Short Communication

Accuracy of Multiplexed Illumina Platform-Based Single-Nucleotide Polymorphism Genotyping Compared between Genomic and Whole Genome Amplified DNA Collected from Multiple Sources

Randi A. Paynter,[1] Danica R. Skibola,[1] Christine F. Skibola,[1] Patricia A. Buffler,[1] Joseph L. Wiemels,[2] and Martyn T. Smith[1]

[1]School of Public Health, University of California, Berkeley, California and [2]Department of Epidemiology and Biostatistics, University of California, San Francisco, California

Abstract

Association studies designed to identify the genetic determinants underlying complex disease increasingly require sustainable high-quality DNA resources for large-scale single-nucleotide polymorphism (SNP) genotyping. Recent studies have shown that genomic DNA (gDNA) suitable for SNP genotyping can be obtained from buccal cells and from dried blood spots on Guthrie cards. Further, successful SNP genotyping has been done using the reaction product of multiple displacement amplification of gDNA. We evaluated genotype consistency on the Illumina genotyping platform for 717 to 1,744 SNP loci between replicate samples of gDNA and whole genome amplified DNA (wgaDNA) from a variety of sources. Nine healthy adults provided peripheral blood via venipuncture and buccal cells by mouth rinse. DNA was also obtained from urothelial cells in urine samples from five of the nine subjects. gDNA was extracted from all samples, wgaDNA was generated from each gDNA, and all samples were genotyped. To assess SNP genotyping accuracy of DNA obtained from dried blood spots, gDNA was extracted, amplified, and genotyped from peripheral blood samples and paired Guthrie card samples were obtained from eight childhood leukemia patients. Call rates and replicate concordances for all sample types, regardless of amplification, were >97%, with most sample types having call rates and replicate concordances >99%. Using the gDNA from blood samples as the reference for concordances calculated for all other sample types, we observed concordances >98% regardless of sample type or amplification. We conclude that highly multiplexed Illumina genotyping may be done on gDNA and wgaDNA obtained from whole blood, buccal samples, dried blood spots on Guthrie cards, and possibly even urine samples, with minimal misclassification. (Cancer Epidemiol Biomarkers Prev 2006;15(12):2533–6)

Introduction

Association studies designed to identify the genetic determinants underlying complex diseases increasingly require a variety of sustainable high-quality DNA resources for large-scale genotyping of single-nucleotide polymorphisms (SNP). Recent studies have shown that genomic DNA (gDNA) suitable for SNP genotyping can be obtained from buccal cells and from dried blood spots on Guthrie cards, in addition to well-established sources for gDNA, such as whole blood and buffy coat samples (1, 2). gDNA also has been extracted and subsequently genotyped from less concentrated sources, such as urine (3-6).

Successful SNP genotyping has been done using the reaction products from whole genome amplification (WGA) of gDNA (7-12). In particular, two WGA methods have produced successful results for SNP genotyping. Multiple displacement amplification (MDA) is a rolling circle amplification, which uses φ29 DNA polymerase and random exonuclease-resistant primers (13). A second method, OmniPlex WGA, involves generating libraries containing fragmented DNA and is especially well suited for amplification of low-concentration gDNA or gDNA that may have undergone some degradation.[3]

Several studies have compared whole genome amplified DNA (wgaDNA) with gDNA, genotyped on the highly multiplexed Illumina BeadArray platform (14, 15). Barker et al. (16) extracted five individual gDNA samples from Centre d'Etude du Polymorphisme Humain (CEPH) cell lines. WGA was done on the gDNA samples using both the MDA and OmniPlex methods. For the 2,320 SNPs evaluated, they observed 99.9% concordance between gDNA- and wgaDNA-derived genotypes for both methods of WGA. Pask et al. (17) made 77 gDNA to wgaDNA (MDA) comparisons, observing 98.8% genotype concordance for 345 SNP markers. A recent study from the International Multiple Sclerosis Genetics Consortium used MDA for generating wgaDNA from gDNA. Among 96 comparisons (92 gDNA to wgaDNA and 4 wgaDNA to wgaDNA), there was >99.9% concordance between duplicate samples for 4,506 SNP markers (18). Based on reported results to date, high-quality wgaDNA does well on the Illumina genotyping platform.

To our knowledge, no reports to date have validated the consistency between DNA genotypes from blood and buccal

[3] http://www.rubicongenomics.com.

cell sources on the Illumina genotyping platform. In this study, we report genotype concordances between blood gDNA and buccal gDNA and their corresponding concordances with wgaDNA. Further, we report genotype success and concordances for DNA extracted from urine samples and for wgaDNA derived from dried blood spots on Guthrie cards.

Materials and Methods

Sample Collection and Preparation. DNA was extracted from fresh peripheral blood obtained from nine healthy adults using the QIAamp DNA Blood Midi kit (Qiagen, Inc., Valencia, CA) according to the manufacturer's instructions. Buccal cells were collected from all nine subjects by scraping the inside of the cheek with a firm toothbrush and rinsing in 10 mL of distilled water followed by a second rinse with 10 mL of distilled water. Both rinses were subsequently added to 5 mL of 99% isopropanol. Additional buccal cells were collected from three of the subjects by swishing orally 10 mL of Original Mint Scope Mouthwash (Proctor & Gamble, Cincinnati, OH) 10 to 20 times, gently scraping the insides of the cheeks with teeth, and expectorating into a 50 mL conical tube. DNA was extracted from buccal cells collected by both methods using Gentra Puregene DNA Isolation kit for mouthwash samples (Gentra Systems, Inc., Minneapolis, MN) according to the manufacturer's instructions. All subjects also provided 50 mL of freshly voided urine, but gDNA was detected from only five of the nine extractions. Each sample was centrifuged at $1,015 \times g$ for 10 min, and the pellet was washed twice with 2 mL $1\times$ PBS with each wash followed by centrifugation at $1,015 \times g$ for 10 min. Cells were resuspended in 1 mL $1\times$ PBS and poured into a sterile 15 mL conical tube. The manufacturer's 2003 preliminary protocol for Gentra Puregene DNA purification from 8 to 10 mL of urine was done beginning with addition of cell lysis solution, scaling up all volumes $10\times$ before a final rehydration of the DNA in 75 μL solution. Peripheral blood from eight childhood leukemia patients was collected and stored at $-80°C$ as part of the Northern California Childhood Leukemia Study. DNA was extracted from thawed 50 μL aliquots using the QIAamp DNA Blood Mini kit (Qiagen) according to the manufacturer's instructions. The leukemia patients' paired Guthrie cards were obtained from the California Department of Health Services. Approximately one-eighth (0.22 cm^2) of the dried blood spot was excised and DNA was extracted using the manufacturer's 2003 protocol for QIAamp DNA Blood Mini kit (Qiagen) dried blood spot protocol. All DNA concentrations were determined using the PicoGreen dsDNA Quantitation kit (Molecular Probes, Inc., Eugene, OR) according to the manufacturer's instructions. DNA (20-50 ng) obtained from childhood leukemia patients' peripheral blood and blood, buccal cells, and urine from volunteer subjects were whole genome amplified using the GenomiPhi DNA Amplification kit (Amersham Biosciences, Piscataway, NJ) according to the manufacturer's instructions. DNA (20 ng) obtained from dried blood spots on Guthrie cards was whole genome amplified using the Repli-g kit (Qiagen) according to the manufacturer's instructions. Both GenomiPhi and Repli-g use MDA technology with random hexamers and ϕ29 polymerase. The study protocol was approved by the Institutional Review Boards of the University of California (Berkeley, CA) and all collaborating institutions, and written informed consent was obtained for all participating subjects.

Genotyping. We attempted genotyping assays for 1,920 SNPs in two oligonucleotide pools (768-plex and 1,152-plex) using the Illumina genotyping platform, with 250 ng of each sample DNA per well and concentrations adjusted to 50 ng/μL. Genotype calls were made using the GenCall version 6.2.0.4 (Illumina, Inc., San Diego, CA) software package. Sample types were analyzed in the following groups: blood gDNA, blood wgaDNA (including wgaDNA from Guthrie cards), buccal gDNA, buccal wgaDNA, and urine DNA (gDNA and wgaDNA were analyzed together due to small sample size). Summary files were generated using the GTS Reports version 5.1.2.0 (Illumina) software package. To confirm and refine clusters, we genotyped duplicate gDNA samples from six CEPH families for which cell line-derived DNA from a brother and sister sibship, their parents, and both paternal and maternal sets of grandparents was available (Coriell Institute, Camden, NJ). Based on clustering failures from CEPH genotyping, we eliminated 51 loci from the 768-plex oligonucleotide pool and 55 loci from the 1,152-plex oligonucleotide pool. Additionally, 70 loci from the 1,152-plex oligonucleotide pool were not evaluated for the CEPH samples, so we eliminated those loci from the analysis of the test samples. Samples with a locus-specific quality measure (GenCall score) <0.1 were reclassified as ''missing'' at that locus (16).

Analysis. Concordances between replicate samples and between paired childhood leukemia peripheral blood and Guthrie card samples were calculated as the number of concordant pairs divided by the number of successfully genotyped pairs. Concordances between comparison samples and blood gDNA were calculated by comparing each of the

Table 1. Call rates and concordances for tissue sources of gDNA and wgaDNA from nine healthy adults

Sample type	Volunteers (n)	Assays run	Assays analyzed	Call rate (%)*	Replicate concordance (%)	Blood gDNA concordance (%)
Blood						
gDNA	9	28,416	25,850	99.88	100.00	NA
wgaDNA	9	28,416	25,850	99.23	99.66	99.65
Buccal (isopropanol)[†]						
gDNA	9	28,416	25,850	97.75	99.35	99.62
wgaDNA	9	28,416	25,850	99.70	99.77	99.78
Buccal (Scope)[‡]						
gDNA	3	6,912	6,162	99.55	99.97	99.95
wgaDNA	3	6,912	6,162	99.20	97.17	98.49
Urine						
gDNA[§]	2	3,456	3,081	99.90	100.00	99.61
wgaDNA	5	11,520	10,270	99.69	99.82	99.01

Abbreviation: NA, not applicable.
*Call rate calculated using ''assays analyzed'' as the denominator.
[†]Buccal cells collected by orally swishing with distilled water and combining 4:1 with 99% isopropanol.
[‡]Buccal cells collected by orally swishing with Scope mouthwash.
[§]gDNA was assayed for two individuals, but duplicate samples were produced from only one individual. Therefore, one individual was not evaluated for replicate concordance but was included in evaluation of blood gDNA concordance.

Table 2. Call rates and concordances for matched samples of DNA from eight childhood leukemia patients

DNA source*	Assays run	Assays analyzed	Call rate (%)[†]	Blood gDNA concordance (%)
Peripheral blood	6,144	5,736	98.68	NA
Guthrie card	6,144	5,736	99.93	99.31

*DNA from one peripheral blood sample was sufficiently concentrated not to require whole genome amplification; all other peripheral blood and Guthrie card samples required amplification.
[†]Call rate calculated using ''assays analyzed'' as the denominator.

duplicate comparison samples with each of the duplicate blood gDNA samples, averaged for up to four comparisons. The concordance calculated for each individual at each locus was weighted according to the number of genotypes successfully scored for comparison.

Results and Discussion

Call Rates. SNPs from all 22 autosomal chromosomes and the X chromosome were interrogated in the custom 768-plex oligonucleotide pool, whereas SNPs in the 1,152-plex oligonucleotide pool provided by Illumina were limited to chromosome 6. For loci not eliminated from analyses (717 SNPs from the 768-plex and 1,027 SNPs from the 1,152-plex), we observed high call rates (>97%) among all sample types (Tables 1 and 2). We note that two wells had very low call rates [<85%, as used by Reich et al. (11) for their exclusion criterion]: subject 1, gDNA buccal (isopropanol), 51.8% and subject 5, wgaDNA blood, 82.3%. Four wells had moderately low call rates [<98%, as used by the International Multiple Sclerosis Genetics Consortium for their exclusion criterion (18)]: childhood leukemia patient F, wgaDNA peripheral blood, 93.4%; subject 3, gDNA buccal (isopropanol), 97.1%; subject 3, wgaDNA buccal (Scope), 97.4%; and childhood leukemia patient C, wgaDNA peripheral blood, 97.9%. All subjects had a second aliquot of each gDNA and wgaDNA sample genotyped. The duplicate aliquot call rates were consistently high, suggesting that the failures observed were well specific and not DNA specific: subject 1, gDNA buccal (isopropanol) duplicate, 99.7%; subject 5, wgaDNA blood duplicate, 99.5%; subject 3, gDNA buccal (isopropanol) duplicate, 99.4%; and subject 3, wgaDNA buccal (Scope) duplicate, 99.5%. Because the peripheral blood samples from childhood leukemia patients were not genotyped in duplicate, we could not determine whether the failure was well specific or due to poor DNA concentration or quality.

Concordance. For samples provided by the nine subjects, we used the genotype calls of gDNA obtained from blood samples as the reference for concordances calculated for all other sample types. wgaDNA from blood samples had 99.7% concordance with their matched samples of gDNA from blood. Buccal samples were collected using the distilled water/isopropanol method for all nine subjects, and additional buccal samples were collected using the Scope method for three of the subjects. For gDNA buccal samples, we observed a concordance of 99.6% comparing the isopropanol-collected samples with their matched gDNA from blood and >99.9% concordance for the Scope-collected buccal samples. The estimated concordance for wgaDNA buccal samples with their matched gDNA blood samples for each collection method was 99.8% (isopropanol) and 98.5% (Scope).

We obtained urine samples with urothelial cells containing DNA from five of the nine subjects. Two samples were sufficiently concentrated to genotype gDNA from urine, resulting in 99.6% concordance with gDNA from the matched blood sample. wgaDNA obtained from urine had a similar concordance of 99.0% to the gDNA from the matched blood sample. Because subjects' DNA sample types were plated in

duplicate, we were able to estimate the concordance between replicates in addition to the concordance to gDNA obtained from blood. All replicate concordances were >99%, except for wgaDNA from buccal cells obtained using the Scope method. We note that one wgaDNA (Scope) from subject 3 had a low call rate (97.4%) and exerted a strong influence on the replicate concordance estimate due to being one of only three subjects evaluated for this sample type. For the eight childhood leukemia samples, we observed 99.1% concordance between genotype calls from DNA obtained from paired peripheral blood and Guthrie card samples.

Cluster Validation. In addition to genotyping the samples described above, we also genotyped CEPH samples in duplicate for the 768-plex reaction: eight samples from each family, representing chromosomes from three generations. Representatives from Illumina provided us with similar CEPH genotyping data for the 1,152-plex reaction. Analysis software from Illumina allows the user to specify parent-child and trio information, if known. We found that using the CEPH samples and including the multigenerational parent-child relationship information aided in determining the clustering patterns at each locus. Additionally, we identified several loci in our custom 768-plex oligonucleotide pool that appeared to cluster clearly into three well-spaced groups but had heritability incompatibilities when evaluated in the CEPH samples and were thus excluded from analyses. We concur with the recommendation of Illumina to validate cluster patterns using heritability information, such as that available from the CEPH samples.

Whole Genome Amplification. Our data support earlier reports of successful genotyping of wgaDNA using Illumina technology (10, 11, 16, 18). Whole genome MDA of gDNA extracted from blood (11, 18) and transformed human cell lines (16-18) previously has been evaluated for Illumina genotyping quality. In this study, we additionally have validated the quality of Illumina genotyping of wgaDNA derived using MDA from DNA obtained from buccal cells collected using either the distilled water/isopropanol method or the Scope method, fresh urine, and dried blood spots. All MDA reactions contained 20 to 50 ng gDNA that was subjected to a minimal number of freeze-thaw cycles. Our starting gDNA was high quality, intact, and had been stored <12 months after extraction. Other studies have emphasized the importance of the quality and concentration of the gDNA input to the WGA reaction for successful highly multiplexed SNP genotyping (16, 18, 19). Although we did not evaluate the OmniPlex method for WGA, it is reported to be an alternative that can be used for low-concentration or partially degraded DNA.[4] Indeed, Barker et al. (16) found no significant differences in reproducibility or gDNA concordance between the MDA and OmniPlex methods of WGA. It must be noted, however, that their input gDNA was ~100 ng and derived from transformed human cell lines (CEPH). Based on our data, we conclude that wgaDNA generated using the MDA method using as little as 20 ng of high-quality, intact gDNA can be successfully and reliably genotyped using the Illumina platform.

---

[4] http://www.rubicongenomics.com.

Buccal Cells. This study includes results from genotyping DNA extracted from buccal cells that were collected by two different buccal cell collection methods: the distilled water/isopropanol method and the Scope method. Both methods resulted in gDNA and wgaDNA that were genotyped with high call rates, high replicate concordance, and, most importantly, high concordance with matched gDNA from blood. We used 20 to 50 ng gDNA input for WGA, although we expected that the relevant human gDNA component of the input was less due to bacterial and food contamination. Even so, 20 to 50 ng was a sufficient amount of input gDNA for successful WGA using MDA as validated by subsequent genotyping on the Illumina platform.

Urothelial Cells. To consider the possibility of genotyping DNA obtained from urine samples when no blood samples are available from epidemiologic study participants, we evaluated DNA extracted from the urine of five of the subjects (four women and one man) who provided samples. DNA extractions from two (both women) of the five samples were sufficiently concentrated (50 ng/μL) to genotype as gDNA. All five gDNA samples, when subjected to whole genome MDA, reached concentrations sufficient for Illumina genotyping. We observed high concordance between replicates for both gDNA and wgaDNA and between genotype calls from both gDNA and wgaDNA and their matched gDNA from blood. It is noteworthy, however, that cluster patterns for urine sometimes differed slightly from the blood gDNA patterns, especially with respect to the θ parameter. In contrast, all other sample types, regardless of amplification treatment, showed consistently similar clustering patterns. Because the CEPH samples cluster similarly to blood and buccal gDNA and wgaDNA but may differ slightly from urine DNA, the CEPH data may not serve as a complete validation for genotyping urine DNA. A possible solution would be to genotype paired DNA samples derived from urine and blood from several individuals to validate genotype clustering in urine. We note that the sample size for urine DNA in this study was limited, particularly for the gDNA. Additionally, DNA in all samples was extracted from fresh urine; epidemiologic studies are more likely to have archived frozen urine samples. It is unclear what effect the time of day at void, freezing of urine, time stored until freezing, duration of frozen storage, and number of freeze-thaw cycles of stored urine have on subsequent yield and quality of extracted DNA; furthermore, the sex of the urine donor may affect DNA yield (4, 6). Urine may be a source of DNA for Illumina genotyping, but we suggest that more data are needed to fully evaluate its use and limitations.

Summary. In summary, successful multiplexed genotyping using the Illumina platform may be done using gDNA and wgaDNA generated from high-quality, intact gDNA. We were able to genotype gDNA and wgaDNA derived from blood, mouthwash, dried blood spots on Guthrie cards, and urine samples, which provided highly replicable and concordant data regardless of the sample source.

The methods by which we collected samples were optimal and may not reflect the conditions of collection or storage in ongoing epidemiologic studies. We note, however, that we were also able to recover high-quality DNA from blood samples collected from childhood leukemia patients and frozen for over a decade. Additionally, we were able to recover high-quality DNA from dried blood spots on Guthrie cards collected at birth. Both the long-term frozen blood samples and dried blood spots provided high call rates in addition to high concordance between paired samples. This additional aspect to our experiment validates Illumina use for genotyping high-quality DNA, whether freshly collected and never frozen or frozen or dried and stored for several years.

References

1. Hannelius U, Lindgren CM, Melen E, Malmberg A, von Dobeln U, Kere J. Phenylketonuria screening registry as a resource for population genetic studies. J Med Genet 2005;42:e60.
2. Steinberg K, Beck J, Nickerson D, et al. DNA banking for epidemiologic studies: a review of current practices. Epidemiology 2002;13:246–54.
3. van der Hel OL, Bueno de Mesquita HB, Sandkuijl L, et al. Rapid *N*-acetyltransferase 2 imputed phenotype and smoking may increase risk of colorectal cancer in women (Netherlands). Cancer Causes Control 2003;14:293–8.
4. van der Hel OL, van der Luijt RB, Bueno de Mesquita HB, et al. Quality and quantity of DNA isolated from frozen urine in population-based research. Anal Biochem 2002;304:206–11.
5. van Duijnhoven FJ, van der Hel OL, van der Luijt RB, Bueno de Mesquita HB, van Noord PA, Peeters PH. Quality of NAT2 genotyping with restriction fragment length polymorphism using DNA isolated from frozen urine. Cancer Epidemiol Biomarkers Prev 2002;11:771–6.
6. Yasuda T, Iida R, Takeshita H, et al. A simple method of DNA extraction and STR typing from urine samples using a commercially available DNA/RNA extraction kit. J Forensic Sci 2003;48:108–10.
7. Hosono S, Faruqi AF, Dean FB, et al. Unbiased whole-genome amplification directly from clinical samples. Genome Res 2003;13:954–64.
8. Lovmar L, Fredriksson M, Liljedahl U, Sigurdsson S, Syvanen AC. Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA. Nucleic Acids Res 2003;31:e129.
9. Luthra R, Medeiros LJ. Isothermal multiple displacement amplification: a highly reliable approach for generating unlimited high molecular weight genomic DNA from clinical specimens. J Mol Diagn 2004;6:236–42.
10. Paez JG, Lin M, Beroukhim R, et al. Genome coverage and sequence fidelity of ϕ29 polymerase-based multiple strand displacement whole genome amplification. Nucleic Acids Res 2004;32:e71.
11. Reich D, Patterson N, De Jager PL, et al. A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. Nat Genet 2005;37:1113–8.
12. Tranah GJ, Lescault PJ, Hunter DJ, De Vivo I. Multiple displacement amplification prior to single nucleotide polymorphism genotyping in epidemiologic studies. Biotechnol Lett 2003;25:1031–6.
13. Dean FB, Hosono S, Fang L, et al. Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci U S A 2002;99:5261–6.
14. Gunderson KL, Kruglyak S, Graige MS, et al. Decoding randomly ordered DNA arrays. Genome Res 2004;14:870–7.
15. Shen R, Fan JB, Campbell D, et al. High-throughput SNP genotyping on universal bead arrays. Mutat Res 2005;573:70–82.
16. Barker DL, Hansen MS, Faruqi AF, et al. Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. Genome Res 2004;14:901–7.
17. Pask R, Rance HE, Barratt BJ, et al. Investigating the utility of combining ϕ29 whole genome amplification and highly multiplexed single nucleotide polymorphism BeadArray genotyping. BMC Biotechnol 2004;4:15.
18. Sawcer S, Ban M, Maranian M, et al. A high-density screen for linkage in multiple sclerosis. Am J Hum Genet 2005;77:454–67.
19. Bergen AW, Haque KA, Qi Y, et al. Comparison of yield and genotyping performance of multiple displacement amplification and OmniPlex whole genome amplified DNA generated from multiple DNA sources. Hum Mutat 2005;26:262–70.