

# SNP-based chromosomal copy number ascertainment following multiple displacement whole-genome amplification

Jason J. Corneveaux, Michael C. Kruer, Diane Hu-Lince, Keri E. Ramsey, Victoria L. Zismann, Dietrich A. Stephan, David W. Craig, and Matthew J. Huentelman

*The Translational Genomics Research Institute (TGen), Phoenix, AZ, USA*

*BioTechniques* 42:77-83 (January 2007)  
doi 10.2144/000112308

*Whole genome amplification by multiple displacement amplification (MDA) offers investigators using precious genomic DNA samples a high fidelity method for amplifying nanogram quantities of DNA several thousandfold. This becomes especially important for the modern day genomics researcher who more and more commonly is applying today's genome scanning technologies to patient cohort samples collected years ago that are irrecoverable and invariably in short supply. We present evidence here that MDA-prepared genomic DNA includes artifacts of chromosomal copy number that resemble copy number polymorphisms (CNPs) upon analysis of the DNA on the Affymetrix 10K GeneChip®. The study of CNPs in both health and disease is a rapidly growing area of research, however our current understanding of the relevance of CNPs is incomplete. Our data indicate that utilization of whole genome-amplified samples for analysis heavily reliant on accurate copy number retention could be confounded if the genomic DNA sample was subjected to MDA. We recommend that small amounts of patient cohort DNA stocks be set aside and not subjected to whole genome amplification in order to facilitate the unbiased determination of chromosomal copy numbers when desired.*

## INTRODUCTION

Retaining sufficient amounts of high-quality genomic DNA is often a challenge for researchers performing genetic-based studies. For each recruited patient cohort, larger quantities of genomic DNA are increasingly required for testing due to the rapidly expanding array of high-throughput analyses at the disposal of the modern day genomics researcher. Additionally, genomic DNA samples are often isolated from archived tissue or blood/buccal cell spot collection cards, two sources that primarily yield relatively low amounts of high-quality genomic DNA for further analysis (1,2). Traditional isolation of DNA from freshly purified lymphocytes yields greater amounts of DNA, yet this relatively invasive procedure is not foolproof, sometimes leading to the isolation of lowered amounts of useable DNA, and in most studies is only able to be performed once without the painstaking re-recruitment of the original patient cohort for an additional blood

draw that can lead to further experimental design issues (i.e., not all patients are able to be present for a second blood draw). These issues of DNA quality and quantity also confront researchers involved in prenatal, forensic, and anthropology genetic study.

To overcome these issues, several methods for the nonspecific amplification of the entire genome, termed whole genome amplification (WGA), from nanogram quantities of genomic DNA have been developed. PCR-based amplification techniques include degenerate oligonucleotide primed PCR (DOP-PCR) or variations thereof [long DOP-PCR and long, low (LL)-DOP-PCR], primer extension preamplification (PEP), tagged PCR, and balanced PCR amplification. These methodologies all suffer from some sort of limitation, including generation of relatively short fragments (DOP-PCR and PEP), low yield (PEP and tagged PCR), increased technical complexity (tagged PCR), and high levels of input DNA (long DOP-PCR), and there are several publications

in the literature detailing modification of these protocols to avoid such limitations (3–7). However, a non-PCR-based method termed multiple displacement amplification (MDA) that uses a  $\phi$ 29 polymerase-mediated reaction carried out in an overnight isothermal reaction, was recently developed (8–10). The  $\phi$ 29 polymerase, isolated from bacteriophage  $\phi$ 29, is a desirable enzyme for WGA due to its high processivity at 30°C, strand displacement capacity, 3'→5' exonuclease activity, and perhaps most importantly, its protein-primed method of replication, which results in the synthesis of DNA fragments >70 kb and error rates as low as 1 mistake in a million (11,12). Studies have shown that MDA produces excellent genotype call concordance between amplified and nonamplified templates (reported error rates of 1 mistake per  $9.5 \times 10^{-6}$  bp and a 99.95% concordance between MD-WGA and non-MD-WGA genotypes on the Affymetrix 10K GeneChip®), and it has quickly become the preferential WGA technique (8,13–17).

As the complexity of the human genome and the variations within are further elucidated, it is becoming increasingly evident that small chromosomal copy number alterations or polymorphisms (CNPs) are present across the genome in healthy individuals, however they have been proposed to play a role in disease (18–22). CNPs of various sizes have been identified, however they are commonly <10 kb in size and generally >200 bp (18,19,23). Some investigators have shown that CNPs appear to be in linkage disequilibrium with neighboring single nucleotide polymorphisms (SNPs) in the genome, thereby suggesting that they both share a similar evolutionary history and could also be assayed by proxy through the genotyping of the linked SNP(s) (19). The mechanism of CNP creation includes retrotransposition among other theories that are currently under investigation (20).

Diseases like Down syndrome, trisomy 21, and Prader-Willi syndrome (PWS), a deletion of the q-arm of chromosome 15, are well-documented large chromosomal aberrations that have been suggested to be localized to a much smaller genetic region [the approximate 5-Mb Down syndrome critical region, or DCR, and the small nuclear ribonu-

cleoprotein polypeptide N (*SNRPN*) locus for a proportion of PWS patients] that can be causative if it alone is altered (24,25). For the analysis of CNPs and disease, it is absolutely critical that if the WGA template is used, there are no artifactual changes in copy number introduced, as such changes would be indistinguishable from any true disease-associated CNPs.

In this study, we performed chromosomal copy number analysis of whole genome-amplified DNA using data derived from the 10K GeneChip Human Mapping Array, a panel of 10,204 SNPs spaced throughout the genome at a median distance of 113 kb that has been successfully used to infer copy number in several studies (17,26–28). The examination of CNPs in the genome of 532 individually genotyped samples following MDA revealed the presence of multiple CNPs (both amplifications and deletions) in every sample examined. In these 532 samples, there was an average of 1278 amplifications and 80 deletions per sample as detected by dChipSNP. One in every approximately 7 SNPs had a copy number change, with duplications occurring nearly 16 times more frequently than deletions. This was not a function of the GeneChip assay performance, as the top eight performing samples, based on SNP call rate, exhibited similar multiple copy number alterations. A commonly amplified region on chromosome 9 (bp 23,651,850–25,716,736) was validated using real-time PCR.

Our findings demonstrate genome-wide copy number changes resulting from MDA and suggest MDA introduces these changes consistently across certain hot spot regions of the genome. The current role of CNPs in health and disease are poorly understood (18,22,29), and our results here stress the need for caution when drawing conclusions regarding chromosomal copy number from MDA-treated genomic DNA on microarrays.

## MATERIALS AND METHODS

### Patient Cohort

The cohort examined in the CNP screen included 532 individuals. These individuals were under study because

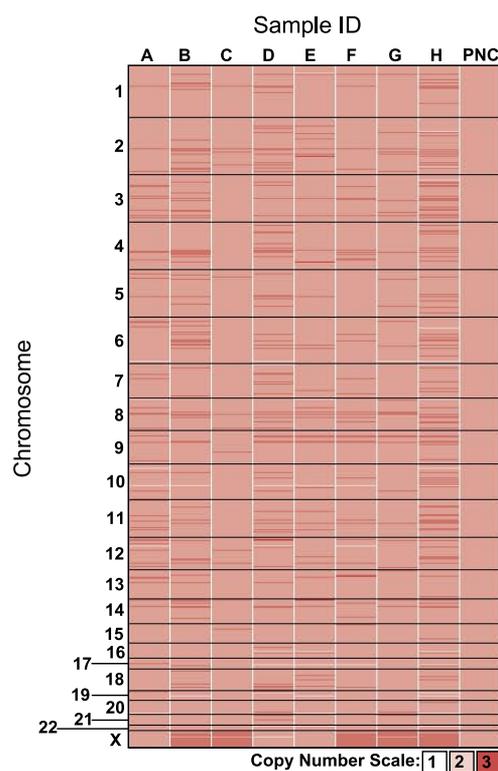
they belonged to an immediate family unit containing at least one autistic sibling or child, therefore, our initial MDA-amplified cohort included a mixture of both autistic children and their healthy siblings and parents.

### REPLI-g® WGA of Purified Genomic DNA

WGA by MDA was performed using the REPLI-g kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. In short, 10 ng template DNA were denatured with KOH-EDTA buffer (0.4 M KOH, 0.01 M EDTA, pH 8.0, in nuclease-free water) at room temperature for 3 min, neutralized with the addition of a 10-fold dilution of Solution B (Qiagen) in nuclease-free water, and amplified [1× REPLI-g buffer, 0.5 µL REPLI-g ( $\phi$ 29) polymerase, in nuclease-free water for a total volume of 50 µL] at 30°C for 8–10 h. After amplification, the  $\phi$ 29 polymerase was inactivated by heating to 65°C for 3 min. One should note that this amount of input DNA was recently reported by Bergen and colleagues to be the minimum needed to achieve satisfactory call rates in an individual TaqMan® SNP genotyping assay (15). Based on the high average SNP call rate for the samples in our cohort, it is unlikely that the use of the lowered amount of input DNA had any major effect on the copy number analysis reported here.

### Genome-Wide SNP Genotyping on the 10K GeneChip

SNP genotyping was performed using the GeneChip Mapping 10K assay kit (Affymetrix, Santa Clara, CA, USA) as previously described (30). According to the manufacturer's protocol, genomic DNA was digested with *Xba*I, ligated to adaptor molecules, and amplified using PCR primers directed against the adaptor sequence. The PCR was performed at the following amplification parameters (35 cycles at 95°C for 20 s, 59°C for 15 s, and 72°C for 15 s)

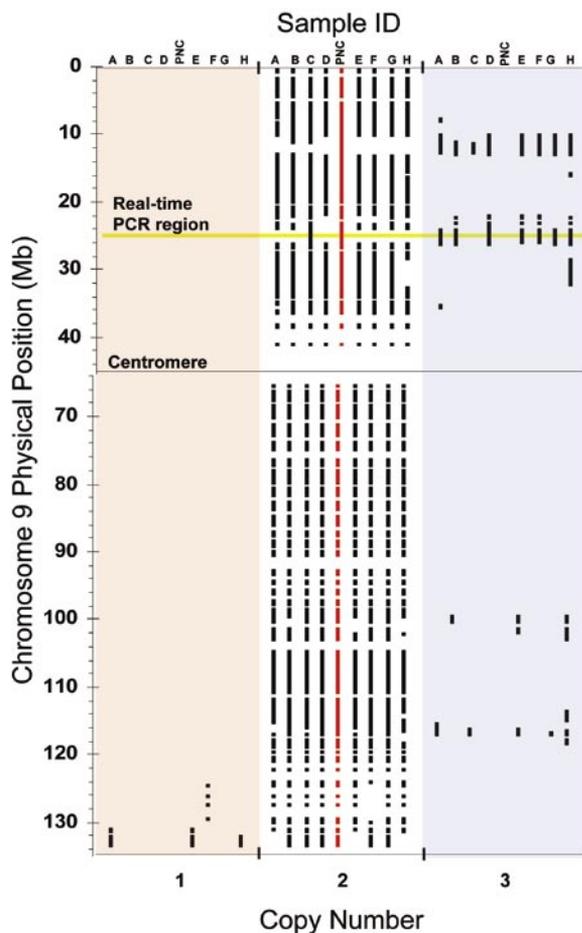


**Figure 1. Genome-wide dChipSNP copy number analysis of the top eight samples based on single nucleotide polymorphism (SNP) call rate following multiple displacement amplification (MDA).** Individual genomic DNA samples were amplified by MDA (REPLI-g), analyzed on the 10K GeneChip, and dChipSNP copy number calculations were performed using 32 pooled normal control (PNC) DNA samples for normalization (see the Materials and Methods section). A representative PNC sample is included for reference. Color scale corresponding to relative copy number is shown at bottom. This image is a direct screen shot of the dChipSNP analysis.

in order to generate products between 250 and 1000 bp in size. Twenty micrograms PCR product was purified, digested with 0.24 U DNase I, and end-labeled with biotin. Biotinylated DNA fragments were hybridized to the surface of the GeneChip for 18 h at 48°C. Microarrays were washed and stained on the Fluidics Station 450 and scanned on the GeneChip Scanner 3000 (both from Affymetrix). Median call rates and signal detection percentages for all arrays were 88.72% and 94.30%, respectively.

### Copy Number Analysis Using dChipSNP

The copy number analysis package within dChipSNP ([www.biostat.harvard.edu/complab/dchip](http://www.biostat.harvard.edu/complab/dchip)) (31) was



**Figure 2. Graphical representation of smoothed dChip-SNP copy number values for chromosome 9.** Samples were processed for copy number analysis as detailed in the Materials and Methods section. Genome-smoothed values were exported from dChip-SNP for creation of this figure. Black bars represent the calculated copy number and actual physical position of the corresponding single nucleotide polymorphism (SNP) on the 10K GeneChip. Pooled normal control (PNC) values for each SNP are illustrated in red. The region chosen for further real-time PCR validation is indicated by the horizontal yellow line. Please note truncation of the figure on both sides of the SNP-less centromeric region indicated by the horizontal black line.

utilized to analyze SNP signal intensity values as previously described (32). First, signal intensity values across each array were normalized using the invariant set normalization method. Raw copy number using the perfect match-mismatch difference model was obtained directly from dChipSNP using the normalized arrays. A Hidden Markov Model was used to infer the whole integer copy number (i.e., genome-smoothed) from the raw copy number values. dChipSNP analysis included experimental samples and pooled normal control (PNC) samples. Thirty-two PNCs were included in each sample group analyzed (group sizes ranged from 36–125 samples). The PNC consists of an equimolar pool of 100 healthy individual genomic DNAs prior to standard GeneChip processing. Because these PNC samples are a mixture of 100 individuals, they should exhibit an average copy number of 2 across the genome, thereby masking any

of the common CNPs recently reported in the literature. We have found that the inclusion of PNC samples results in higher quality dChipSNP analysis.

### Copy Number Analysis Using Real-Time PCR

Real-time PCR was performed on a DNA Engine Opticon® System (MJ Research, Waltham, MA, USA) on the MDA-treated DNA from the top eight performing (by call rate) samples. Reaction parameters were as follows: 95°C for a 5-min initial denaturation and activation of enzyme, then cycled at 95°C for 15 s, 66°C for 15 s, and 72°C for 30 s for a total of 40 cycles. Each real-time PCR consisted of 10 ng WGA DNA, 2.25 mM MgCl<sub>2</sub>, 25 μM each primer, DNase-free water, and 1× LightCycler™ FastStart DNA Master SYBR® Green I mix (contains FastStart Taq DNA polymerase; Roche, Indianapolis, IN, USA). Template

DNA concentrations were measured using PicoGreen® reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions to ensure equal amounts of template across all samples. Primers were designed using Primer3 (frodo.wi.mit.edu/cgi-bin/primer3/primer3\_www.cgi) and were synthesized by Invitrogen. Product sizes were limited to <200 bp and included a single base pair GC clamp. Sequences are as follows, 5'-TGCTCTGGTTAGGGCTTCC-3' and 5'-TCTGCTCAATTTTGCTGTGG-3' directed against chromosome 9p21.3 to detect the CNP, 5'-TGAGCAAAGGGACAAAAGC-3' and 5'-GAGGTGTGAACATGCACAGG-3' directed against chromosome 11q12.2, and 5'-TGGATGTTTCCTTG TGATCG-3' and 5'-ATTCCAGAAACC CAAAACC-3' directed against chromosome 17q12. Copy number was calculated using the  $2^{-\Delta\Delta C_T}$  method (where  $C_T$  is the cycle threshold) using both the 11q12.2 and 17q12 regions as the internal (two-copy) reference controls (33).

## RESULTS AND DISCUSSION

### MDA Introduces Copy Number Polymorphism Artifacts

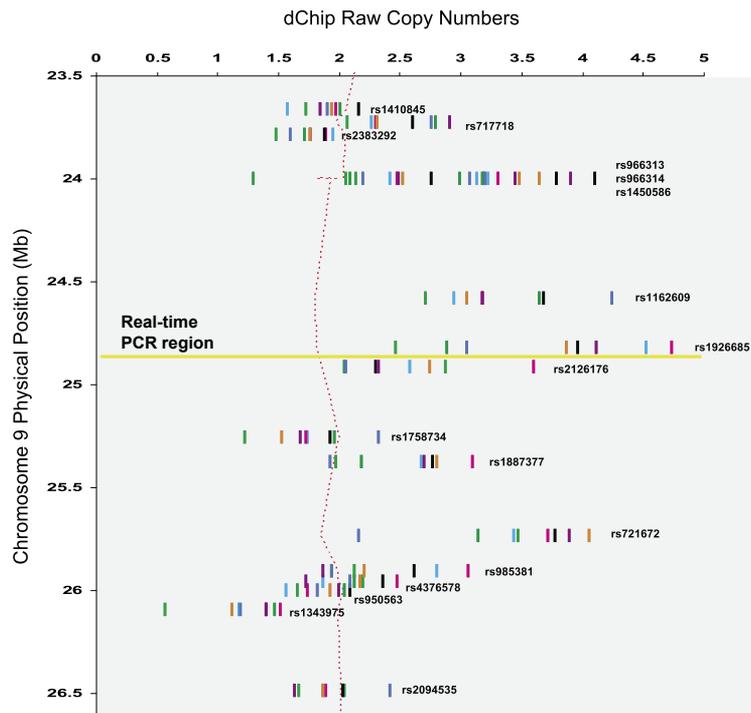
REPLI-g amplified samples (532) were analyzed using dChipSNP as detailed in the Materials and Methods section and visually inspected for CNPs (full data set not shown). This initial analysis clearly indicated that (i) MDA-processed genomic DNA applied to the 10K GeneChip yielded a wide array of CNPs upon analysis in dChipSNP and (ii) several regions across the genome were consistently represented as amplified or deleted. For illustrative purposes, Figure 1 shows the dChipSNP genome-smoothed chromosome screenshot of the top eight performing arrays with regard to overall SNP call rates (call rates >98.7%). The top performing arrays were chosen to minimize any variation due to performance of the GeneChip assay. The median number of CNPs identified per individual using this type of analysis approach was 438 CNPs. This is in stark contrast to the 1–2 total number

## Short Technical Reports

of CNPs recently found in healthy individuals in several publications.

Our cohort is one of the largest investigated to date for CNPs using microarray data, and these results suggest that the MDA procedure can introduce artifactual CNP changes in the resulting data. In a recent study conducted in our laboratory to detect CNPs in both healthy and diseased individuals, we were forced to exclude any individual with greater than three CNPs (as determined using identical analysis procedures as in this manuscript), as this indicates a sample with below average (i.e., low SNP call rate) microarray data (data not shown). If we applied that same exclusion criterion to the present study, we would have excluded 72% of the individuals from further analysis. For the 532 samples analyzed in this report, the median SNP call rate was 88.72%, clearly indicating that the samples were performing quite well for SNP genotype calling, but were failing the quality control metrics for use in CNP study.

Figure 2 illustrates the relative genome-smoothed copy numbers across chromosome 9 for the top eight samples, including a PNC for reference. It is immediately clear from this figure that the MDA procedure is preferentially amplifying distinct regions of the chromosome across several samples (the 12 and 25 Mb region). Additionally, individual samples show alterations at other distinct regions (the 126 Mb region for sample F and the 36 Mb region for sample A). Shown in yellow is the common region that was investigated further using real-time PCR for validation of copy number. Figure 3 illustrates a further detailed view of the approximately 3 Mb region surrounding the real-time PCR region (yellow line).



**Figure 3. Individual raw copy number results from dChipSNP for the 3-Mb region flanking the real-time PCR validated area.** The rs identification numbers for each single nucleotide polymorphism (SNP) are indicated along with the raw calculated copy number for each of the eight (A, turquoise; B, pink; C, dark green; D, black; E, purple; F, bright green; G, orange; H, blue) samples with the highest call rates on the 10K GeneChip following multiple displacement amplification (MDA). The red dotted line indicates the median score for the respective SNP, as determined by the pooled normal control (PNC) samples. The region of real-time PCR is indicated by the yellow line.

The raw copy number value for each SNP within each sample is represented in the context of the PNC values. Table 1 reports the values obtained from the real-time PCR analysis of the specified region. Both the dChip raw and inferred (or genome-smoothed) values are  $>2$ , with the exception of the inferred value for sample C, while the average real-time PCR values are all  $>5$ , with 3 out of 8 samples yielding copy number values  $>10$ . Sample C showed a discrepancy between smoothed dChip and real-

time PCR results; however this is most likely due to an inability of the Hidden Markov Model used by dChipSNP to detect the exact boundary region of each CNP. These values were calculated using the well known  $\Delta\Delta C_T$  methodology using two different regions with dChipSNP predicated copy number of 2 as the internal control regions (see the Materials and Methods section). These results indicate that our analysis methods probably underestimate the true copy number changes induced by MDA and

**Table 1.  $\Delta\Delta C_T$  and dChip Copy Number Values for the Chromosome 9p21.3 Region**

Sample ID	A	B	C	D	E	F	G	I
dChip Raw	4.52	4.73	2.88	3.96	4.11	3.87	3.86	4.51
dChip Inferred	3.00	3.00	2.00	3.00	3.00	3.00	3.00	3.00
Real-Time PCR 1	5.47	14.73	8.95	7.55	11.71	5.76	7.42	15.52
Real-Time PCR 2	4.62	26.55	5.80	7.44	10.39	4.41	11.69	17.51
Real-Time PCR 3	5.64	13.10	5.10	6.14	10.95	4.92	8.26	15.11
Average	5.24 $\pm$ 0.55	18.13 $\pm$ 7.34	6.62 $\pm$ 2.05	7.04 $\pm$ 0.78	11.01 $\pm$ 0.66	5.03 $\pm$ 0.69	9.12 $\pm$ 2.26	16.04 $\pm$ 1.28

The  $\Delta\Delta C_T$  and dChip copy number values are for the chromosome 9 region (bp 24,864,403–24,864,594) in the top eight samples according to overall SNP call rate.  $C_T$ , cycle threshold; SNP, single nucleotide polymorphism. Average is  $\pm$  sd.

are, if anything, a conservative measure of the effects of the WGA process on chromosomal copy number.

The alterations in copy number reported here are likely solely due to the initial MDA treatment of the isolated genomic DNA. This is clear for two key reasons. First, since the majority of the identified CNPs in this study have not been reported in the Database of Genomic Variants (34), they are likely not true common CNPs. Secondly, overall SNP performance (i.e., frequency of genotype calling) cannot be to blame for the frequent identification of regions as CNPs, as the overall median call rate for the cohort was 88.72%. Therefore, the large number of CNPs identified in this study are apt to be due to the MDA treatment alone.

We strongly believe that CNPs should be considered an important part of the study of genetic diversity and disease. Recently, an article detailing that 100 common intermediate length deletion CNPs are in linkage disequilibrium (LD) with their surrounding SNPs illustrates the point that some CNPs probably arose during a single event in human history, like SNPs, and share similar evolutionary inheritance patterns (19). Therefore, CNPs could conceivably be used as genetic markers in addition to representing a probably important component to genetic/phenotypic variation of complex traits and disease. The importance of preserving precious genomic DNA samples collected from the recruited cohort of patients for genetic study cannot be understated. The ultimate goal of any clinical genetic study should be to collect not only enough biological material in the proper way for immediate study, but also to create a bank of genetic material for future investigation as technology and the genetic understanding of disease advances. This is becoming apparent as our understanding of CNPs in relation to disease improves. Although our results herein were illustrated on the 10K GeneChip platform, it is expected that the current generation of higher density SNP microarrays (100K and 500K chipsets) would perform similarly on MDA-WGA treated samples. As these chipsets are expected only to increase in usage as whole genome association studies become more commonplace, the

results discussed here have the possibility to create a bigger impact on the field.

Our findings may have further implications toward some designs of genome-wide SNP association (GWA) studies. GWA studies are of increasing prominence, since they allow a researcher to identify common variants that predispose one to disease. Arguably, GWA studies are the primary reason for development of ultra-high density SNP genotyping microarrays, such as the 500K GeneChip Mapping arrays (Affymetrix) and the 320K/550K HumanHap arrays (Illumina, San Diego, CA, USA). Due to the exorbitant costs of these studies, some groups are performing pooling-based GWA studies, whereby hundreds of individuals are added to pools in equimolar amounts under the assumption that each individual has contributed equal copy number (35–37). Because the representative case-control DNA pools for this type of study must consist of equimolar contributions of DNA from each individual comprising the pools, this approach would also be sensitive to the MDA-WGA effects reported in this manuscript, and caution should be exercised by those researchers using pooled case-control study to ensure that no MDA-amplified samples are included in either pool. Consequentially, power is unknowingly lost and substantially uneven throughout the genome.

The results reported here contrast with those reported in 2004 by Paez et al. (10). One of the aims of their study was to examine the effects of MDA treatment on cancer-associated copy number using the 10K GeneChip. They reported no significant alterations in copy number after MDA treatment. Our study differs in several key respects including sample size ( $n = 532$  versus  $n = 16$  for Paez et al.) and the type of disease under investigation. One of the hallmarks of cancer are alterations in chromosomal copy numbers, and it is possible that MDA treatment might exhibit a much smaller effect in that setting because several regions in the genome already have an elevated copy number, whereas in our cohort of primarily healthy individuals, any copy number alterations are rare, and therefore any artifact introduced by MDA might be greatly amplified.

Additionally, the majority of the conclusions by Paez et al. (10) were drawn by comparing non-MDA-treated versus MDA-treated DNA from the same sample. We agree that this is the ideal control when using MDA, however this type of comparison is typically not “real world,” as the primary reason for using MDA is to amplify DNA from clinical samples that were collected previously and that the researcher has no way of recovering additional DNA.

It is foreseeable that the reanalysis of DNA samples, collected purely for microsatellite or SNP genotyping, will be undertaken with a renewed focus on CNP discovery. It is established that individual and genome-wide SNP genotyping and product-based resequencing approaches can succeed on MDA-treated DNA samples. However, MDAs affects on copy number analysis are well understood, and our results here stress the need to, when possible, reserve some of the original (i.e., non-MDA-treated) sample DNA in order to perform unbiased CNP analysis.

#### ACKNOWLEDGMENTS

*Funding was provided in part by the Bisgrove charitable donation (D.A.S and D.W.C.) and the National Institutes of Health (NIH) Neuroscience Blueprint (U24NS051872 to D.A.S.).*

#### COMPETING INTERESTS STATEMENT

*The authors declare no competing interests.*

#### REFERENCES

1. Sigurdson, A.J., M. Ha, M. Cosentino, T. Franklin, K.A. Haque, Y. Qi, C. Glaser, Y. Reid, et al. 2006. Long-term storage and recovery of buccal cell DNA from treated cards. *Cancer Epidemiol. Biomarkers Prev.* 15:385-388.
2. Mulot, C., I. Stucker, J. Clavel, P. Beaune, and M.A. Loriot. 2005. Collection of human genomic DNA from buccal cells for genetics studies: comparison between cytobrush, mouthwash, and treated card. *J. Biomed. Biotechnol.* 2005:291-296.
3. Wang, G., C. Brennan, M. Rook, J.L. Wolfe, C. Leo, L. Chin, H. Pan, W.H. Liu, et al. 2004. Balanced-PCR amplification allows unbiased identification of genomic copy changes in minute cell and tissue samples. *Nucleic Acids Res.* 32:e76.
4. Ng, G., I. Roberts, and N. Coleman. 2005. Evaluation of 3 methods of whole-genome

- amplification for subsequent metaphase comparative genomic hybridization. *Diagn. Mol. Pathol.* 14:203-212.
5. **Kittler, R., M. Stoneking, and M. Kayser.** 2002. A whole genome amplification method to generate long fragments from low quantities of genomic DNA. *Anal. Biochem.* 300:237-244.
  6. **Casas, E. and B.W. Kirkpatrick.** 1996. Evaluation of different amplification protocols for use in primer-extension preamplification. *BioTechniques* 20:219-225.
  7. **Telenius, H., N.P. Carter, C.E. Bebb, M. Nordenskjold, B.A. Ponder, and A. Tunnacliffe.** 1992. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 13:718-725.
  8. **Dean, F.B., S. Hosono, L. Fang, X. Wu, A.F. Faruqi, P. Bray-Ward, Z. Sun, Q. Zong, et al.** 2002. Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA* 99:5261-5266.
  9. **Hosono, S., A.F. Faruqi, F.B. Dean, Y. Du, Z. Sun, X. Wu, J. Du, S.F. Kingsmore, et al.** 2003. Unbiased whole-genome amplification directly from clinical samples. *Genome Res.* 13:954-964.
  10. **Paez, J.G., M. Lin, R. Beroukhim, J.C. Lee, X. Zhao, D.J. Richter, S. Gabriel, P. Herman, et al.** 2004. Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Res.* 32:e71.
  11. **Blanco, L., A. Bernad, J.M. Lazaro, G. Martin, C. Garmendia, and M. Salas.** 1989. Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J. Biol. Chem.* 264:8935-8940.
  12. **Esteban, J.A., M. Salas, and L. Blanco.** 1993. Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *J. Biol. Chem.* 268:2719-2726.
  13. **Holbrook, J.F., D. Stabley, and K. Sol-Church.** 2005. Exploring whole genome amplification as a DNA recovery tool for molecular genetic studies. *J. Biomol. Tech.* 16:125-133.
  14. **Sun, G., R. Kaushal, P. Pal, M. Wolujewicz, D. Smelser, H. Cheng, M. Lu, R. Chakraborty, et al.** 2005. Whole-genome amplification: relative efficiencies of the current methods. *Leg. Med. (Tokyo)* 7:279-286.
  15. **Bergen, A.W., Y. Qi, K.A. Haque, R.A. Welch, and S.J. Chanock.** 2005. Effects of DNA mass on multiple displacement whole genome amplification and genotyping performance. *BMC Biotechnol.* 5:24.
  16. **Tzvetkov, M.V., C. Becker, B. Kulle, P. Nurnberg, J. Brockmoller, and L. Wojnowski.** 2005. Genome-wide single-nucleotide polymorphism arrays demonstrate high fidelity of multiple displacement-based whole-genome amplification. *Electrophoresis* 26:710-715.
  17. **Wong, K.K., Y.T. Tsang, J. Shen, R.S. Cheng, Y.M. Chang, T.K. Man, and C.C. Lau.** 2004. Allelic imbalance analysis by high-density single-nucleotide polymorphic allele (SNP) array with whole genome amplified DNA. *Nucleic Acids Res.* 32:e69.
  18. **McCarroll, S.A., T.N. Hadnott, G.H. Perry, P.C. Sabeti, M.C. Zody, J.C. Barrett, S. Dallaire, S.B. Gabriel, et al.** 2006. Common deletion polymorphisms in the human genome. *Nat. Genet.* 38:86-92.
  19. **Hinds, D.A., A.P. Kloek, M. Jen, X. Chen, and K.A. Frazer.** 2006. Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nat. Genet.* 38:82-85.
  20. **Tuzun, E., A.J. Sharp, J.A. Bailey, R. Kaul, V.A. Morrison, L.M. Pertz, E. Haugen, H. Hayden, et al.** 2005. Fine-scale structural variation of the human genome. *Nat. Genet.* 37:727-732.
  21. **Newman, T.L., E. Tuzun, V.A. Morrison, K.E. Hayden, M. Ventura, S.D. McGrath, M. Rocchi, and E.E. Eichler.** 2005. A genome-wide survey of structural variation between human and chimpanzee. *Genome Res.* 15:1344-1356.
  22. **Hinds, D.A., L.L. Stuve, G.B. Nilsen, E. Halperin, E. Eskin, D.G. Ballinger, K.A. Frazer, and D.R. Cox.** 2005. Whole-genome patterns of common DNA variation in three human populations. *Science* 307:1072-1079.
  23. **Conrad, D.F., T.D. Andrews, N.P. Carter, M.E. Hurler, and J.K. Pritchard.** 2006. A high-resolution survey of deletion polymorphism in the human genome. *Nat. Genet.* 38:75-81.
  24. **Korenberg, J.R., X.N. Chen, R. Schipper, Z. Sun, R. Gonsky, S. Gerwehr, N. Carpenter, C. Daumer, et al.** 1994. Down syndrome phenotypes: the consequences of chromosomal imbalance. *Proc. Natl. Acad. Sci. USA* 91:4997-5001.
  25. **Cassidy, S.B.** 1997. Prader-Willi syndrome. *J. Med. Genet.* 34:917-923.
  26. **Matsuzaki, H., H. Loi, S. Dong, Y.Y. Tsai, J. Fang, J. Law, X. Di, W.M. Liu, et al.** 2004. Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Res.* 14:414-425.
  27. **Zhou, X., S.C. Mok, Z. Chen, Y. Li, and D.T. Wong.** 2004. Concurrent analysis of loss of heterozygosity (LOH) and copy number abnormality (CNA) for oral premalignancy progression using the Affymetrix 10K SNP mapping array. *Hum. Genet.* 115:327-330.
  28. **Thompson, E.R., S.C. Herbert, S.M. Forrest, and I.G. Campbell.** 2005. Whole genome SNP arrays using DNA derived from formalin-fixed, paraffin-embedded ovarian tumor tissue. *Hum. Mutat.* 26:384-389.
  29. **Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, H. Massa, et al.** 2004. Large-scale copy number polymorphism in the human genome. *Science* 305:525-528.
  30. **Huentelman, M.J., D.W. Craig, A.D. Shieh, J.J. Corneveaux, D. Hu-Lince, J.V. Pearson, and D.A. Stephan.** 2005. SNIper: improved SNP genotype calling for Affymetrix 10K GeneChip microarray data. *BMC Genomics* 6:149.
  31. **Li, C. and W.H. Wong.** 2001. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl. Acad. Sci. USA* 98:31-36.
  32. **Beroukhim, R., M. Lin, Y. Park, K. Hao, X. Zhao, L.A. Garraway, E.A. Fox, E.P. Hochberg, et al.** 2006. Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. *PLoS Comput. Biol.* 2:e41.
  33. **Livak, K.J. and T.D. Schmittgen.** 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)). *Methods* 25:402-408.
  34. **Iafate, A.J., L. Feuk, M.N. Rivera, M.L. Listewnik, P.K. Donahoe, Y. Qi, S.W. Scherer, and C. Lee.** 2004. Detection of large-scale variation in the human genome. *Nat. Genet.* 36:949-951.
  35. **Butcher, L.M., E. Meaburn, J. Knight, P.C. Sham, L.C. Schalkwyk, I.W. Craig, and R. Plomin.** 2005. SNPs, microarrays and pooled DNA: identification of four loci associated with mild mental impairment in a sample of 6000 children. *Hum. Mol. Genet.* 14:1315-1325.
  36. **Craig, D.W., M.J. Huentelman, D. Hu-Lince, V.L. Zismann, M.C. Krueger, A.M. Lee, E.G. Puffenberger, J.M. Pearson, et al.** 2005. Identification of disease causing loci using an array-based genotyping approach on pooled DNA. *BMC Genomics* 6:138.
  37. **Spector, T.D., R.H. Reneland, S. Mah, A.M. Valdes, D.J. Hart, S. Kammerer, M. Langdown, C.R. Hoyal, et al.** 2006. Association between a variation in LRCH1 and knee osteoarthritis: a genome-wide single-nucleotide polymorphism association study using DNA pooling. *Arthritis Rheum.* 54:524-532.

Received 7 July 2006; accepted 21 September 2006.

Address correspondence to Matthew J. Huentelman, Neurogenomics Division, The Translational Genomics Research Institute, 445 North Fifth Street, Phoenix, AZ 85004, USA. e-mail: mhuentelman@tgen.org

To purchase reprints of this article, contact: Reprints@BioTechniques.com