Performance of Whole-Genome Amplified DNA Isolated from Serum and Plasma on High-Density Single Nucleotide Polymorphism Arrays

Daniel T. Croft, Jr.,* Rick M. Jordan,*
Heather L. Patney,* Craig D. Shriver,[†]
Marina N. Vernalis,[‡] Trevor J. Orchard,[§]
and Darrell L. Ellsworth*[¶]

From the Clinical Breast Care Project* and the Integrative Cardiac and Metabolic Health Program,[¶] Windber Research Institute, Windber, Pennsylvania; General Surgery Services[‡] and the Integrative Cardiac Health Program,[‡] Walter Reed Army Medical Center, Washington, District of Columbia; and the Department of Epidemiology,[§] University of Pittsburgh, Pittsburgh, Pennsylvania

Defining genetic variation associated with complex human diseases requires standards based on high-quality DNA from well-characterized patients. With the development of robust technologies for whole-genome amplification, sample repositories such as serum banks now represent a potentially valuable source of DNA for both genomic studies and clinical diagnostics. We assessed the performance of whole-genome amplified DNA (wgaDNA) derived from stored serum/plasma on high-density single nucleotide polymorphism arrays. Neither storage time nor usage history affected either DNA extraction or whole-genome amplification yields; however, samples that were thawed and refrozen showed significantly lower call rates ($73.9 \pm 7.8\%$) than samples that were never thawed ($92.0 \pm 3.3\%$) ($P < 0.001$). Genotype call rates did not differ significantly ($P = 0.13$) between wgaDNA from never-thawed serum/plasma ($92.9 \pm 2.6\%$) and genomic DNA ($97.5 \pm 0.3\%$) isolated from whole blood. Approximately 400,000 genotypes were consistent between wgaDNA and genomic DNA, but the overall discordance rate of $4.4 \pm 3.8\%$ reflected an average of $11,110 \pm 9502$ genotyping errors per sample. No distinct patterns of chromosomal clustering were observed for single nucleotide polymorphisms showing discordant genotypes or homozygote conversion. Because the effects of genotyping errors on whole-genome studies are not well defined, we recommend caution when applying wgaDNA from serum/plasma to high-density single nucleotide polymorphism arrays in addition to the use of stringent quality control requirements for the resulting genotype data. (*J Mol Diagn* 2008, 10:249–257; DOI: 10.2353/jmoldx.2008.070155)

Recent technical advances have produced platforms to assay hundreds of thousands of genetic polymorphisms across the human genome. High-density single nucleotide polymorphism (SNP) genotyping panels with automated allele calling are rapidly becoming powerful tools for identifying new genes that contribute to common diseases,^{1,2} detecting copy number variations,³ and for novel applications in forensics⁴ and anthropology.^{5,6} Genome-wide SNP typing requires sufficient amounts of high-quality DNA, which is usually obtained from whole blood, cell lines, and surgical biopsies. Unfortunately, the amount of DNA is often limited in many repositories where stored serum, buccal swabs, or dried blood spots are the only sources of DNA available.

Whole-genome amplification is an important technology for genomic and molecular diagnostic studies that permits analysis of samples containing only minute amounts of DNA.⁷ Early protocols for whole-genome amplification using TaqDNA polymerase with degenerate primers^{8,9} were hampered by amplification artifacts, allelic dropout, and localized enrichment leading to incomplete genomic coverage. New amplification systems based on multiple displacement amplification and the highly processive $\phi 29$ DNA polymerase¹⁰ provide highly uniform representation across the genome regardless of DNA sequence or secondary structure, high-fidelity amplification with low error rates of only 1 in 10^6 to 10^7 nucleotides, and high yields of amplified DNA, as much as 10^7 -fold amplification.^{11–13} Resulting whole-genome amplified DNA (wgaDNA) can be directly used in microarray-based methods for genomic analysis, such as SNP genotyping¹⁴ and comparative genomic hybridization.¹⁵

Previous studies of relatively small numbers of polymorphisms have shown high concordance between

Supported by the U.S. Department of Defense (Military Molecular Medicine Initiative MDA W81XWH-05-2-0075) through the U.S. Army Medical Research and Materiel Command/Telemedicine and Advanced Technology Research Center, Fort Detrick, MD, and the Henry M. Jackson Foundation for the Advancement of Military Medicine, Rockville, MD.

Accepted for publication January 4, 2008.

The opinion and assertions contained herein are the private views of the authors and are not to be construed as official or as representing the views of the Department of the Army or the Department of Defense.

Address reprint requests to Darrell L. Ellsworth, PhD, Windber Research Institute, 620 Seventh Street, Windber, PA 15963. E-mail: d.ellsworth@wriwindber.org.

genomic DNA (gDNA) and wgaDNA from a variety of sources,^{11,16–18} but to our knowledge, no studies have evaluated performance of wgaDNA on very large numbers of SNPs using high-density arrays. To assess the potential utility of serum/plasma DNA in molecular epidemiology studies and diagnostic applications, we examined performance of wgaDNA on Affymetrix 500K arrays containing 500,568 SNPs. Our specific objectives were to assess call rates of wgaDNA from samples that had been stored for extended periods of time (up to 20 years), to determine rates of concordance in genotyping data between gDNA and wgaDNA, and to characterize the nature and genomic distribution of discordant genotypes. Information on potential uses and limitations of wgaDNA obtained from stored serum/plasma samples will be important for guiding future uses of wgaDNA in large-scale genomic studies.

Materials and Methods

Sample Collection

All research was conducted under protocols approved by the Windber Medical Center and University of Pittsburgh Institutional Review Boards. Fresh serum and plasma samples were obtained from subjects who voluntarily agreed to participate in this study and gave written informed consent. Peripheral blood was collected in serum separator or 10-ml heparinized BD vacutainers (Thermo Fisher Scientific, Waltham, MA) and placed at room temperature for 30 minutes (serum) or on wet ice (plasma). All vacutainers were centrifuged at $1300 \times g$ for 10 minutes at 4°C, then divided into 500- μ l aliquots and stored at –80°C. Stored serum/plasma samples that were originally collected 3, 10, or 20 years ago were obtained from the sample archives of the Integrative Cardiac and Metabolic Health Program at Windber Research Institute and the Epidemiology of Diabetes Complications Study at the Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh.

DNA Isolation from Serum and Plasma

Before DNA isolation, samples were removed from –80°C, allowed to thaw on ice for 30 minutes, then mixed thoroughly. To simulate sample usage over time, which is often characteristic of banked serum/plasma samples, aliquots of two recently collected serum samples were thawed and refrozen five times. DNA was extracted from 200 μ l of serum or plasma using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). Samples were treated with Qiagen protease, then buffer AL (lysis solution) at 56°C for 10 minutes using a TropicCooler (Boekel Scientific, Feasterville, PA). DNA was isolated in QIAamp spin columns, then eluted, dried, and reconstituted in a final volume of 100 μ l of AE buffer. Initial DNA concentrations and purity were determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). Visual inspection of representative serum/plasma DNA aliquots on agarose gels showed a

prominent band >10 kb in size with minimal detectable degradation.

For all samples where whole blood was available, gDNA was isolated from peripheral blood mononuclear cells with the Puregene DNA purification kit (Gentra Systems, Minneapolis, MN) according to the manufacturer's protocol. Concentration and purity of the gDNA was determined as above on the NanoDrop ND-1000 spectrophotometer.

Whole-Genome Amplification of Serum/Plasma DNA

Serum or plasma DNA aliquots (100 μ l) were concentrated to a final volume of 10 μ l at 45°C in an Eppendorf Vacufuge concentrator (Brinkman Instruments, Westbury, NY).¹⁹ Whole-genome amplification was then performed on 2.5 μ l of concentrated serum/plasma DNA subjected only to alkaline denaturation using the REPLI-g whole-genome amplification kit (Qiagen). For samples that had been thawed and refrozen, a second independent whole-genome amplification was performed using template DNA that had been subjected to both alkaline and heat (3 minutes at 95°C) denaturation to determine whether multiple whole-genome amplifications could be pooled to increase SNP call rates on 500K arrays. All amplifications were incubated for 16 hours at 30°C in the TropicCooler, followed by inactivation of the REPLI-g DNA polymerase at 65°C for 3 minutes. Amplified serum/plasma DNA was then purified using the ethanol precipitation purification protocol in the GenomiPhi DNA amplification kit (GE HealthCare, Buckinghamshire, UK). Purified wgaDNA was re-suspended in 30 μ l of TE buffer, assessed on the NanoDrop ND-1000, and then stored at –20°C until used for genome-wide SNP analysis.

SNP Analysis

Purified wgaDNA samples (diluted to 50 ng/ μ l) from serum/plasma and gDNA from whole blood were applied to Affymetrix GeneChip Human Mapping 500K arrays following the manufacturer's protocol. The DNA was cleaved with either *NspI* or *StyI* restriction enzymes, ligated to specific linkers, and amplified with the Titanium DNA amplification kit (Clontech, Mountain View, CA). Polymerase chain reaction products were purified using the Clontech DNA amplification clean-up kit, diluted to 90 μ g in 45 μ l of Clontech RB buffer, fragmented, and labeled with biotinylated GeneChip DNA labeling reagent. Samples were then hybridized to the Affymetrix GeneChip Human Mapping 500K arrays. Following hybridization, arrays were stained and scanned on a GeneChip scanner 3000. To assess the effects of a second whole-genome amplification on call rates for thawed-refrozen samples, independent reactions were pooled before SNP analysis.

Data Analysis

For all analyses, $P < 0.05$ was used to determine statistical significance. Genotypes (AA, AB, BB, or no call) were determined using the dynamic model mapping algorithm in the Affymetrix GeneChip genotyping analysis software (GTTYPE 4.0) package²⁰ at a 0.33 confidence score. For wgaDNA, SNP call rates were compared between *Nspl* and *Styl* arrays across all samples using a paired *t*-test. As no significant difference ($P = 0.10$) in call rates was detected, *Nspl* and *Styl* results were combined in subsequent analyses. We used a Wilcoxon signed-rank test, which accounts for a non-normal distribution of differences between pairs, to compare genotype call rates for gDNA versus wgaDNA and single versus multiple whole-genome amplifications from the same participant. A Pearson correlation was used to examine variability in call rates as a function of storage time in never-thawed samples and Student's *t*-test was used to assess call rates for never-thawed serum/plasma aliquots versus samples that had been thawed and re-frozen at least once over time.

Genotypes were imported into Access 2003 (Microsoft, Redmond, WA) and queried to determine levels and patterns of concordance between gDNA and wgaDNA. Spotfire DecisionSite (TIBCO Software, Palo Alto, CA) was used to examine the chromosomal distribution of SNPs that were not scored consistently between gDNA and wgaDNA. Approximate chromosomal locations for SNPs showing discordant genotypes or homozygote conversion were obtained by NetAffx (available from Affymetrix) batch queries.

Results

DNA Extractions and Whole-Genome Amplifications

DNA was extracted from 200- μ l aliquots of plasma or serum collected over a period of 20 years from 12 individuals (Figure 1). Average DNA yields ($\pm$ SD) were 366 ± 27 ng for recently collected samples, 454 ± 77 ng for samples that had been stored for 3 years, 412 ± 14 ng for 10-year samples, and 322 ± 85 ng for 20-year-old aliquots. Yields were not correlated with storage time of the sample ($r^2 = 0.18$, $P = 0.17$). The OD 260/280 ratio, a measure of DNA purity with respect to protein contamination, ranged from ~ 1.3 to ~ 1.7 (average 1.4 ± 0.1).

When aliquots of serum/plasma DNA (~ 100 ng) were subjected to whole-genome amplification, with the exception of one sample that may have been compromised during storage and thus failed to amplify, all reactions generated more than 30 μ g of wgaDNA (Figure 1). Average yield of the amplified products ($\pm$ SD) from alkaline-denatured template was 37.5 ± 3.8 μ g; average yield in the subset of thawed-refrozen samples with a second amplification from template that had been subjected to alkaline and heat denaturation was 32.6 ± 9.1 μ g. All corresponding OD 260/280 ratios were consistent at 1.9 ± 0.02 . Multiple freeze-thaw cycles did not reduce

Figure 1. DNA extraction (A) and whole-genome amplification (B) yields from 200- μ l aliquots of serum or plasma collected over a 20-year period from 12 individuals.

the efficiency of DNA extraction from serum (yield ≥ 400 ng of DNA) or the whole-genome amplification process (amplifications generated > 40 μ g wgaDNA).

SNP Genotyping Quality Control

The mapping algorithm report in GTTYPE contains several parameters that are useful for assessing quality of 500K SNP genotyping results. Call rates for the reference genomic DNA 103 (96.5% for *Nspl*, 97.8% for *Styl*) indicated good overall performance of the assays. The modified partitioning around medoids (MPAM) algorithm, developed to make genotype determinations on the GeneChip Human Mapping 10K arrays, genotypes a subset of ~ 8000 SNPs on the 500K arrays. For high-quality genomic DNA, the MPAM call rate (MCR) should be $\geq 94\%$ and the MPAM detection rate (MDR) should be $\sim 99\%$. In our gDNA samples, all parameters were within the manufacturer's recommended specifications: MCRs were $\geq 94.2\%$ and MDRs were $\geq 98.7\%$ (data not shown). Although comparable QC parameters have not been developed for wgaDNA, our wgaDNA samples showed MCR values ranging from $\sim 20\%$ to 93% and MDR values of $\sim 31\%$ to 99% (Table 1). Both MCR and MDR values were significantly higher ($P < 0.001$) in never-thawed serum/plasma samples compared to previously thawed and refrozen samples.

Table 1. Parameters for Assessing Quality of 500K Data Generated from wgaDNA

Sample	Storage time	Date of collection	QC parameter			
			MCR		MDR	
			Nspl	Styl	Nspl	Styl
109022 serum	Recent	11/02/2006	73.6	71.9	92.3	91.3
109023 serum	Recent	04/17/2007	76.8	79.2	94.6	96.0
109025 serum	Recent	04/17/2007	72.0	75.2	91.2	92.7
0270 plasma	~3 years	01/12/2004	90.2	82.5	98.4	95.3
0271 plasma	~3 years	01/12/2004	93.1	92.5	98.8	98.2
0905 serum	~10 years	02/08/1997	63.5	68.6	88.5	93.0
8773 serum	~10 years	03/05/1997	79.3	81.7	96.5	97.4
0625 serum	~20 years	09/17/1988	85.9	92.6	96.7	98.8
0531 plasma	~20 years-TR	04/04/1987	19.9	38.1	31.4	49.2
1166 plasma	~20 years-TR	04/01/1987	53.2	58.7	66.5	70.6
109023 serum	Recent-5X	04/17/2007	34.6	38.1	44.6	49.2
109025 serum	Recent-5X	04/17/2007	57.3	58.7	74.2	70.6

From the dynamic mapping algorithm report in GTYPE 4.0. ~20 years-TR, archived samples that were thawed and refrozen for other purposes at least once over time; recent-5X, samples experimentally thawed and refrozen five times to simulate heavy sample usage. Nspl arrays contain 262,264 SNPs, Styl arrays contain 238,304 SNPs.

Genotype Call Rates

Call rates for all gDNA samples were >97%, well above the Affymetrix recommended minimum call rate of 93% (Table 2). In wgaDNA samples, call rates showed more variability, ranging from ~63% in a 20-year-old sample that had been thawed and refrozen at least once to ~97% in a 20-year-old never-thawed serum sample. The percentage of genotype calls did not differ significantly between gDNA and wgaDNA ($P = 0.13$) for samples that were recently collected or stored for approximately 3 years (Table 2). For wgaDNA from never-thawed serum/plasma samples collected at various times over a 20-year period, no significant correlation was observed between call rates and age of the sample ($r^2 = 0.08$, $P = 0.51$), indicating that call rates did not diminish with increasing storage time. However, samples that had been thawed and refrozen, either experimentally or due to usage over time, showed significantly lower call rates ($P < 0.001$) than never-thawed samples. Pooling of two independent whole-genome amplifications with different initial dena-

turation conditions improved call rates in the thawed-refrozen samples by $2.3 \pm 3.5\%$ ($P < 0.05$).

Concordance Between wgaDNA and gDNA

To evaluate whether wgaDNA gives robust genotype data on high-density SNP arrays, we determined concordance between wgaDNA and gDNA considering only SNPs that gave genotype calls in both amplified and unamplified samples. For all individuals where gDNA was available for study, approximately 400,000 or more SNPs (~80%+ of all SNPs on the 500K arrays) yielded concordant genotype calls between gDNA and wgaDNA (Table 3). In all cases, discordance rates were <10%; average discordance ($\pm$ SD) was $4.4 \pm 3.8\%$. Heterozygote drop-out in wgaDNA accounted for the majority of discordant genotypes, ranging from 60% to 91% of all discrepancies (average $78.5 \pm 14.1\%$; see Table 4). All other differences comprised only a minor component of the overall

Table 2. SNP Call Rates (%) on 500K Arrays from gDNA and wgaDNA

Sample	Storage time	Date of collection	DNA			
			gDNA		wgaDNA	
			Nspl	Styl	Nspl	Styl
109022 serum	Recent	11/02/2006	—	—	89.1	89.0
109023 serum	Recent	04/17/2007	97.4	97.3	91.4	93.0
109025 serum	Recent	04/17/2007	97.3	97.7	89.0	90.9
0270 plasma	~3 years	01/12/2004	97.9	97.1	95.4	91.4
0271 plasma	~3 years	01/12/2004	97.8	97.8	96.1	95.8
0905 serum	~10 years	02/08/1997			85.1	88.2
8773 serum	~10 years	03/05/1997			92.6	93.6
0625 serum	~20 years	09/17/1988			94.1	96.8
0531 plasma	~20 years-TR	04/04/1987			62.9	65.0
1166 plasma	~20 years-TR	04/01/1987			79.2	81.6
109023 serum	Recent-5X	04/17/2007			69.7	70.5
109025 serum	Recent-5X	04/17/2007			80.9	81.6

~20 years-TR, archived samples that were thawed and refrozen for other purposes at least once over time; recent-5X, samples experimentally thawed and refrozen five times to simulate heavy sample usage. Nspl arrays contain 262,264 SNPs, Styl arrays contain 238,304 SNPs.

Table 3. Concordance Rates for SNP Genotyping on 500K Arrays Between gDNA and wgaDNA

Sample	No. unscored SNPs in gDNA	No. unscored SNPs in wgaDNA	No. concordant genotypes	% concordance	No. discordant genotypes	% discordance
109023– <i>Nspl</i>	6862	22,537	219,491	83.7	17,376	6.6
109023– <i>Styl</i>	6365	16,652	203,331	85.3	15,298	6.4
109025– <i>Nspl</i>	7027	28,945	205,991	78.5	24,663	9.4
109025– <i>Styl</i>	5468	21,690	193,427	81.2	21,055	8.8
0270– <i>Nspl</i>	3957	10,610	244,516	93.2	3181	1.2
0270– <i>Styl</i>	6912	20,549	212,074	89.0	4081	1.7
0271– <i>Nspl</i>	5714	10,276	247,696	94.4	1656	0.6
0271– <i>Styl</i>	5156	10,086	224,844	94.4	1572	0.7

Nspl arrays contain 262,264 SNPs; *Styl* arrays contain 238,304 SNPs. The number of concordant and discordant genotypes includes only those SNPs that gave genotype calls in both amplified (wgaDNA) and unamplified (gDNA) samples, thus % concordance + % discordance does not equal 100%.

discordance in genotype calls between gDNA and wgaDNA.

Analysis of the genomic distribution of SNPs yielding discordant genotypes showed no distinct patterns of chromosomal clustering. When all discordant genotypes (including allelic dropout) across multiple patients were combined, the distribution appeared relatively uniform throughout the genome (Figure 2). Likewise, SNPs showing homozygote conversion, a pattern of discordance involving homozygous genotypes for the alternate allele (eg, AA in gDNA, BB in wgaDNA), appeared to be evenly distributed (Figure 3). Complete data have been deposited to the NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_viewTable.cgi?handle=WINDBER_RESEARCH).

Discussion

Major efforts to define genetic variation associated with complex human diseases require high-quality DNA from well-characterized patients. Many studies with abundant long-term demographic and risk factor information, however, have only limited amounts of DNA available for genomic studies. Serum banks²¹ for example, often contain tens of thousands of samples and traditionally have been used for plasma-based screening in diverse epidemiological surveys.^{22,23} With the development of robust technologies for whole-genome amplification, these repositories now represent a potentially valuable source of DNA for genomic studies and molecular diagnostic uses.

Whole-genome amplification has been performed successfully on DNA recovered from several different body fluids including saliva,²⁴ urine,¹⁸ amniotic fluid,²⁵ and plasma,¹⁹ and in general, the resulting SNP genotype data from wgaDNA were >97% concordant with SNP data from gDNA for small numbers of SNPs. Studies using whole-genome amplification to rescue DNA from plasma samples stored under optimal conditions have observed genotyping success rates of >93% and error rates of <1% for eight SNPs.¹⁹ Despite the robust SNP genotyping results attained with high-quality samples, little information exists regarding the effects of suboptimal long-term storage and thawing and refreezing of serum/plasma samples on the genotyping performance of wgaDNA using very high density SNP arrays.

Collections of archived biological samples normally experience a gradual deterioration in DNA quality over time, which may be accelerated by repeated thawing and refreezing as samples are used to assess other biomarkers. Overall call rates and accuracy of genotyping results in studies that rely on whole-genome amplification can be highly variable,²⁶ but generally correlate with the quality and/or quantity of the template DNA being used in the amplification reactions. Poor-quality DNA when used as a template for whole-genome amplification can reduce the efficiency of the reactions and increase the prevalence of artifacts such as loss of uniform genomic coverage and unbalanced amplification of alleles at specific sites (allelic dropout).^{26–29} Such genotyping errors de-

Table 4. Nature of Discordant Genotypes from 500K Arrays Between gDNA and wgaDNA

Nature of discordance		109023			109025			0270			0271		
gDNA genotype	wgaDNA genotype	<i>Nspl</i>	<i>Styl</i>	Total % of discordant genotypes	<i>Nspl</i>	<i>Styl</i>	Total % of discordant genotypes	<i>Nspl</i>	<i>Styl</i>	Total % of discordant genotypes	<i>Nspl</i>	<i>Styl</i>	Total % of discordant genotypes
AA	BB	64	43	0.3	214	134	0.8	1	26	0.4	2	1	0.1
BB	AA	72	53	0.4	257	195	1.0	5	22	0.4	3	2	0.1
AA	AB	700	607	4.0	1389	990	5.2	375	447	11.3	278	278	17.2
BB	AB	735	558	4.0	1508	1068	5.6	452	445	12.4	361	371	22.7
AB	AA	8302	7365	47.9	11,330	9773	46.2	1279	1710	41.1	529	525	32.7
AB	BB	7503	6672	43.4	9965	8895	41.2	1069	1431	34.4	483	395	27.2
Total		17,376	15,298		24,663	21,055		3181	4081		1656	1572	

Nspl arrays contain 262,264 SNPs; *Styl* arrays contain 238,304 SNPs. Total % of discordant genotypes includes discordant genotypes from both *Nspl* and *Styl* chips for each sample.

Figure 2. Genomic distribution of SNPs yielding discordant genotypes between wgaDNA derived from stored serum/plasma and gDNA isolated from peripheral blood on *NspI* (A) and *StyI* (B) arrays. The centromeres are indicated by gray bars.

crease statistical power for detecting associations between genetic variants and complex disease phenotypes and can strongly affect individual identification in forensic cases.³⁰

To our knowledge, this study is the first to evaluate performance of wgaDNA derived from serum/plasma on large-scale SNP genotyping and examine the effects of sample storage/usage on genotyping performance. We observed that frozen serum/plasma collected in different laboratories over a 20-year period produced DNA yields within the range of plasma DNA concentrations previously obtained in other laboratories using similar protocols.^{31,32} Whole-genome amplification of the resulting serum/plasma DNA generated ample amounts (>30 μ g) of relatively pure (OD 260/280 ratio ~1.9) wgaDNA, comparable to previous investigations using similar technology.¹⁹ When applied to Human Mapping 500K arrays however, the wgaDNA samples showed significant variability in SNP call rates and this variability appeared to be affected by long-term storage conditions. In our study of a limited number of samples, overall call rates for wgaDNA from never-thawed serum/plasma did not differ significantly from call rates derived from high-quality gDNA and did not diminish with increasing storage time.

Conversely, samples that had been thawed and refrozen showed significant deterioration in call rates compared to never-thawed samples.

A high degree of concordance in SNP genotype calls has been observed between unamplified (genomic) DNA and wgaDNA when high-quality gDNA was used as the template,^{28,33} but degraded or very low copy number templates usually reduce the yield of wgaDNA and result in a higher frequency of genotyping errors.^{34,35} Thawing and refreezing may alter the structural integrity of extracted DNA, leading to artifacts in the wgaDNA. In our experience, high yields of wgaDNA can be routinely generated from thawed and refrozen serum/plasma samples, but these samples have significantly lower call rates than never thawed samples. Therefore, multiple freeze-thaw cycles do not necessarily reduce the efficiency of DNA extraction from serum/plasma or the whole-genome amplification process, but the resulting wgaDNA is not likely to be of sufficient quality for accurate large-scale SNP genotyping.

Never-thawed serum/plasma, even when stored for extended periods of time, can yield accurate genotype data for very small numbers of markers,³⁶ but large numbers of SNPs on high density arrays may produce unac-

Figure 3. Genomic distribution of SNPs showing homozygote conversion between wgaDNA derived from stored serum/plasma and gDNA isolated from peripheral blood on *Nsp I* (A) and *Sty I* (B) arrays. The centromeres are indicated by gray bars.

ceptably high error rates even with well-preserved samples. Our data suggest that wgaDNA from serum/plasma may be useful for genomic studies and molecular diagnostic applications using Affymetrix 500K arrays, but stringent quality requirements need to be established. Because the discordance rate is closely associated with the overall call rate, the selection criterion could be a minimum call rate of 93%, as recommended by the manufacturer for genomic DNA. Using this selection strategy, only four of the eight never-thawed samples, and none of the thawed/refrozen samples, would be deemed acceptable. The average discordance rate for acceptable samples with gDNA available (0270 and 0271, Table 3) was 1.1%, which was equivalent to an average of 5245 genotyping errors per sample compared to genomic DNA. Unfortunately, the effects of genotyping error rates on whole-genome studies have not been well defined. Further research is needed to determine the number of genotyping errors and the distribution of these errors across samples that would lead to erroneous conclusions such as false positive or false negative associations.

Critical requirements for reliable SNP genotyping from wgaDNA include balanced amplification of different regions (uniform genomic coverage) and balanced ampli-

fication of both alleles at each genetic locus (no allelic dropout).²⁷ Previous studies have observed that specific regions of the genome may not be replicated efficiently during whole-genome amplification due to local base composition (GC content) and the presence of repetitive elements.³³ However, these under-represented regions do not appear to be restricted to specific areas such as centromeres and telomeres. When we examined the genomic location of SNPs showing discordant genotypes between gDNA and wgaDNA and SNPs showing homozygote conversion across multiple patients, the distribution appeared relatively uniform throughout the genome, suggesting that no genomic regions were more prone to genotyping errors than other regions. Therefore, we believe that 500K SNP data generated from high-quality wgaDNA will be unlikely to show localized loss of information in genome-wide association studies.

In conclusion, whole-genome amplification is a promising technology for a growing number of research and diagnostic applications, but the amount and quality of input DNA may limit applicability in many cases.³⁷ Archived samples with storage and usage histories that promote degradation of high molecular weight DNA appear more likely to produce suboptimal genotyping per-

formance following whole genome amplification.³⁸ In this study, wgaDNA from serum/plasma samples stored for up to 20 years, but not subjected to thawing and refreezing, produced the best SNP genotyping results on Affymetrix 500K arrays; however, samples that had been thawed and refrozen showed significant deterioration in call rates compared to never-thawed samples. Numerous studies have investigated ways to preserve low-concentration solutions of DNA that are prone to degradation³⁹ and to improve the genotyping performance of wgaDNA from limited/degraded templates.^{26,28,29,32,40–42} In our hands, an independent whole-genome amplification significantly increased call rates for poorly performing samples, but only by ~2%. We recommend using caution when applying wgaDNA from serum/plasma to high-density SNP arrays because even small discordance rates may lead to an unacceptably high number of genotyping errors. Given the enormous potential of serum banks for genetic testing, forensics, and genome-wide association studies, there is a critical need to further optimize whole-genome amplification methods to improve fidelity when amplifying partially degraded templates.^{43–46}

References

- Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, Struwing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, the SEARCH collaborators, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen C-Y, Wu P-E, Wang H-C, Eccles D, Evans DG, Peto J, Fletcher O, Johnson N, Seal S, Stratton MR, Rahman N, Chenevix-Trench G, Bojesen SE, Nordestgaard BG, Axelsson CK, Garcia-Closas M, Brinton L, Chanock S, Lissowska J, Peplonska B, Nevanlinna H, Fagerholm R, Eerola H, Kang D, Yoo K-Y, Noh D-Y, Ahn S-H, Hunter DJ, Hankinson SE, Cox DG, Hall P, Wedren S, Liu J, Low Y-L, Bogdanova N, Schürmann P, Dörk T, Tollenaar RAEM, Jacobi CE, Devilee P, Klijn JGM, Sigurdson AJ, Doody MM, Alexander BH, Zhang J, Cox A, Brock IW, MacPherson G, Reed MWR, Couch FJ, Goode EL, Olson JE, Meijers-Heijboer H, van den Ouweland A, Uitterlinden A, Rivadeneira F, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Hopper JL, McCredie M, Southey M, Giles GG, Schroen C, Justenhoven C, Brauch H, Hamann U, Ko Y-D, Spurdle AB, Beesley J, Chen X, kConFab, AOCs Management Group, Mannermaa A, Kosma V-M, Kataja V, Hartikainen J, Day NE, Cox DR, Ponder BAJ: Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007, 447:1087–1093
- Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding C-J, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li X-Y, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007, 316:1341–1345
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwork C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME: Global variation in copy number in the human genome. *Nature* 2006, 444:444–454
- Sobrinho B, Brión M, Carracedo A: SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int* 2005, 154:181–194
- Tian C, Hinds DA, Shigeta R, Kittles R, Ballinger DG, Seldin MF: A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. *Am J Hum Genet* 2006, 79:640–649
- Mao X, Bigham AW, Mei R, Gutierrez G, Weiss KM, Brutsaert TD, Leon-Velarde F, Moore LG, Vargas E, McKeigue PM, Shriver MD, Parra EJ: A genomewide admixture mapping panel for Hispanic/Latino populations. *Am J Hum Genet* 2007, 80:1171–1178
- Lovmar L, Syvänen A-C: Multiple displacement amplification to create a long-lasting source of DNA for genetic studies. *Hum Mutat* 2006, 27:603–614
- Telenius H, Carter NP, Bebb CE, Nordenskjöld M, Ponder BAJ, Tunnacliffe A: Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 1992, 13:718–725
- 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
- Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, Lasken RS: Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci USA* 2002, 99:5261–5266
- Hosono S, Faruqi AF, Dean FB, Du Y, Sun Z, Wu X, Du J, Kingsmore SF, Egholm M, Lasken RS: Unbiased whole-genome amplification directly from clinical samples. *Genome Res* 2003, 13:954–964
- Nelson JR, Cai YC, Giesler TL, Farchaus JW, Sundaram ST, Ortiz-Rivera M, Hosta LP, Hewitt PL, Mamone JA, Palaniappan C, Fuller CW: TempliPhi, Φ29 DNA polymerase based rolling circle amplification of templates for DNA sequencing. *BioTechniques* 2002, Suppl:44–47
- Yan J, Feng J, Hosono S, Sommer SS: Assessment of multiple displacement amplification in molecular epidemiology. *BioTechniques* 2004, 37:136–138, 140–143
- Pask R, Rance HE, Barratt BJ, Nutland S, Smyth DJ, Sebastian M, Twells RCJ, Smith A, Lam AC, Smink LJ, Walker NM, Todd JA: Investigating the utility of combining Φ29 whole genome amplification and highly multiplexed single nucleotide polymorphism BeadArray™ genotyping. *BMC Biotechnol* 2004, 4:15
- Lage JM, Leamon JH, Pejovic T, Hamann S, Lacey M, Dillon D, Segraves R, Vossbrinck B, González A, Pinkel D, Albertson DG, Costa J, Lizardi PM: Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. *Genome Res* 2003, 13:294–307
- Barker DL, Hansen MST, Faruqi AF, Giannola D, Irsula OR, Lasken RS, Latterich M, Makarov V, Oliphant A, Pinter JH, Shen R, Sleptsova I, Ziehler W, Lai E: Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. *Genome Res* 2004, 14:901–907
- Bergen AW, Haque KA, Qi Y, Beerman MB, Garcia-Closas M, Rothman N, Chanock SJ: Comparison of yield and genotyping performance of multiple displacement amplification and OmniPlex™ whole genome amplified DNA generated from multiple DNA sources. *Hum Mutat* 2005, 26:262–270
- Paynter RA, Skibola DR, Skibola CF, Buffler PA, Wiemels JL, Smith MT: Accuracy of multiplexed Illumina platform-based single-nucleotide polymorphism genotyping compared between genomic and whole genome amplified DNA collected from multiple sources. *Cancer Epidemiol Biomarkers Prev* 2006, 15:2533–2536
- Lu Y, Gioia-Patricola L, Gomez JV, Plummer M, Franceschi S, Kato I, Canzian F: Use of whole genome amplification to rescue DNA from plasma samples. *BioTechniques* 2005, 39:511–515
- Di X, Matsuzaki H, Webster TA, Hubbell E, Liu G, Dong S, Bartell D, Huang J, Chiles R, Yang G, Shen M, Kulp D, Kennedy GC, Mei R, Jones KW, Cawley S: Dynamic model based algorithms for screening and genotyping over 100K SNPs on oligonucleotide microarrays. *Bioinformatics* 2005, 21:1958–1963
- Jellum E, Andersen A, Lund-Larsen P, Theodorsen L, Ørjasæter H: Experiences of the Janus Serum Bank in Norway. *Environ Health Perspect* 1995, 103(Suppl 3):85–88
- Lettieri C, Moon J, Hickey P, Gray M, Berg B, Hospenthal D: Prevalence of leptospira antibodies in U.S. Army blood bank donors in Hawaii. *Mil Med* 2004, 169:687–690

23. Thomas K, Bannon G, Herouet-Guicheney C, Ladics G, Lee L, Lee S-I, Privalle L, Ballmer-Weber B, Vieths S: The utility of an international sera bank for use in evaluating the potential human allergenicity of novel proteins. *Toxicol Sci* 2007, 97:27–31
24. Rylander-Rudqvist T, Håkansson N, Tybring G, Wolk A: Quality and quantity of saliva DNA obtained from the self-administrated oragene method—a pilot study on the cohort of Swedish men. *Cancer Epidemiol Biomarkers Prev* 2006, 15:1742–1745
25. Sahoo T, Cheung SW, Ward P, Darilek S, Patel A, del Gaudio D, Kang SHL, Lalani SR, Li J, McAdoo S, Burke A, Shaw CA, Stankiewicz P, Chinault AC, Van den Veyver IB, Roa BB, Beaudet AL, Eng CM: Prenatal diagnosis of chromosomal abnormalities using array-based comparative genomic hybridization. *Genet Med* 2006, 8:719–727
26. Montgomery GW, Campbell MJ, Dickson P, Herbert S, Siemering K, Ewen-White KR, Visscher PM, Martin NG: Estimation of the rate of SNP genotyping errors from DNA extracted from different tissues. *Twin Res Hum Genet* 2005, 8:346–352
27. Lovmar L, Fredriksson M, Liljedahl U, Sigurdsson S, Syvänen A-C: Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA. *Nucleic Acids Res* 2003, 31:e129
28. Tzvetkov MV, Becker C, Kulle B, Nürnberg P, Brockmüller J, Wojnowski L: Genome-wide single-nucleotide polymorphism arrays demonstrate high fidelity of multiple displacement-based whole-genome amplification. *Electrophoresis* 2005, 26:710–715
29. Ballantyne KN, van Oorschot RAH, Mitchell RJ: Comparison of two whole genome amplification methods for STR genotyping of LCN and degraded DNA samples. *Forensic Sci Int* 2007, 166:35–41
30. Pompanon F, Bonin A, Bellemain E, Taberlet P: Genotyping errors: causes, consequences and solutions. *Nat Rev Genet* 2005, 6:847–859
31. Gormally E, Hainaut P, Caboux E, Airoldi L, Autrup H, Malaveille C, Dunning A, Garte S, Matullo G, Overvad K, Tjonneland A, Clavel-Chapelon F, Boffetta P, Boeing H, Trichopoulou A, Palli D, Krogh V, Tumino R, Panico S, Bueno-de-Mesquita HB, Peeters PH, Lund E, Gonzalez CA, Martinez C, Dorronsoro M, Barricarte A, Tormo MJ, Quiros JR, Berglund G, Hallmans G, Day NE, Key TJ, Veglia F, Peluso M, Norat T, Saracci R, Kaaks R, Riboli E, Vineis P: Amount of DNA in plasma and cancer risk: a prospective study. *Int J Cancer* 2004, 111:746–749
32. Schoenborn V, Gohlke H, Heid IM, Illig T, Utermann G, Kronenberg F: Sample selection algorithm to improve quality of genotyping from plasma-derived DNA: to separate the wheat from the chaff. *Hum Mutat* 2007, 28:1141–1149
33. Paez JG, Lin M, Beroukhi R, Lee JC, Zhao X, Richter DJ, Gabriel S, Herman P, Sasaki H, Altshuler D, Li C, Meyerson M, Sellers WR: Genome coverage and sequence fidelity of Φ 29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Res* 2004, 32:e71
34. Schneider PM, Bender K, Mayr WR, Parson W, Hoste B, Decorte R, Cordonnier J, Vanek D, Morling N, Karjalainen M, Carlotti CMP, Sabatier M, Hohoff C, Schmitter H, Pflug W, Wenzel R, Patzelt D, Lessig R, Dobrowolski P, O'Donnell G, Garafano L, Dobosz M, de Knijff P, Mevag B, Pawlowski R, Gusmão L, Conceicao Vide M, Alonso Alonso A, García Fernández O, Sanz Nicolás P, Kihlgren A, Bär W, Meier V, Teyssier A, Coquoz R, Brandt C, Germann U, Gill P, Hallett J, Greenhalgh M: STR analysis of artificially degraded DNA—results of a collaborative European exercise. *Forensic Sci Int* 2004, 139:123–134
35. Bergen AW, Qi Y, Haque KA, Welch RA, Garcia-Closas M, Chanock SJ, Vaught J, Castle PE: Effects of electron-beam irradiation on whole genome amplification. *Cancer Epidemiol Biomarkers Prev* 2005, 14:1016–1019
36. Sjöholm MIL, Hoffmann G, Lindgren S, Dillner J, Carlson J: Comparison of archival plasma and formalin-fixed paraffin-embedded tissue for genotyping in hepatocellular carcinoma. *Cancer Epidemiol Biomarkers Prev* 2005, 14:251–255
37. Lasken RS, Egholm M: Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. *Trends Biotechnol* 2003, 21:531–535
38. Bergen AW, Qi Y, Haque KA, Welch RA, Chanock SJ: Effects of DNA mass on multiple displacement whole genome amplification and genotyping performance. *BMC Biotechnol* 2005, 5:24
39. Smith S, Morin PA: Optimal storage conditions for highly dilute DNA samples: a role for trehalose as a preserving agent. *J Forensic Sci* 2005, 50:1101–1108
40. Rook MS, Delach SM, Deyneko G, Worlock A, Wolfe JL: Whole genome amplification of DNA from laser capture-microdissected tissue for high-throughput single nucleotide polymorphism and short tandem repeat genotyping. *Am J Pathol* 2004, 164:23–33
41. Ballantyne KN, van Oorschot RAH, Mitchell RJ, Koukoulas I: Molecular crowding increases the amplification success of multiple displacement amplification and short tandem repeat genotyping. *Anal Biochem* 2006, 355:298–303
42. Ballantyne KN, van Oorschot RAH, Muharam I, van Daal A, Mitchell RJ: Decreasing amplification bias associated with multiple displacement amplification and short tandem repeat genotyping. *Anal Biochem* 2007, 368:222–229
43. Wang G, Maher E, Brennan C, Chin L, Leo C, Kaur M, Zhu P, Rook M, Wolfe JL, Makrigiorgos GM: DNA amplification method tolerant to sample degradation. *Genome Res* 2004, 14:2357–2366
44. Ogino S, Kawasaki T, Brahmandam M, Yan L, Cantor M, Namgyal C, Mino-Kenudson M, Lauwers GY, Loda M, Fuchs CS: Sensitive sequencing method for KRAS mutation detection by pyrosequencing. *J Mol Diagn* 2005, 7:413–421
45. Li J, Harris L, Mamon H, Kulke MH, Liu W-H, Zhu P, Makrigiorgos GM: Whole genome amplification of plasma-circulating DNA enables expanded screening for allelic imbalance in plasma. *J Mol Diagn* 2006, 8:22–30
46. Wang F, Wang L, Briggs C, Sicinska E, Gaston SM, Mamon H, Kulke MH, Zamponi R, Loda M, Maher E, Ogino S, Fuchs CS, Li J, Hader C, Makrigiorgos GM: DNA degradation test predicts success in whole-genome amplification from diverse clinical samples. *J Mol Diagn* 2007, 9:441–451