but also to assess the significance of such variation in disease-association studies. The first human CNV map was constructed from a study of 270 normal individuals with a total of 1447 CNV regions in the whole genome (9); more than 15 000 CNVs have been found in the human genome (http://projects.tcag.ca/variation). A recent paper demonstrated the presence of 525 novel insertion sequences across the genomes of eight unrelated individuals, which were not present in the human reference genome, and showed that many of these have different copy numbers (10). However, the current CNV analysis is mainly dependent upon microarray-based SNP and comparative genomic hybridization (CGH) platforms, or DNA sequencing, and is therefore subject to low sensitivity and low resolution. These techniques are high throughput but lack the flexibility of analyzing individual genes or sequences of interest. Other existing technologies, such as quantitative polymerase chain reaction (PCR), are limited because of their inability to reliably distinguish less than a twofold difference in copy number of a particular gene in DNA samples (11–13).

In this study we demonstrate the use of a unique integrated nanofluidic system, the digital array, in the study of CNVs. The digital array (14,15) is able to accurately quantitate DNA samples based on the fact that single DNA molecules are randomly distributed in more than 9000 reaction chambers and then PCR amplified. The concentration of any sequence in a DNA sample (copies/µl) can be calculated using the numbers of positive chambers that contain at least one copy of that sequence. In order to ensure that the apparent difference in gene copy numbers in different samples are real, and not distorted by differences in sample amounts, we use the expression 'relative copy number'. The relative copy number of a gene is the number of copies of that gene per haploid genome. It can be easily expressed as the ratio of the copy number of a target gene to the copy number of a single copy reference gene (two copies per cell) in a DNA sample, which is always 1 per haploid genome. By using two assays for the two genes (the gene of interest and the reference gene)

*To whom correspondence should be addressed. Tel: +1 650 266 6084; Fax: +1 650 871 7152; Email: ramesh.ramakrishnan@fluidigm.com

with two fluorescent dyes on the same digital array, we are able to simultaneously quantitate both genes in the same DNA sample. The ratio of the numbers of molecules of these two genes is the relative copy number of the gene of interest in a DNA sample. A single copy gene should have a relative copy number of 1. A relative copy number greater than 1 indicates the presence of duplication of the target gene while a number smaller than 1 implies deletion of this gene.

Our data show that the digital array is able to distinguish less than twofold differences in gene copy number and differentiate between 1, 2, 3, 4, 5, 6 and 7 copies of a gene with great accuracy. It provides a reliable and robust platform to study copy number variations and has great advantages over conventional techniques.

MATERIALS AND METHODS

Construct, primer and probe sequences

The sequence of the RPP30 synthetic construct and the sequences of the primers and probe used to amplify this construct are shown in Supplementary Table 1, while the primers and probes for the CYP2D6 and ERBB2 genes are shown in Supplementary Table 2. All primers, probes and the synthetic construct were ordered from Biosearch Technologies (Novato, CA) and Integrated DNA Technologies (Coralville, IA, USA).

The TaqMan assay for the RNase P gene (VIC) was ordered from Applied Biosystems (Foster City, CA).

Digital array as a quantitating tool

The feasibility of digital PCR has previously been demonstrated by performing PCR on a single DNA sample obtained by a serial dilution process (16,17). Target molecules in a DNA sample could be quantitated by counting the number of positive reactions. We utilize the principle of partitioning instead of dilution in order to identify and quantitate individual DNA molecules.

The Fluidigm digital array is a novel nanofluidic biochip where digital PCR reactions can be performed (14,15). Utilizing nanoscale valves and pumps, the digital array delivers up to 12 mixtures of sample and PCR reagents into 12 individual panels. Each panel contains 765 independent 6-nl chambers. This nanofluidic platform utilizes soft lithography and silicone rubber to create nanoscale valves and pumps that can be used in serial or parallel applications. The digital array is composed of a PDMS (silicone rubber) Integrated Fluidic Circuit, an Integrated Heat Spreader to ensure rapid heat transfer and temperature uniformity within the array and an SBS-formatted carrier with inputs and pressure accumulator to act as an interface between the user and the PDMS chip. There are 12 carrier inputs corresponding to 12 separate sample inputs to the chip. Individual samples of a minimum volume of 8 μl each are delivered into 765 6-nl preprogrammed partitioning chambers in the chip by pressure-driven 'blind filling' in the PDMS. Control lines are primed with control fluid and are pressurized to actuate valves between the reaction chambers. The valves

partition individual chambers that are kept closed during the PCR experiment.

One of the important applications of the digital array is absolute quantitation (14,15). The DNA molecules in each mixture are randomly partitioned into the 765 chambers of each panel. The chip is then thermocycled on Fluidigm's BioMark system and the positive chambers that originally contained one or more molecules will generate fluorescent signals and can be counted by the Digital PCR Analysis software. Since the volumes and dilution factors of the DNA samples are known prior to loading into the digital array, the DNA concentrations can be accurately calculated. The precision of this test is only dependent upon the sampling randomness and, like any biological experiments, will improve with multiple tests (panels). Digital array has been routinely used by us to quantitate DNA samples of unknown concentration and, especially, cDNA samples whose concentrations of the sequences of interest are hard to determine otherwise.

Specific Target Amplification separates the linked copies of a target gene

When duplication occurs, multiple copies of a gene might be closely linked on the same chromosome and therefore might not be separated from each other, even on the digital array. As a result, multiple copies might behave as a single molecule and the total number of copies of the gene would be underestimated. When two copies are separated by a large genomic distance, some of them might be separated when DNA molecules are fragmented during purification. However, in most cases this would not be sufficient (see Table 2, sample NA11994 genomic DNA data). Specific target amplification (STA) is a good solution to this problem. STA is a simple PCR reaction with primers for both the reference gene and the gene of interest. It is typically performed for a limited number of thermal cycles (five in this study). The copy numbers of both genes are proportionally increased. Using this process, multiple copies of the gene of interest will be amplified separately and later randomly partitioned into chambers in the digital array. Since the newly generated molecules of both genes reflect the original ratio and they are not linked any more, a digital chip analysis can quantitate the molecules of the two genes and measure their ratio, and therefore the copy number of the gene of interest, very accurately (Figure 3). It is very important that the amplification efficiencies of the two pairs of primers be approximately equal in order not to introduce any bias in the ratio of the two gene copy numbers in the limited number of STA thermal cycles, although this is likely to have an insignificant effect on our results since we utilized only five cycles of preamplification. The amplification efficiency of any pair of primers can be easily measured using real-time PCR (18).

STA was performed on a GeneAmp PCR 9700 system (Applied Biosystems, Foster City, CA) in a 5 μl reaction containing 1 × TaqMan PreAmp master mix (Applied Biosystems, Foster City, CA), 225 nM of primers for both RNase P and the target gene and 10–50 ng DNA. Thermocycling conditions were 95°C, 10 min hot start and five cycles of 95°C for 15 s and 60°C for 2 min. The

products were diluted prior to the copy number analysis on the digital array based on their initial concentrations so that there would be about 500–600 RNase P molecules per panel.

Copy number analysis using the digital array on the BioMark system

Each panel of a digital array contains a total of 4.59 μl (6 nl × 765 chambers) PCR reaction mix. However, 10 μl reaction mixes were normally prepared for each panel, containing 1 × TaqMan gene expression master mix (Applied Biosystems, Foster City, CA), 1 × RNase P-VIC TaqMan assay, 1 × TaqMan assay (900 nM primers and 200 nM probe) for the target gene, 1 × sample loading reagent (Fluidigm, South San Francisco, CA) and DNA with about 1100–1300 copies of the RNase P gene. The reaction mix was uniformly partitioned into the 765 reaction chambers of each panel and the digital array was thermocycled on the BioMark system (http://www.fluidigm.com/products/biomark-main.html). Thermocycling conditions included a 95°C, 10 min hot start followed by 40 cycles of two-step PCR: 15 s at 95°C for denaturing and 1 min at 60°C for annealing and extension. Molecules of the two genes were independently amplified. FAM and VIC signals of all chambers were recorded at the end of each PCR cycle. After the reaction was completed, Digital PCR Analysis software (Fluidigm, South San Francisco, CA) was used to process the data and count the numbers of both FAM-positive chambers (target gene) and VIC-positive chambers (RNase P) in each panel.

Mathematical analysis of the digital array data

There are 765 chambers in each of the 12 panels in a digital array. When single DNA molecules are randomly partitioned into these chambers, it is possible that multiple molecules could partition into the same chamber. As a result there could be more molecules in each panel than positive chambers. The true number of molecules per chamber can be estimated using a simple Poisson distribution equation as described by Sindelka *et al.* (15). We have developed a more robust computational algorithm to analyze CNV data obtained from the digital array. This algorithm has been integrated into the Digital PCR Analysis software and is detailed in (19).

RESULTS

Establishment of a CNV model system

A proof-of-principle spike-in experiment was performed using a synthetic construct to explore the digital array's feasibility as a robust platform for the CNV study. A 65-base oligonucleotide that is identical to a fragment of the human RPP30 was ordered from Integrated DNA Technologies (Coralville, IA, USA). RNase P, a single copy gene, is used as reference in this study (20,21). Both RPP30 synthetic construct and human genomic DNA NA10860 from the Coriell Cell Repositories (Camden, NJ, USA) were quantitated using the RPP30

assay on a digital array. Different amounts of RPP30 synthetic construct were then spiked into the genomic DNA so that mixtures with ratios of RPP30 and RNase P of 1:1 (no spike-in), 1:1.5, 1:2, 1:2.5, 1:3 and 1:3.5 were made, simulating DNA samples containing two to seven copies of the RPP30 gene.

These spike-in mixtures were analyzed on the digital arrays. Five panels were used for each mixture and 400–500 RNase P molecules were present in each panel. The ratios of RPP30/RNase P of all samples were calculated and are plotted against the expected ratios in Figure 1. A good linear relationship can be observed. Also shown in Figure 2 is an example of a typical digital array experiment.

CNVs of the CYP2D6 gene

CYP2D6 belongs to the cytochrome P450 system responsible for the metabolism of many commonly prescribed medications (22,23). The CYP2D6 gene is highly polymorphic and this can significantly influence the metabolic activity of the enzyme it codes for (debrisoquine 4-hydroxylase) and the therapeutic efficacy of the drugs. Therefore, the pharmacogenetic polymorphism information of this gene would be of great clinical importance in therapeutic decision-making (24–27). More than 100 alleles of the CYP2D6 gene have been identified (http://www.cypalleles.ki.se/cyp2d6.htm). Allele-associated variations in the activity of the CYP2D6 enzyme have been observed and individuals carrying these alleles are classified into poor, intermediate, extensive and ultrarapid metabolizers (28,29).

Genotyping patients would be able to identify those who are at risk of severe toxic responses (poor metabolizer) or in need of more than standard level of drugs (ultra rapid metabolizer). It has been shown that some poor metabolizers and ultra rapid metabolizers are caused by the deletion or duplication of the entire CYP2D6 gene (30,31).

These large structural changes can be detected using conventional technologies such as Southern blot and long-range PCR. However, it is believed that real-time PCR is
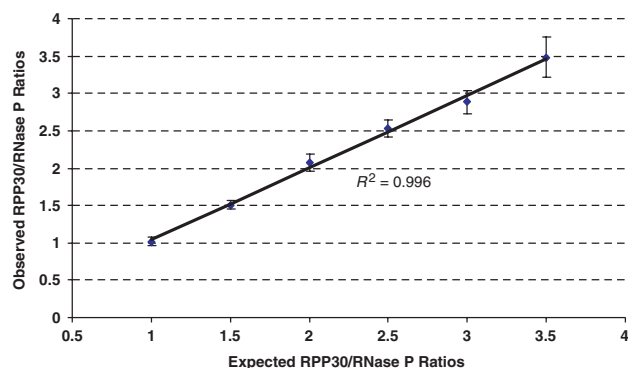

Figure 1. Quantitation of the RPP30 copy number in spike-in samples that contain two to seven copies of the RPP30 molecules per two haploid genomes. The x-axis shows the expected ratio of the numbers of RPP30 molecules to RNase P molecules. The y-axis shows the observed ratios. Each value is calculated using five panels of the same sample mix and the error bars represent standard errors. A good linear correlation can be seen with a coefficient of determination (R^2) of 0.996.
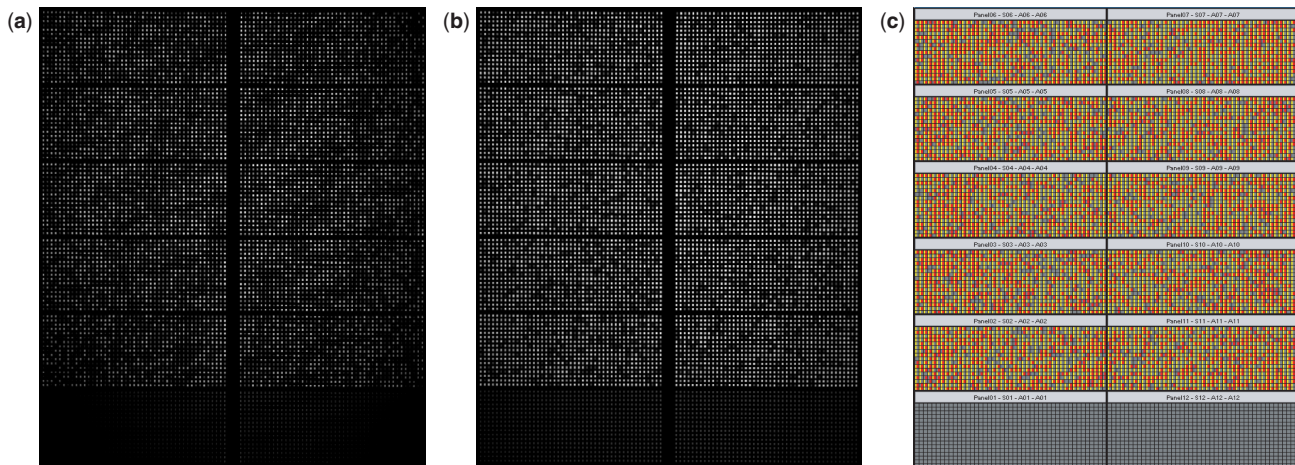
Figure 2. Five panels of each of the 6- and 7-copy mixtures were analyzed in this digital array for the RPP30 gene (FAM TaqMan assay) and the RNase P gene (VIC TaqMan assay). The RPP30/RNase P ratio of each panel was calculated using the numbers of molecules of the two genes in that panel. The two bottom panels were NTC (no template control). (**a**) and (**b**) the VIC (RNase P) and FAM (RPP30) images of the same digital array taken at the end of the PCR reaction, (**c**) the software-generated composite heat map showing the chambers with positive signals for either or both genes, each labeled with a different fluorescent dye (red for VIC and yellow for FAM). The Digital PCR Analysis software is able to count the number of positive chambers for each gene and calculate the RPP30 to RNase P ratio and its 95% CI (19).

Table 1. CYP2D6 genotypes and relative copy numbers of three DNA samples of known CYP2D6 genotypes

Genotype	CYP2D6/RNASE *P* (SE)
CYP2D6 *5/*5	0.00 (0.00)
CYP2D6 *1/*5	0.49 (0.03)
CYP2D6 *1A/*1A	0.98 (0.05)

*1 and *1A are both CYP2D6 wild-type alleles and *5 is the deletion of the entire gene. *1 is a new allele characterized and defined by ParagonDx (Morrisville, NC). Contact ParagonDx for more details.

currently the only promising technique that is able to provide information about the exact copy number of the CYP2D6 gene in a routine clinical setting (32–34).

We used the digital array to measure the CYP2D6 copy numbers of three DNA samples from ParagonDX (Morrisville, NC). The CYP2D6 genotypes of these DNA samples had been characterized (Table 1). The samples were STA-treated (see Figure 3 and Materials and methods section) and the products were analyzed using five panels each on the digital arrays. The relative copy numbers of these three samples are 0, 0.49 and 0.98, respectively, highly consistent with their assumed CYP2D6 diploid copy numbers (0, 1 and 2) based upon their genotypes.

We also studied five cell line DNA samples from Coriell Cell Repositories (Camden, NJ). First, we measured their relative copy numbers using genomic DNA. The results showed that two of them have a single copy and two have two copies of the CYP2D6 gene per cell (Table 2). One sample had a relative copy number of about 1.17, equal to a diploid copy number of 2.34. We then STA-treated these five samples and ran the products on digital arrays. The relative copy numbers of the 1- and 2-copy samples remained the same and the fifth sample showed a relative copy number of about 1.5 or a diploid copy number of 3. Apparently this sample had a duplication of the CYP2D6 gene on one of the two chromosomes (35). It has been
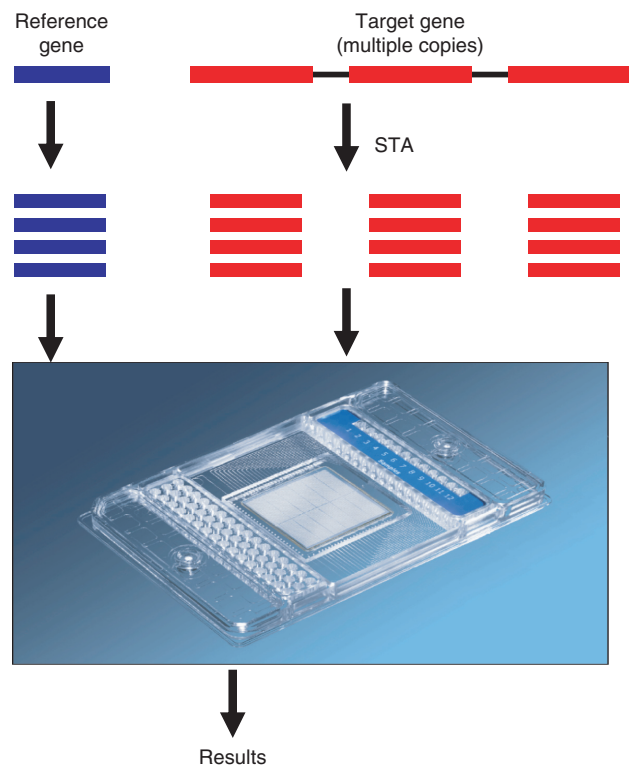

Figure 3. STA separates linked copies of the target gene on the same chromosome.

previously demonstrated (31,36) that when CYP2D6 duplication occurs, the two copies are separated by 12.1 kb. Therefore, the diploid copy number of 2.34 obtained when genomic DNA was used is likely the result of DNA breakage in this 12.1 kb genomic region in some DNA molecules that separated the two CYP2D6 copies. To confirm this, we ran a long range PCR [see (31) for primers and PCR conditions]. An extra band characteristic of

Table 2. CYP2D6 relative copy numbers of 5 Coriell DNA samples

	Genomic DNA (SE)	STA Product (SE)
NA12155	0.50 (0.02)	0.53 (0.01)
NA12873	0.54 (0.03)	0.48 (0.03)
NA07357	1.01 (0.05)	0.98 (0.10)
NA12872	0.95 (0.06)	0.97 (0.03)
NA11994	1.17 (0.06)	1.51 (0.07)

The results of both genomic DNA and STA products are shown. The ratios of the CYP2D6 gene to the RNase P gene should be close to multiples of 0.5. The genomic ratio of 1.17 for sample NA11994 (corresponding to a diploid copy number of 2.34) reflects the partial separation of the duplication alleles in the genomic DNA. A ratio of 1.51 (diploid copy number of 3) was obtained when the sample was subjected to STA prior to the digital PCR analysis.
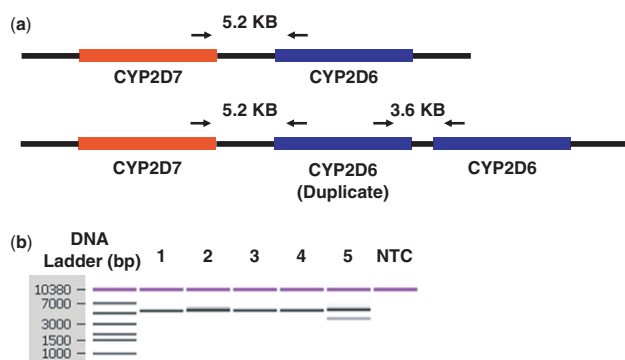

Figure 4. Long-range PCR assay confirms the digital array findings. A long-range PCR reaction was performed on the five Coriell DNA samples for the detection of the CYP2D6 duplication allele and the PCR products were analyzed using the Agilent 2100 Bioanalyzer (Santa Clara, CA). (**a**) The left PCR primer can anneal to both the CYP2D6 gene and CYP2D7, a pseudogene. A 5.2-kb fragment from the CYP2D7–CYP2D6 intergenic region should be obtained from every sample. An extra 3.6-kb PCR product can also be observed in individuals with CYP2D6 duplication. (**b**) Lanes 1–5: PCR products of samples NA12155, NA12873, NA07357, NA12872 and NA11994, respectively. NTC: no template control. The results show that NA11994 has a duplication of the CYP2D6 gene.

CYP2D6 duplication was observed only in the sample with a relative copy number of 1.5 (Figure 4).

CNVs of the ERBB2 gene in breast cancer and normal samples

ERBB2 (also known as HER2) is a receptor tyrosine kinase gene overexpressed in up to 30% of invasive breast cancer, resulting in a loss of normal cellular growth control. Most of these cases (97%) are caused by the amplification of this gene and the number of extra copies is closely related to the protein expression level (37–40).

ERBB2 amplification is well correlated with an aggressive phenotype characterized by reduced response to chemotherapy, high recurrence rate and short survival time and serves as a significant prognostic predictor for breast cancer patients (37,41). Trastuzumab (Herceptin), an FDA-approved monoclonal antibody against the ERBB2 protein, has been shown to dramatically increase response rate and extend survival in breast cancer patients with ERBB2 amplification. Given Trastuzumab's proven efficacy and substantial benefit in multiple clinical trials, detection of ERBB2 amplification has become critical (42–45).

There are different methodologies of determining the ERBB2 status in breast cancer. Immunohistochemistry (IHC) and fluorescence *in situ* hybridization (FISH) are two FDA-approved technologies for the detection of ERBB2 amplification. The former detects overexpression of the ERBB2 receptor on the cell membrane while the latter detects the copy number of the gene itself relative to the chromosome 17 centromere.

IHC is less expensive and easy to perform but is prone to a high rate of inaccuracies due to variations in tissue preparation, protein stability, antibody sensitivity and scoring subjectivity. On the other hand, FISH is accurate with good clinical correlation but it is expensive, time consuming, and labor intensive and requires very experienced personnel. Therefore, suggestions have been made to use a combination of IHC and FISH, where IHC is used as a screening procedure followed by a FISH confirmation if necessary (46,47).

We used digital arrays to analyze the ERBB2 copy numbers of 40 breast cancer and 8 normal breast tissue DNA samples from BioChain (Hayward, CA). All DNA samples were from Asian individuals except one normal sample that was from a Caucasian. Of the 40 breast cancer samples, 3 are adenocarcinoma, 1 is fibroadenoma, 2 are invasive lobular carcinoma, 1 is infiltrative ductal carcinoma and 33 are invasive ductal carcinoma. The samples were STA-treated and, for screening purpose, the products were analyzed using only two panels for each sample on digital arrays. The results are shown in Figure 5. Fourteen breast cancer samples (35%) had a diploid ERBB2 copy number of more than five while all control samples were below five copies [an absolute number of ERBB2 copies greater than 4.0 per cell is considered amplification in FISH analysis (47). Here we use five as the threshold]. The copy numbers shown are not all integers due to (i) heterogeneity of the cancer cells and (ii) sampling variations as only two panels were used for each sample.

A real-time PCR reaction was also performed on these 48 samples. Twenty-four replicates were used for each sample. Although the average copy numbers were close to the digital array data, large fluctuations (SDs of up to 0.5) were observed in the 24 reactions of each sample. Studies on other genes (for example, CYP2D6) showed that real-time PCR does not always produce accurate results (data not shown).

DISCUSSION

Genomewide analyses have shown the existence of large numbers of CNVs in the entire human genome with large interindividual diversity (48–53). Many of these CNVs colocalize with genes involved in a variety of diseases or disease susceptibility and are believed to play some role in pathogenesis (54–57). The first Mendelian disorder associated with the amplification of a 750 kb DNA fragment was reported recently (54). It appears to only be a
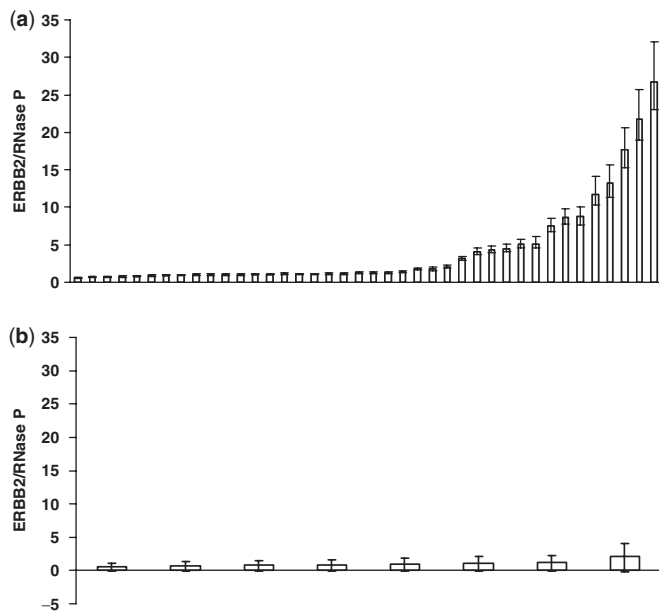
Figure 5. Relative copy numbers of the ERBB2 gene of (**a**) 40 breast cancer samples and (**b**) 8 normal breast samples. Error bars show the 95% CI (19).

question of time before more genetic conditions related to CNV are identified.

Two standard genomewide scanning methods for CNV detection are array-based CGH and high-density SNP genotyping arrays and both were employed in the construction of the first human CNV map (58). These microarray techniques are able to generate whole-genome CNV data and are important in CNV discovery. Their resolution is also improving with the development of new probes. However, since they are both based on hybridization, the detection of copy number changes largely depends on signal-to-noise ratio, which is sensitive to reagent and manufacturing variability. Therefore, false positive and false negative results are sometimes inevitable (59). Additionally, the lack of standard reference genomes in the studies using these technologies further complicates the interpretation of the results (60).

On many occasions, gene- or locus-specific (other than the whole genome) copy number information is required. This is especially true in the cases of CYP2D6 and ERBB2 described above in which therapeutic decision needs to be made based upon the copy numbers of these genes. In addition to other conventional methods (Southern blot, long-range PCR and FISH), the possibility of using quantitative PCR in the CNV study of these two genes has been previously explored (61–65).

Quantitative PCR is simple and easy to perform. However, since the copy number of the target gene is derived from the Ct difference between the target gene and a reference gene, the results are very sensitive to the efficiency of the amplification reaction. Even if one compensates for the amplification efficiency, it is considered difficult to obtain a discrimination power of better than twofold (66).

The digital array has the ability to absolutely quantitate any type of DNA sample. In a multiplex PCR reaction with two assays, the quantitation of two or more genes/sequences in a single sample becomes possible, effectively eliminating pipetting variations inherently occurring in any quantitation experiment. The accuracy of the results is only subject to the random distribution of the molecules and, like any biological experiments, can improve with the use of multiple replicates for each sample. STA can efficiently separate the linked copies of a gene on the same chromosome when duplication occurs while other methods, such as restriction digestion are also valid (data not shown).

We performed three experiments to test the feasibility of the digital array in the CNV study. First we measured the copy numbers of the RPP30 gene of a series of mixtures made of a human genomic DNA and a synthetic RPP30 construct. We observed a very good correlation between the results and the expected outcome. We then studied the CYP2D6 copy numbers of some DNA samples that were either genotyped elsewhere or characterized by us using conventional techniques. The results were also consistent. Lastly, we screened 40 breast cancer samples for the amplification of the ERBB2 gene. Although the clinical data (other than pathological classification) of these samples were lacking, about 35% of the samples had an increased number of this gene above 5, very close to the ERBB2 amplification frequency reported in the literature (67).

In conclusion, this study shows that the digital array provides a new and robust technology to study gene- and sequence-specific CNV and is able to detect gene copy numbers with great accuracy. Digital arrays provide a much greater discrimination power than quantitative PCR. CNV studies on the digital array are easy to perform, fast and the data obtained is easy to interpret. Furthermore, the platform is very flexible and can be tailored to any gene/sequence. It can also serve as an independent measure to verify results from the whole-genome scans using array technologies. The digital array is an excellent CNV platform for both basic research and clinical investigation.

REFERENCES

1. Brookes,A.J. (1999) The essence of SNPs. *Gene*, **234**, 177–186.
2. Eichler,E.E. (2001) Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet.*, **17**, 661–669.
3. Iafrate,A.J., Feuk,L., Rivera,M.N., Listewnik,M.L., Donahoe,P.K., Qi,Y., Scherer,S.W. and Lee,C. (2004) Detection of large-scale variation in the human genome. *Nat. Genet.*, **36**, 949–951.
4. Sebat,J., Lakshmi,B., Troge,J., Alexander,J., Young,J., Lundin,P., Månér,S., Massa,H., Walker,M., Chi,M. *et al.* (2004) Large-scale copy number polymorphism in the human genome. *Science*, **305**, 525–528.
5. Sharp,A.J., Locke,D.P., McGrath,S.D., Cheng,Z., Bailey,J.A., Vallente,R.U., Pertz,L.M., Clark,R.A., Schwartz,S., Segraves,R. *et al.* (2005) Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.*, **77**, 78–88.
6. Feuk,L., Carson,A.R. and Scherer,S.W. (2006) Structural variation in the human genome. *Nat. Rev. Genet.*, **7**, 85–97.
7. Ropers,H.H. (2007) New perspectives for the elucidation of genetic disorders. *Am. J. Hum. Genet.*, **81**, 199–207.
8. Lupski,J.R. (2007) Genomic rearrangements and sporadic disease. *Nat. Genet.*, **39**, S43–S47.
9. Redon,R., Ishikawa,S., Fitch,K.R., Feuk,L., Perry,G.H., Andrews,T.D., Fiegler,H., Shapero,M.H., Carson,A.R., Chen,W. *et al.* (2006) Global variation in copy number in the human genome. *Nature*, **444**, 444–454.
10. Kidd,J.M., Cooper,G.M., Donahue,W.F., Hayden,H.S., Sampas,N., Graves,T., Hansen,N., Teague,B., Alkan,C., Antonacci,F. *et al.* (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature*, **453**, 56–64.
11. Zimmermann,B., Holzgreve,W., Wenzel,F. and Hahn,S. (2002) Novel real-time quantitative PCR test for trisomy 21. *Clin. Chem.*, **48**, 362–363.
12. Lo,Y.M., Lun,F.M., Chan,K.C., Tsui,N.B., Chong,K.C., Lau,T.K., Leung,T.Y., Zee,B.C., Cantor,C.R. and Chiu,R.W. (2007) Digital PCR for the molecular detection of fetal chromosomal aneuploidy. *Proc. Natl Acad. Sci. USA*, **104**, 13116–13121.
13. Bubner,B., Gase,K. and Balwin,I.T. (2004) Two-fold differences are the detection limit for determining transgene copy numbers in plants by real-time PCR. *BMC Biotechnol.*, **4**, 1–11.
14. Spurgeon,S.L., Jones,R.C. and Ramakrishnan,R. (2008) High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. *PLoS ONE*, **3**, e1662.
15. Sindelka,R., Jonak,J., Hands,R., Bustin,S.A. and Kubista,M. (2008) Intracellular expression profiles measured by real-time PCR tomography in the *Xenopus laevis* oocyte. *Nucleic Acids Res.*, **36**, 387–392.
16. Kalinina,O., Lebedeva,I., Brown,J. and Silver,J. (1997) Nanoliter scale PCR with TaqMan detection. *Nucleic Acids Res.*, **25**, 1999–2004.
17. Vogelstein,B. and Kinzler,K.W. (1999) Digital PCR. *Proc. Natl Acad. Sci. USA*, **96**, 9236–9241.
18. Furtado,M.R., Petrauskene,O.V. and Livak,K.J. (2004) Application of real-time quantitative PCR in the analysis of gene expression. DNA Amplification: Current Technologies and Applications. In Demidov,V.V. and Broude,N.E. (eds), *Horizon Bioscience*. Norfolk, UK, Wymondham, pp. 131–145.
19. Dube,S., Qin,J. and Ramakrishnan,R. (2008) Mathematical analysis of copy number variation in a DNA sample using digital PCR on a nanofluidic device. *PLoSONE*, **3**, e2876, doi:10.1371/journal.pone.0002876.
20. Emery,S.L., Erdman,D.D., Bowen,M.D., Newton,B.R., Winchell,J.M., Meyer,R.F., Tong,S., Cook,B.T., Holloway,B.P., McCaustland,K.A. *et al.* (2004) Real-time reverse transcription-polymerase chain reaction assay for SARS-associated coronavirus. *Emerg. Infect. Dis.*, **10**, 311–316.
21. Baer,M., Nilsen,T.W., Costigan,C. and Altman,S. (1990) Structure and transcription of a human gene for H1 RNA, the RNA component of human RNase P. *Nucleic Acids Res.*, **18**, 97–103.
22. Lynch,T. and Price,A. (2007) The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. *Am. Fam. Physician*, **76**, 391–396.
23. Meyer,U.A. (1996) Overview of enzymes of drug metabolism. *J. Pharmacokinet. Biopharm.*, **24**, 449–459.
24. Tomalik-Scharte,D., Lazar,A., Fuhr,U. and Kirchheiner,J. (2008) The clinical role of genetic polymorphisms in drug-metabolizing enzymes. *Pharmacogenomics J.*, **8**, 4–15.
25. Daly,A.K. (2007) Individualized drug therapy. *Curr. Opin. Drug. Discov. Develop.*, **10**, 29–36.
26. Wijnen,P.A., Op den Buijsch,R.A., Drent,M., Kuipers,P.M., Neef,C., Bast,A., Bekers,O. and Koek,G.H. (2007) The prevalence and clinical relevance of cytochrome P450 polymorphisms. *Aliment. Pharmacol. Ther.*, **26 (**Suppl. 2), 211–219.
27. Meyer,U.A. (2000) Pharmacogenetics and adverse drug reactions. *Lancet*, **356**, 1667–1671.
28. Beverage,J.N., Sissung,T.M., Sion,A.M., Danesi,R. and Figg,W.D. (2007) CYP2D6 polymorphisms and the impact on tamoxifen therapy. *J. Pharm. Sci.*, **96**, 2224–2231.
29. Dorado,P., Berecz,R., Peñas-Lledó,E.M., Cáceres,M.C. and Llerena,A. (2006) Clinical implications of CYP2D6 genetic polymorphism during treatment with antipsychotic drugs. *Curr. Drug Targets*, **7**, 1671–1680.
30. Gaedigk,A., Blum,M., Gaedigk,R., Eichelbaum,M. and Meyer,U.A. (1991) Deletion of the entire cytochrome P450 CYP2D6 gene as a cause of impaired drug metabolism in poor metabolizers of the debrisoquine/sparteine polymorphism. *Am. J. Hum. Genet.*, **48**, 943–950.
31. Løvlie,R., Daly,A.K., Molven,A., Idle,J.R. and Steen,V.M. (1996) Ultrarapid metabolizers of debrisoquine: characterization and PCR-based detection of alleles with duplication of the CYP2D6 gene. *FEBS Lett.*, **392**, 30–34.
32. Schaeffeler,E., Schwab,M., Eichelbaum,M. and Zanger,U.M. (2003) CYP2D6 genotyping strategy based on gene copy number determination by TaqMan real-time PCR. *Hum. Mutat.*, **22**, 476–485.
33. Bodin,L., Beaune,P.H. and Loriot,M.A. (2005) Determination of cytochrome P450 2D6 (CYP2D6) gene copy number by real-time quantitative PCR. *J. Biomed. Biotechnol.*, **2005**, 248–53.
34. Meijerman,I., Sanderson,L.M., Smits,P.H., Beijnen,J.H. and Schellens,J.H. (2007) Pharmacogenetic screening of the gene deletion and duplications of CYP2D6. *Drug Metab. Rev.*, **39**, 45–60.
35. Kimura,S., Umeno,M., Skoda,R.C., Meyer,U.A. and Gonzalez,F.J. (1989) The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene. *Am. J. Hum. Genet.*, **45**, 889–904.
36. Steijns,L.S. and Van Der Weide,J. (1998) Ultrarapid drug metabolism: PCR-based detection of CYP2D6 gene duplication. *Clin. Chem.*, **44**, 914–917.
37. Slamon,D.J., Clark,G.M., Wong,S.G., Levin,W.J., Ullrich,A. and McGuire,W.L. (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science*, **235**, 177–182.
38. Slamon,D.J., Godolphin,W., Jones,L.A., Holt,J.A., Wong,S.G., Keith,D.E., Levin,W.J., Stuart,S.G., Udove,J., Ullrich,A. *et al.* (1989) Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science*, **244**, 707–712.
39. Pauletti,G., Godolphin,W., Press,M.F. and Slamon,D.J. (1996) Detection and quantitation of HER-2/neu gene amplification in human breast cancer archival material using fluorescence in situ hybridization. *Oncogene*, **13**, 63–72.
40. Masood,S. and Bui,M.M. (2002) Prognostic and predictive value of HER2/neu oncogene in breast cancer. *Microsc. Res. Tech.*, **59**, 102–108.
41. Révillion,F., Bonneterre,J. and Peyrat,J.P. (1998) ERBB2 oncogene in human breast cancer and its clinical significance. *Eur. J. Cancer*, **34**, 791–808.
42. Slamon,D.J., Leyland-Jones,B., Shak,S., Fuchs,H., Paton,V., Bajamonde,A., Fleming,T., Eiermann,W., Wolter,J., Pegram,M. *et al.* (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.*, **344**, 783–792.
43. Tan,A.R. and Swain,S.M. (2003) Ongoing adjuvant trials with trastuzumab in breast cancer. *Semin. Oncol.*, **30 (**Suppl. 16), 54–64.
44. Piccart-Gebhart,M.J., Procter,M., Leyland-Jones,B., Goldhirsch,A., Untch,M., Smith,I., Gianni,L., Baselga,J., Bell,R., Jackisch,C. *et al.* (2005) Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N. Engl. J. Med.*, **353**, 1659–1672.

45. Smith,I., Procter,M., Gelber,R.D., Guillaume,S., Feyereislova,A., Dowsett,M., Goldhirsch,A., Untch,M., Mariani,G., Baselga,J. *et al.* (2007) 2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial. *Lancet*, **369**, 29–36.

46. Masood,S. and Bui,M.M. (2002) Prognostic and predictive value of HER2/neu oncogene in breast cancer. *Microsc. Res. Tech.*, **59**, 102–108.

47. Laudadio,J., Quigley,D.I., Tubbs,R. and Wolff,D.J. (2007) HER2 testing: a review of detection methodologies and their clinical performance. *Expert Rev. Mol. Diagn.*, **7**, 53–64.

48. Eichler,E.E. (2001) Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet.*, **17**, 661–669.

49. Iafrate,A.J., Feuk,L., Rivera,M.N., Listewnik,M.L., Donahoe,P.K., Qi,Y., Scherer,S.W. and Lee,C. (2004) Detection of large-scale variation in the human genome. *Nat. Genet.*, **36**, 949–951.

50. Sebat,J., Lakshmi,B., Troge,J., Alexander,J., Young,J., Lundin,P., Månér,S., Massa,H., Walker,M., Chi,M. *et al.* (2004) Large-scale copy number polymorphism in the human genome. *Science*, **305**, 525–528.

51. Sharp,A.J., Locke,D.P., McGrath,S.D., Cheng,Z., Bailey,J.A., Vallente,R.U., Pertz,L.M., Clark,R.A., Schwartz,S., Segraves,R. *et al.* (2005) Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.*, **77**, 78–88 .

52. Feuk,L., Carson,A.R. and Scherer,S.W. (2006) Structural variation in the human genome. *Nat. Rev. Genet.*, **7**, 85–97.

53. Scherer,S.W., Lee,C., Birney,E., Altshuler,D.M., Eichler,E.E., Carter,N.P., Hurles,M.E. and Feuk,L. (2007) Challenges and standards in integrating surveys of structural variation. *Nat. Genet.*, **39**, S7–S15.

54. Balikova,I., Martens,K., Melotte,C., Amyere,M., Van Vooren,S., Moreau,Y., Vetrie,D., Fiegler,H., Carter,N.P., Liehr,T. *et al.* (2008) Autosomal-dominant microtia linked to five tandem copies of a copy-number-variable region at chromosome 4p16. *Am. J. Hum. Genet.*, **82**, 181–187.

55. Wong,K.K., deLeeuw,R.J., Dosanjh,N.S., Kimm,L.R., Cheng,Z., Horsman,D.E., MacAulay,C., Ng,R.T., Brown,C.J., Eichler,E.E. *et al.* (2007) A comprehensive analysis of common copy-number variations in the human genome. *Am. J. Hum. Genet.*, **80**, 91–104.

56. Ropers,H.H. (2007) New perspectives for the elucidation of genetic disorders. *Am. J. Hum. Genet.*, **81**, 199–207.

57. Lupski,J.R. (2007) Genomic rearrangements and sporadic disease. *Nat. Genet.*, **39**, S43–47.

58. Redon,R., Ishikawa,S., Fitch,K.R., Feuk,L., Perry,G.H., Andrews,T.D., Fiegler,H., Shapero,M.H., Carson,A.R., Chen,W. *et al.* (2006) Global variation in copy number in the human genome. *Nature*, **444**, 444–454.

59. Carter,N.P. (2007) Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat. Genet.*, **39**, S16–S21.

60. Scherer,S.W., Lee,C., Birney,E., Altshuler,D.M., Eichler,E.E., Carter,N.P., Hurles,M.E. and Feuk,L. (2007) Challenges and standards in integrating surveys of structural variation. *Nat. Genet.*, **39** (Suppl. 7), S7–S15.

61. Schaeffeler,E., Schwab,M., Eichelbaum,M. and Zanger,U.M. (2003) CYP2D6 genotyping strategy based on gene copy number determination by TaqMan real-time PCR. *Hum. Mutat.*, **22**, 476–485.

62. Bodin,L., Beaune,P.H. and Loriot,M.A. (2005) Determination of cytochrome P450 2D6 (CYP2D6) gene copy number by real-time quantitative PCR. *J. Biomed. Biotechnol.*, **2005**, 248–253.

63. Meijerman,I., Sanderson,L.M., Smits,P.H., Beijnen,J.H. and Schellens,J.H. (2007) Pharmacogenetic screening of the gene deletion and duplications of CYP2D6. *Drug Metab. Rev.*, **39**, 45–60.

64. Königshoff,M., Wilhelm,J., Bohle,R.M., Pingoud,A. and Hahn,M. (2003) HER-2/neu gene copy number quantified by real-time PCR: comparison of gene amplification, heterozygosity, and immunohistochemical status in breast cancer tissue. *Clin. Chem.*, **49**, 219–229.

65. Lamy,P.J., Nanni,I., Fina,F., Bibeau,F., Romain,S., Dussert,C., Penault Llorca,F., Grenier,J., Ouafik,L.H., Martin,P.M. *et al.* (2006) Reliability and discriminant validity of HER2 gene quantification and chromosome 17 aneusomy analysis by real-time PCR in primary breast cancer. *Int. J. Biol. Markers*, **21**, 20–29.

66. Lo,Y.M., Lun,F.M., Chan,K.C., Tsui,N.B., Chong,K.C., Lau,T.K., Leung,T.Y., Zee,B.C., Cantor,C.R. and Chiu,R.W. (2007) Digital PCR for the molecular detection of fetal chromosomal aneuploidy. *Proc. Natl Acad. Sci. USA*, **104**, 13116–13121.

67. Slamon,D.J., Clark,G.M., Wong,S.G., Levin,W.J., Ullrich,A. and McGuire,W.L. (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177–182.