Evaluating Whole Genome Amplification via Multiply-Primed Rolling Circle Amplification for SNP Genotyping of Samples With Low DNA Yield

Kaisa Silander,[1] Kati Komulainen,[1] Pekka Ellonen,[1] Minttu Jussila,[1] Mervi Alanne,[1] Minna Levander,[1] Päivi Tainola,[1] Kari Kuulasmaa,[2] Veikko Salomaa,[2] Markus Perola,[1] Leena Peltonen,[1] and Janna Saarela[1]

[1]*Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland*
[2]*Department of Epidemiology and Health Promotion, National Public Health Institute, Helsinki, Finland*

The amount of available DNA is often a limiting factor in pursuing genetic analyses of large-scale population cohorts. An association between higher DNA yield from blood and several phenotypes associated with inflammatory states has recently been demonstrated, suggesting that exclusion of samples with very low DNA yield may lead to biased results in statistical analyses. Whole genome amplification (WGA) could present a solution to the DNA concentration-dependent sample selection. The aim was to thoroughly assess WGA for samples with low DNA yield, using the multiply-primed rolling circle amplification method. Fifty-nine samples were selected with the lowest DNA yield (less than 7.5μg) among 799 samples obtained for one population cohort. The genotypes obtained from two replicate WGA samples and the original genomic DNA were compared by typing 24 single nucleotide polymorphisms (SNPs). Multiple genotype discrepancies were identified for 13 of the 59 samples. The largest portion of discrepancies was due to allele dropout in heterozygous genotypes in WGA samples. Pooling the WGA DNA replicates prior to genotyping markedly improved genotyping reproducibility for the samples, with only 7 discrepancies identified in 4 samples. The nature of discrepancies was mostly homozygote genotypes in the genomic DNA and heterozygote genotypes in the WGA sample, suggesting possible allele dropout in the genomic DNA sample due to very low amounts of DNA template. Thus, WGA is applicable for low DNA yield samples, especially if using pooled WGA samples. A higher rate of genotyping errors requires that increased attention be paid to genotyping quality control, and caution when interpreting results.

The amount of available DNA is often a limiting factor in pursuing large-scale genetic analyses of population cohorts or epidemiological study samples, especially when considering old archived materials whose value is often increased by the long follow-up times. Many large study cohorts use blood leucocytes as the DNA source. An association between higher DNA yield from blood and several clinical phenotypes associated with inflammatory states, such as diabetes or metabolic syndrome, has recently been shown (Alanne et al., 2004). The DNA amount-based study sample selection may thus lead to biased results in statistical analysis (Alanne et al., 2004). Whole genome amplification (WGA) of genomic DNA could present a solution to these problems. Several methods have been developed for WGA, including degenerate oligonucleotide-primed polymerase chain reaction (PCR), primer extension preamplification, and multiply-primed rolling circle amplification (MPRCA; reviewed by Hawkins et al., 2002). The first two methods are PCR-based and result in products less than 3 kb in length, while the latter is non-PCR based, using Φ29 DNA polymerase and random hexamer primers, and results in products greater than 10 kb in length. Additional advantages of the multiply-primed rolling circle amplification (MPRCA) approach include the lower error rate of the DNA polymerase and an unbiased amplification of the original genome (Dean et al., 2001; Hawkins et al., 2002).

The genotyping accuracy using DNA samples amplified via MPRCA-WGA has been previously evaluated, using single nucleotide polymorphism (SNP) markers and several different genotyping platforms. These studies showed that MPRCA-WGA leads to a balanced amplification of the genome, with high correlation between genotypes obtained using genomic and WGA DNA samples (Barker et al., 2004; Lovmar et al., 2003; Paez et al., 2004;

Received 14 March 2005; accepted 25 May, 2005.

Address for correspondence: Kaisa Silander, PhD, Department of Molecular Medicine, National Public Health Institute, Biomedicum, Haartmaninkatu 8, 00251 Helsinki, Finland. E-mail: kaisa.silander@ktl.fi

Tranah et al., 2003). In all of these studies the DNA samples used for amplification were of high yield and quality. As these are not samples typically requiring amplification, we aimed to thoroughly assess the use of MPRCA-WGA for samples with low DNA yield, by evaluating the reproducibility of SNP genotypes obtained from these samples. This study design better reflects the real-life situation in the genotyping laboratories. Among 799 DNA samples obtained for a single epidemiological study and isolated from 10 ml of whole blood, 59 samples that had a DNA amount of less than a preset limit of 7.5 µg were identified. Each sample was amplified twice using MPRCA. The genotypes obtained for the two WGA replicates were compared to the genotypes obtained from genomic DNA, typing a total of 24 SNPs on two genotyping platforms. The effect of pooling the WGA DNA products prior to genotyping on genotyping accuracy was also studied. As a comparison, 36 good quality and high yield DNA samples were studied.

Materials and Methods

DNA Isolation and WGA

DNA was isolated from 10 ml of peripheral whole blood samples (EDTA) using a standard phenol-chroloform extraction or salt precipitation. The 59 DNA samples studied were chosen from among 799 samples collected during 1992 as part of the FINRISK study (Vartiainen et al., 2000). The 799 samples were selected from among 5827 samples for a case-cohort study that aims to identify environmental and genetic risk factors predisposing to cardiovascular disease (Evans et al., 2005). All 5827 samples were isolated following storage for between 6 and 11 years at –20°C. The DNA samples were originally quantitated and quality checked with UV-absorbance measurement, and later quantitated during the DNA aliquoting process using PicoGreen dsDNA Quantitation Kit (Molecular Probes, Eugene, OR). The motivation for selecting the 59 samples for WGA was an available DNA amount less than or equal to 7.5 µg following PicoGreen measurement. Thirty-six good quality, high-DNA yield samples were chosen as controls from a set of 52 anonymous DNA samples that are in regular use as laboratory controls in the laboratory.

Genomic DNA (10 ng when available) was amplified using GenomePhi DNA amplification kit (Amersham Biosciences, Piscataway, NJ) as specified in the manufacturer's instructions. WGA samples were cleaned by Spin column chromatography using MicroSpin G-50 Columns (Amersham Biosciences). WGA DNA concentration was measured with PicoGreen. For each of the 59 low-yield DNA samples, a pooled-WGA sample was also made by combining equal amounts of DNA from each of the two WGA replicates originating from the same sample.

SNP Genotyping

For all samples, 16 SNPs were genotyped using the MassARRAY System (Sequenom, San Diego, CA), and eight SNPs were genotyped by allele-specific primer extension on microarrays (Pastinen et al., 2000). For the control DNA samples of high quality, eight additional SNPs were genotyped using the MassARRAY system. These SNPs are part of the MORGAM study, an ongoing association study to evaluate the role of specific candidate genes as cardiovascular disease risk factors (Evans et al., 2005). All WGA DNA samples were genotyped in duplicates, while the genomic DNA was genotyped once for each sample-SNP combination. Genotypes were compared between the two WGA replicates for each sample, the WGA-pool DNA sample and also the genomic DNA sample when available (only 40 of the 59 samples were also genotyped using genomic DNA).

For MassARRAY genotyping, WGA DNA samples were diluted to 2.5 ng/µl, and 4 µl were pipetted onto a 384-microtiter plate, and dried down at 37°C. Similarly, for WGA-pool DNA samples 3.5 µl of 3.0 ng/µl DNA were used, and for genomic DNA samples 2.0 µl of 2.5 ng/µl DNA were used. SNP assays were designed using SpectroDESIGNER (Sequenom), and PCR and extension reactions were done as specified by the manufacturer. Genotypes were automatically called with the SpectroCALLER software (Sequenom), and manually checked as described (Silander et al., 2003). The 16 SNPs were typed in four tetraplexes (tetraplex 1: rs405509, rs1035550, rs710898, rs10908821; tetraplex 2: rs1800792, rs16944, rs1143637, rs1143634; tetraplex 3: rs5960, s1044291, rs2066860, rs1401296; tetraplex 4: hCV9884916, hCV1872266, rs1409433, rs1554286). The good quality, high-DNA yield control DNA samples were typed for the four tetraplex groups above, and for two additional tetraplexes (tetraplex 5: rs970741, rs6030, rs6019, rs762637; tetraplex 6: rs6025, hCV1872237, rs2227819, rs250733). The minor allele frequencies of the SNPs were between .05 and .50. The SNPs were located on the following genes: apoliprotein E, fibrinogen beta chain, fibrinogen gamma chain, interleukin 1 beta, interleukin 10, coagulation factor V, coagulation factor VII, coagulation factor X, protein C, thrombin receptor, thrombin-activable fibrinolysis inhibitor, upstream stimulatory factor 1, forkhead box C2, and thioredoxin-interacting protein.

For allele-specific primer extension on microarrays, a modification of the protocol described by Pastinen et al. (2000) was used. WGA DNA samples were diluted to 2.5ng/µl, and 10 ng were pipetted onto a 96-microtiter plate, and dried at 37°C until all liquid had evaporated. Genomic DNA samples were diluted to 5 ng/µl, and 10 ng of DNA were pipetted onto a 96-microtiter plate, and dried at 37°C. Primers for PCR were designed using the Primer3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/

https://doi.org/10.1375/twin.8.4.368 Published online by Cambridge University Press

primer3_www.cgi; Rozen & Skaletsky, 2000). Allele-specific extension primers were designed using the Oligo Analyzer 3.0 (Integrated DNA Technologies web site; http://scitools.idtdna.com/Analyzer/). The SNP loci for the sample were amplified by PCR and *in vitro* transcribed into ssRNA. The SNP genotype of the RNA sample was determined by the primer extension reaction of the allele-specific oligonucleotides (ASOs) on the microarray surface. Each biallelic SNP is represented by two 5' amine-modified oligonucleotides defining the alleles by their 3' nucleotides on the array surface. ASOs having complete hybridization with sample RNA are extended by a MMLV-RT catalyzed primer extension reaction in optimized conditions, whereas incomplete hybridization shows no or residual primer extension. Microarray slides were manufactured in-house using the Omnigrid microarrayer (GeneMachines, San Carlos, CA) and 20 µM allele-specific oligonucleotides in 0.4 M sodium acetate buffer, pH 9, onto chemically activated microscopic slides. Oligonucleotides attach to the slide surface via 5' amine-modification, which covalently bonds with the aminosilane-coated slide surface (Guo et al., 1994). Each microarray slide contained 80 identical subarrays with ASOs for each SNP printed in duplicates. The 'array of arrays' layout allowed a genotyping capacity of 80 samples per slide with low reaction volumes.

For multiplex PCR, reaction conditions were as follows: 0.4 U of AmpliTaq GOLD (Applied Biosystems, Foster City, CA), 200 µM dNTP, 3.0 mM $MgCl_2$ and 200 nM of each PCR primer in 1X AmpliTaq buffer. One primer of each amplicon was 5' tailed with the T3 RNA polymerase promoter sequence (AATTAACCCTCACTAAAGGGAGA). A 2.0 µl aliquote of PCR product was used as a template for in-vitro transcription by Ampliscribe T3 RNA polymerase (Epicentre Technologies, Madison, WI) utilizing the promoter sequence introduced to the amplicons via the PCR primers. Following transcription, PCR products were degraded by DNAseI treatment to yield a clean single-stranded RNA template to facilitate hybridization to ASOs on the microarray surface. Two microliters of the RNA sample were hybridized on each subarray in 1.6 M NaCl in a humid chamber at 42°C for 20 minutes. After hybridization, slides were washed in a buffer containing 0.5X TE, 0.3 M NaCl and 0.1% Triton X-100, rinsed with cold dH_2O and dried by a stream of air. The primer extension reaction mix (2µl) was immediately applied on each subarray and the slide was incubated at 52°C for 20 minutes in a humid chamber. The primer extension mix consisted of 1X MMLV-RT buffer; 2U MMLV-RT enzyme; 0.5 mM dATP, dGTP, ddATP and ddGTP; 1.0 mM Cy5-dCTP and Cy5-dUTP, in 0.5 mM Trehalose/glycerol. The slides were washed after incubation as after hybridization. Microarray slides were then scanned with a ScanArray4000 laser scanner instrument (PerkinElmer Life and Analytical Sciences, Boston, MA) and fluorescent signals were quantitated by QuantArray 3.0 software. Raw signals were converted to genotypes with the SNPSnapper software developed by Dr Juha Saharinen (National Public Health Institute, Helsinki, Finland; http://www.bioinfo.helsinki.fi/SNPSnapper/default.htm). The following eight SNPs were genotyped in one multiplex using the allele-specific primer extension method: rs4843395, rs7199663, rs9245, rs7211, rs2073658, rs2516839, rs2774279 and one SNP not available in public databases. SNP minor allele frequencies were between .05 and .40. The SNPs were located on the following genes: forkhead box c2, thioredoxin-interacting protein, and upstream stimulatory factor 1.

Sequencing

Two SNPs, rs405509 and rs1143637, showing genotype discrepancies, were reanalyzed in five and six samples respectively, using sequencing as a reference method. PCR products from both WGA replicates and the genomic DNA were directly sequenced by BigDye 3.1 sequencing kit (Applied Biosystems) with a dilution ratio of 1:32, and protocol as recommended by the manufacturer. The sequencing products were separated on ABI3730 DNA Analyzer (Applied Biosystems) and analyzed using Sequencher 4.1.4 (GeneCodes Corporation, Ann Arbor, MI).

DNA Quality Tests

The integrity of the DNA was tested using Agilent 2100 Bioanalyzer, DNA 12000 LabChip kit (Agilent Technologies, Palo Alto, CA). Samples were run according to manufacturer's instructions using 1 µl of the DNA dilution. A total of 12 low-yield DNA samples were studied, 8 in which multiple genotype discrepancies were detected, and 4 in which no genotype discrepancies were detected. For each of the samples, both WGA repeats were also studied.

To control for sample handling errors and as an initial quality check, each of the WGA replicate samples were studied using three autosomal multiallelic markers (*GDB371492*, *D20S448*, and *AC00818_1*), one X chromosome–specific multiallelic marker (*DXS1220*), one Y chromsome–specific locus (*SRY*), and one X chromosome–specific locus (*HTR2C*). The markers were PCR-amplified using fluorescently labeled primers, and the PCR fragments were separated on ABI3730 DNA Analyzer and genotyped with GeneMapper 3.0 (Applied Biosystems). Primer information is available from the authors upon request.

Results

Quality and Reproducibility of SNP Genotyping for WGA Samples

For the high-quality control DNA samples, we typed two WGA replicates of 36 samples on 24 SNPs using the MassARRAY System, and two WGA replicates of 23 samples on eight SNPs using the microarray system. The allele call rate for the WGA DNA samples on the MassARRAY system was generally slightly lower than

https://doi.org/10.1375/twin.8.4.368 Published online by Cambridge University Press

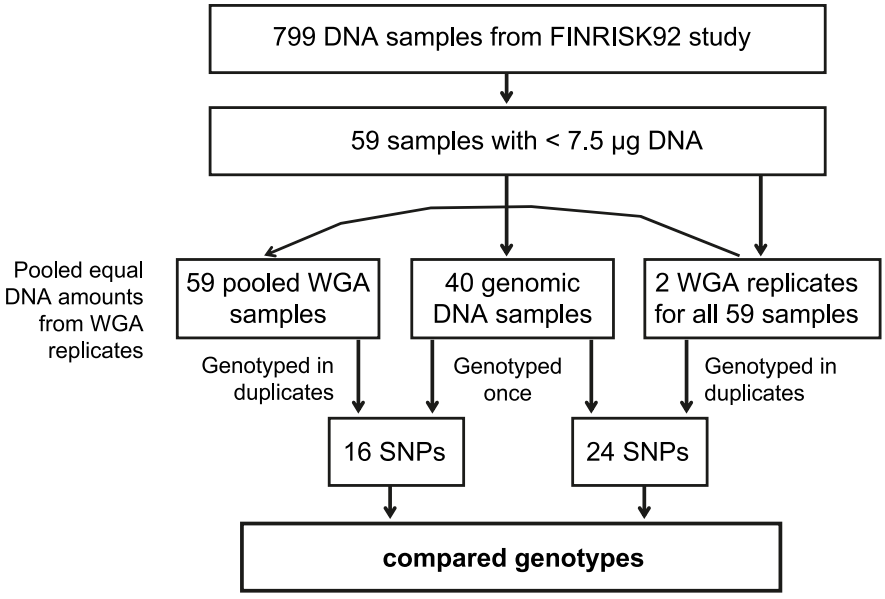

Figure 1

Flow chart of the study design, aimed to evaluate whole genome amplification of DNA samples with low DNA yield.

for the genomic DNA samples, and one of the 24 SNPs failed completely in the WGA samples, possibly due to a technical problem. The allele call rate for the WGA samples on the microarray system was 99%. The genotypes obtained from the WGA and genomic DNA samples were in perfect agreement (~700 successful WGA-genomic DNA genotype comparisons), and no substantial difference in the quality of the genotype profiles could be observed.

A flow chart detailing the study design for the low DNA yield samples is shown in Figure 1. The individual WGA replicates of the 59 samples with low DNA yield were tested by typing a total of 24 SNPs, 16 using the MassARRAY System and 8 using the microarray system. Each replicate WGA sample was genotyped in duplicate. A majority of the samples were also typed once using genomic DNA. The allele call rate was systematically slightly lower (5% to 10%) for the WGA samples when compared to the allele call rate for these SNPs in the full FINRISK92 case-cohort study (less than 95%). Multiple genotype discrepancies were detected for 13 of the 59 samples (Table 1), both between WGA replicates and between WGA product and genomic DNA. Eight of these discrepancies were identified among duplicates of the same replicate WGA sample. In addition, a single genotype discrepancy was observed for four samples. In most of the discrepancies, heterozygote genotypes in genomic DNA appeared as homozygote genotypes in WGA DNA. This suggests, as expected, that genotyping errors of WGA DNA are mostly caused by allele dropout. For 11 instances in which discrepancies were identified between WGA replicates of the same genomic DNA sample, the PCR products were directly sequenced as a reference genotyping method. Seven of the discrepancies were reproduced using direct sequencing (an example shown in Figure 2), while for four other samples, sequencing gave identical genotypes between WGA replicates. Mixed sequences and unexpected additional sequences were identified in

Table 1

Concentration of DNA Used for WGA and Distribution of Multiple Genotype Discrepancies for 13 Problematic Samples Genotyped on a Total of 24 SNPs, Comparing Genotypes Obtained From WGA Replicates of the Same Sample and Genomic DNA

	# Discrepant SNPs			DNA concentration
Sample	MassARRAY	Microarray	Total	ng/µl
1	2		2	7
2	2		2	11
3		2	2	90
4	1	2	3	31
5	1	2	3	> 0
6	3	2	5	1
7	5		5	13
8	4	2	6	23
9	6		6	30
10	5	2	7	3
11	7	1	8	18
12	8	1	9	> 0
13	5	6	11	2

Note: Four samples showing discrepancies following pooling of the WGA replicates are in boldface. DNA concentrations were measured with PicoGreen prior to WGA. For two samples, DNA concentrations were just slightly above zero.

https://doi.org/10.1375/twin.8.4.368 Published online by Cambridge University Press

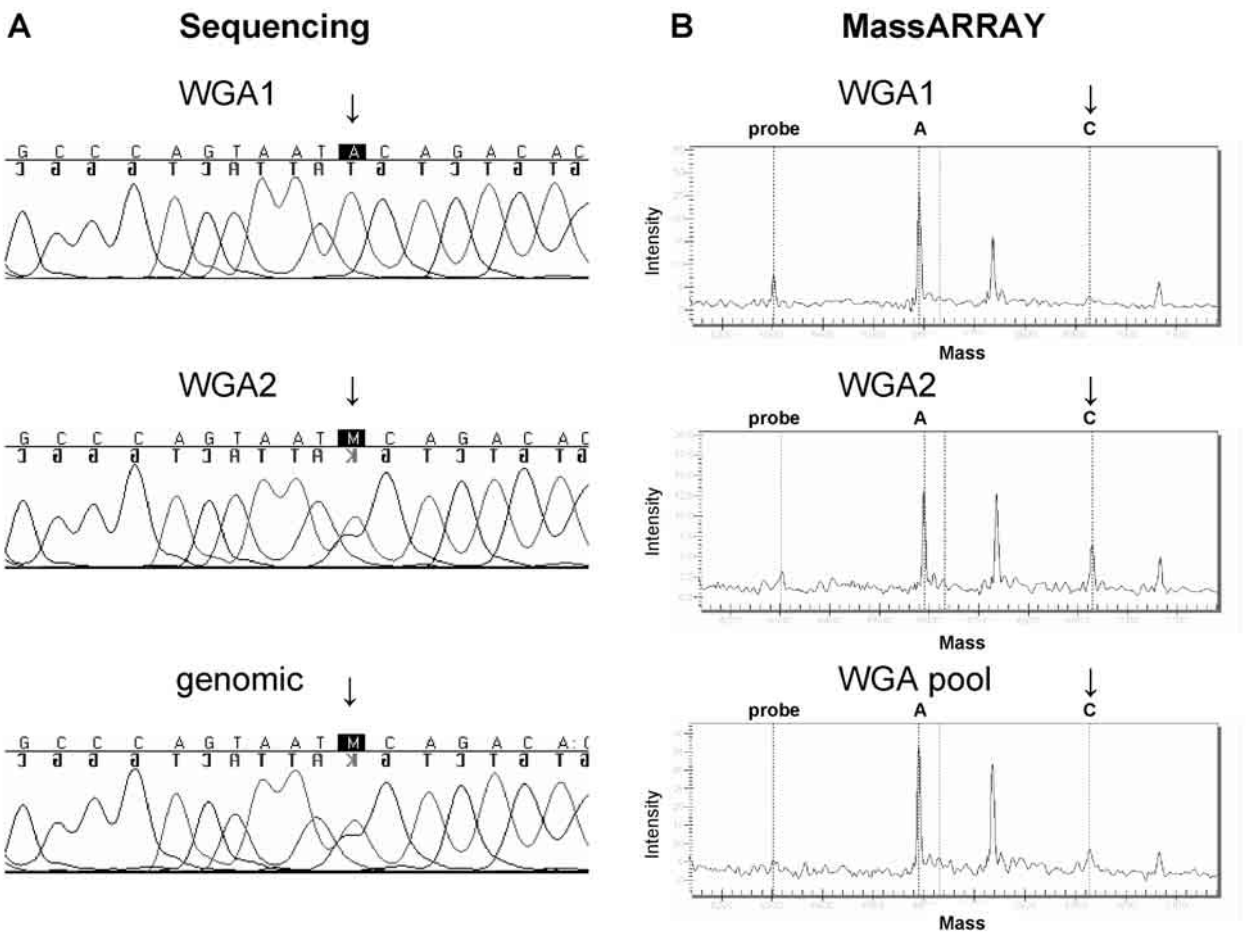

Figure 2

Example for discrepant genotypes between WGA replicates of the same sample for SNP rs405509, using either (A) direct sequencing or (B) primer extension with the MassARRAY system. The genomic DNA and the WGA-pool sample are shown for comparison.

four of the discrepant samples, suggesting unspecific sequence amplification in these WGA samples.

The allele ratio for heterozygote genotypes was more variable for the WGA samples when compared to the genomic DNA samples. An example of one multiplex for the MassARRAY system is detailed in Table 2. The skewed allele ratio in heterozygotes was most notable for the 13 problematic WGA samples, as illustrated in Figure 3. For these samples, many genotypes were discarded due to allele calling problems, such as low signal intensities or genotypes escaping from genotype clusters.

Analysis of Genomic DNA and WGA DNA Quality

The DNA amplification efficiency of the WGA samples, as assessed by PicoGreen measurements, was similar for the 59 low-yield DNA samples and for the 36 control samples. For the 59 low-yield DNA samples, analysis evaluating the sizes of the original

Table 2

Comparison of Allele Ratio of Heterozygote Genotypes for WGA and Genomic DNA samples, the MassARRAY System

		Heterozygote allele ratio		
SNP	DNA type	*M* ± *SD*	median	minimum–maximum
rs1800792	genomic	1.17 ± 0.15	1.17	0.78 – 1.75
	WGA	1.2 ± 0.47	1.13	0.43 – 3.63
rs16944	genomic	0.92 ± 0.15	0.9	0.61 – 1.43
	WGA	1.44 ± 0.62	1.3	0.45 – 3.26
rs1143637	genomic	1.56 ± 0.21	1.56	0.96 – 2.33
	WGA	1.56 ± 0.43	1.58	0.63 – 2.62
rs1143634	genomic	1.26 ± 0.17	1.27	0.8 – 1.79
	WGA	1.28 ± 0.67	1.08	0.55 – 3.55

Note: Genomic DNA based on heterozygotes identified among 373 samples; WGA DNA based on two WGA replicates of 59 low yield DNA samples, genotyped twice.

https://doi.org/10.1375/twin.8.4.368 Published online by Cambridge University Press

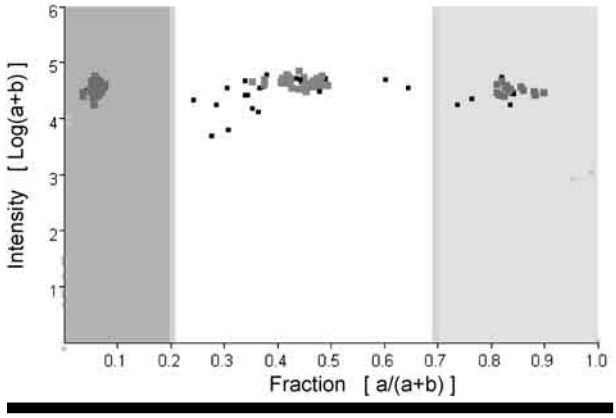

Figure 3

Microarray genotyping for SNP rs2073658 for 59 WGA DNA samples. The 13 samples that show multiple genotype discrepancies are shown as black dots, while the 3 genotype clusters for the other 46 samples are shown as grey dots. Two dots are present for each sample since genotyping was done in duplicates.

genomic DNA molecules did not reveal differences between the samples showing multiple discrepancies compared to the samples showing no discrepancies. All the genomic DNA samples analyzed were above the optimal measuring range of the DNA 12,000 chip, but we saw only high molecular weight DNA molecules, larger than the largest size standard (17.0 kb) and no degradation products. The size range of the WGA amplification products was 1.5kb to 20.0 kb with most of the products in the 7.0 to 9.0 kb range. The size range did not differ between the problem samples and those samples showing no discrepancies.

No correlation was found between the DNA amount, concentration or purity (based on UV-absorbance measurement at 260 nm and 280 nm) and the presence/absence of genotype discrepancies. However, the allele call rate was correlated with the DNA concentration of the genomic DNA serving as the template for the WGA reaction: higher starting DNA concentrations showing higher genotype call rates. The genomic DNA concentration (based on PicoGreen measurement) for the samples with multiple genotype discrepancies was variable, as shown in Table 1. Details on the DNA extraction method, yield and purity for the whole FINRISK92 study and for the 59 samples tested in the study are shown in Table 3.

Analysis of three autosomal and one X-chromosomal multiallelic markers, and two sex chromosome–specific loci revealed discrepancies between the two WGA replicates for five of the 59 samples. An extra allele was identified for one sample, allele dropout for two other samples, and allelic imbalance for one sample. In addition, for one sample successful multiallelic PCR products were not able to be obtained from the WGA repeats, suggesting that no or only residual amplification of the genomic DNA had occurred. All five samples also showed multiple SNP genotype discrepancies when comparing WGA and genomic DNA. Among the additional eight samples showing multiple SNP genotyping discrepancies, seven did not work or worked poorly in multiallelic PCR for one of the WGA replicates. In contrast, 38 out of 42 samples with no SNP genotype discrepancies worked well in multiallelic PCR and the genotypes were fully correlated.

Pooling WGA Products Prior to Genotyping

Next, equal amounts of DNA were pooled from the two WGA replicates for each of the 59 low DNA yield samples into a single tube prior to genotyping. A total of 16 MassARRAY SNPs were genotyped in duplicates using the pooled WGA samples (Figure 1).

Table 3

Comparison of Genomic DNA Characteristics Between the Whole FINRISK92 Sample Set (n = 5827) and the Samples With a DNA Amount Less Than 7.5 µg Chosen for WGA Amplification

	FINRISK 92	FINRISK 92 subsample	FINRISK92 for WGA
N	5722	105	55
Quantity of DNA obtained			
Mean (µg) [*SD*]	271.4 [195.4]	158.0 [132.1]	81.2 [67.8]
Min (µg)	10	11	10
Max (µg)	3204	736	301
Mean concentration (ng/µl) [*SD*]	542.7 [390.8]	316.0 [264.1]	162.4 [135.6]
Purity			
Mean [*SD*]	1.833 [0.32]	1.374 [0.16]	1.729 [0.33]
Extraction method	Phenol-chloroform	Salt precipitation	both
Storage time (–20 °C)	6–11 years	6–11 years	—

Note: DNA quantity and purity obtained from UV-absorbance measurement at 260 and 280 nm immediately following DNA extraction, while the actual selection of samples with DNA amount less than 7.5 µg was based on an additional PicoGreen measurement of diluted sample, which is more accurate.

FINRISK92 subamples were isolated in a different laboratory.

FINRISK92 WGA: Data available for 55 of 59 samples.

https://doi.org/10.1375/twin.8.4.368 Published online by Cambridge University Press

The genotyping success rate for the pooled WGA samples was very high, greater than 98%, and the duplicate genotypes obtained for each of the pooled samples were fully consistent. In addition, the number of genotype discrepancies between WGA and genomic DNA samples was markedly reduced, as shown in Table 4. The four samples that still showed genotype discrepancies were ones in which the original number of discrepancies were 3, 5, 9 and 11. Six of the seven genotype discrepancies were such that a homozygote genotype in the genomic DNA sample was heterozygote in the pooled WGA sample. These four samples had a very low DNA concentration as measured from the stock genomic DNA with PicoGreen (less than or equal to 2 ng/μl; Table 1), but the WGA DNA yield did not differ from the rest of the samples.

Table 4

Comparison of WGA-Genomic Genotype Discrepancies Obtained for Single WGA Samples and Pooled WGA Samples

	# samples with discrepancies	Total # discrepancies	# Unique discrepancies
WGA single reactions	12/40 (30%)	97/3148 (3.1%)	60/847 (7.1%)
WGA pool	4/40 (10%)	12/1125 (1.1%)	7/572 (1.2%)

Note: Genotypes for genomic DNA available only for 40 out of 59 samples.
Total # of discrepancies counted from 2 genotyping attempts for each replicate WGA-sample (i.e., a total of 4 genotyping attempts per sample for WGA single reactions, and 2 genotyping attempts for WGA pool), and one genotyping attempt for genomic DNA sample.
unique discrepancies: each SNP-sample comparison counted once.

Discussion

We have here evaluated whether MPRCA-WGA is applicable for the amplification of low DNA yield samples by analyzing a subset of 59 DNA samples out of a total study sample of 799 DNAs, collected for the population cohort of MORGAM (MORGAM study http://www.ktl.fi/morgam/index.html). This subset of samples represented the lowest seventh percentile and had an initial DNA amount of less than 7.5 μg. Twenty-four SNPs were typed on one of two genotyping platforms, the MassARRAY System and the allele-specific primer extension on microarrays (Pastinen et al., 2000). The results show that MPRCA-WGA works well for the majority of samples (n = 46). For these samples, reproducibility of genotypes was very high and the obtained genotypes were in perfect agreement with those obtained for the genomic DNA, when available. However, for 13 samples, multiple genotype discrepancies were observed. The genotype discrepancies were identified between genotyping repeats of the same WGA sample, between the two replicate WGA samples, and between the WGA sample and the genomic DNA. No correlation was found between DNA purity or concentration and the absence/presence of genotype discrepancies. For the majority of samples showing discrepancies, there were sufficient amounts of DNA template for WGA. Further, no difference was found in the integrity or quantity of the WGA DNA and the presence/absence of genotyping discrepancies. This suggests that such problematic samples cannot be identified a priori based on the characteristics typically determined for extracted DNA samples, or based on WGA DNA quantity.

We tested if genotyping reproducibility could be improved by combining two replicate WGA reactions of the same sample, and using the pooled DNA as a template for PCR and genotyping (Rook et al., 2004). Pooling the WGA products reduced the number of samples showing genotype discrepancies by 67%, and markedly reduced the number of discrepancies. As there is a degree of randomness in the WGA process, the pooled WGA samples may have a better representation of the genome and thus the genotyping errors will be minimized. The results of this study are in agreement with the study by Rook et al. (2004), in which WGA was used to amplify minute amounts of DNA extracted from laser capture microdissection cells, and pooling the WGA products significantly reduced genotyping inconsistencies (Rook et al., 2004). Interestingly, six of the seven discrepancies that were present following pooling were of the nature of homozygote genotypes in the genomic DNA and a heterozygote genotype in the WGA sample, suggesting a possible allele dropout not in the WGA, but in the genomic DNA sample due to the very low amounts of DNA template (less than 2 ng/μl concentration in undiluted stock DNA tube). For such samples, increasing the DNA template amount for the WGA reaction might lead to better quality WGA DNA.

To summarize, WGA should be considered when doing large-scale epidemiological studies to eliminate bias from exclusion of samples with low amounts of DNA. MPRCA-WGA is applicable for low DNA yield samples, especially if pooling at least two WGA replicates for each sample. A higher rate of genotyping errors requires that increased attention should be given to the quality control of genotyping. An effort could also be made to identify and exclude problematic WGA samples by studying a small amount of polymorphisms before advancing to high throughput genotyping, paying special attention to samples with extremely low amounts of DNA. The possible increased rate of genotyping errors should be accounted for when doing statistical analysis using genotype data from the WGA of low yield DNA samples.

Acknowledgments

The study was supported by a grant number 53646 from the Finnish Academy and grants from the Sigrid Juselius Foundation, and the Jenny and Antti Wihuri Foundation. Mervi Alanne was supported by a grant from Aarne Koskelo Foundation, and Kati Komulainen was supported by a grant from the Jenny

and Antti Wihuri Foundation. The GenomEUtwin project is supported by the European Commission under the program 'Quality of Life and Management of the Living Resources' of the 5th Framework Programme (No. QLG2-CT-2002-01254).

References

Alanne, M., Salomaa, V., Stengård, J., Reunanen, A., Saarela, J., Peltonen, L., & Perola, M. (2004). DNA extraction yield is associated with several phenotypic characteristics: Results from two large population surveys. *Journal of Thrombosis and Haemostasis, 2,* 2069–2071.

Barker, D. L., Hansen, M. S., Faruqi, A. F., Giannola, D., Irsula, O. R., Lasken, R. S., Latterich, M., Makarov, V., Oliphant, A., Pinter, J. H., Shen, R., Sleptsova, I., Ziehler, W., & Lai, E. (2004). Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. *Genome Research, 14,* 901–907.

Dean, F. B., Nelson, J. R., Giesler, T. L., & Lasken, R. S. (2001). Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Research, 11,* 1095–1099.

Evans, A., Salomaa, V., Kulathinal, S., Asplund, K., Cambien, F., Ferrario, M., Perola, M., Peltonen, L., Shields, D., Tunstall-Pedoe, H., & Kuulasmaa, K., for the MORGAM Project. (2005). MORGAM (an international pooling of cardiovascular cohorts). *International Journal of Epidemiology, 34,* 21–27.

Guo, Z., Guilfoyle, R. A., Thiel, A. J., Wang, R., & Smith, L. M. (1994). Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Research, 22,* 5456–5465.

Hawkins, T. L., Detter, J. C., & Richardson, P. M. (2002). Whole genome amplification: Applications and advances. *Current Opinion in Biotechnology, 13,* 65–67.

Lovmar, L., Fredriksson, M., Liljedahl, U., Sigurdsson, S., & Syvanen, A. C. (2003). Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA. *Nucleic Acids Research, 31,* e129.

Paez, J. G., Lin, M., Beroukhim, R., Lee, J. C., Zhao, X., Richter, D. J., Gabriel, S., Herman, P., Sasaki, H., Altshuler, D., Li, C., Meyerson, M., & Sellers, W. R. (2004). Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Research, 32,* e71.

Pastinen, T., Raitio, M., Lindroos, K., Tainola, P., Peltonen, L., & Syvänen, A. C. (2000). A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Research, 10,* 1031–1042.

Rook, M. S., Delach, S. M., Deyneko, G., Worlock, A., & Wolfe, J. L. (2004). Whole genome amplification of DNA from laser capture-microdissected tissue for high-throughput single nucleotide polymorphism and short tandem repeat genotyping. *American Journal of Pathology, 164,* 23–33.

Rozen, S., & Skaletsky, H. J. (2000). Primer3 on the WWW for general users and for biologist programmers. In S. Krawetz & S. Misener (Eds.), *Bioinformatics methods and protocols: Methods in molecular biology* (pp. 365–386). Totowa, NJ: Humana Press.

Silander, K., Axelsson, T., Widen, E., Dahlgren, A., Palotie, A., & Syvanen, A. C. (2003). Analysis of genetic variation in the GenomEUtwin project. *Twin Research, 6,* 391–398.

Tranah, G. J., Lescault, P. J., Hunter, D. J., & De Vivo, I. (2003). Multiple displacement amplification prior to single nucleotide polymorphism genotyping in epidemiologic studies. *Biotechnology Letters, 25,* 1031–1036.

Vartiainen, E., Jousilahti, P., Alfthan, G., Sundvall, J., Pietinen, P., & Puska, P. (2000). Cardiovascular risk factor changes in Finland, 1972–1997. *International Journal of Epidemiology, 29,* 49–56.

https://doi.org/10.1375/twin.8.4.368 Published online by Cambridge University Press