Short Communication

Successful Genome-Wide Scan in Paired Blood and Buccal Samples

Heather Spencer Feigelson,[1] Carmen Rodriguez,[1] Robert Welch,[2,3] Amy Hutchinson,[2,3] Wen Shao,[2] Kevin Jacobs,[2] W. Ryan Diver,[1] Eugenia E. Calle,[1] Michael J. Thun,[1] David J. Hunter,[4,6,7] Gilles Thomas,[3,4] and Stephen J. Chanock[3,4,5]

[1]Department of Epidemiology and Surveillance Research, American Cancer Society, NE, Atlanta, Georgia; [2]Intramural Research Support Program Science Applications International Corporation-Frederick, National Cancer Institute-Frederick, Frederick, Maryland; [3]Core Genotyping Facility, [4]Division of Cancer Epidemiology and Genetics, and [5]Section on Genomic Variation, Pediatric Oncology Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland; and [6]Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, and [7]Channing Laboratory, Brigham and Women's Hospital, Boston, Massachusetts

Abstract

Interest in genome-wide association studies to identify susceptibility alleles for cancer is growing, and several are currently planned or under way. Although the feasibility of collecting buccal cell samples as an alternative to venous blood samples as a source of genomic DNA has been shown, the validity of using DNA from buccal cells for genome-wide scans has not been assessed. We used 46 paired buffy coat and buccal cell samples to test the feasibility of using DNA from buccal cells for genotyping with the HumanHap300 Bead Chip (v.1.0.0) on the Illumina Infinium II platform. Genotyping was successful in every sample, regardless of DNA yield or sample type. Of the 317,502 genotypes attempted, 315,314 (99.3%) were successfully called. Completion rates were similar for buffy coat and buccal cell samples (99.63% and 99.44%, respectively; $P = 0.15$). Completion rates <99% were observed in only four samples and did not differ by specimen type. The paired samples showed exceptionally high concordance (99.96%). These results show that buccal cell samples collected and processed under optimal conditions can be used for genome-wide association studies with results comparable to those obtained from DNA extracted from buffy coat. (Cancer Epidemiol Biomarkers Prev 2007;16(5):1023–5)

Introduction

Contemporary genotyping technologies allow for thousands, and even hundreds of thousands, of genotypes to be done quickly using a relatively small amount of DNA (1-3). Interest in genome-wide scans to identify susceptibility alleles for cancer is growing, and the use of these technologies to perform genome-wide association studies is under way (2). One such study is the Cancer Genetic Markers of Susceptibility initiative that will use a genome-wide scan to identify genetic susceptibility to prostate and breast cancer.[8] Although the amount of DNA for these genotyping platforms is much smaller than the amount required by older technologies, DNA collected in epidemiologic studies from carefully characterized populations is still a finite resource that needs to be managed and preserved. Moreover, the source of the DNA can be critical, especially for high-throughput technologies that require optimal quality DNA to ensure high completion rates and fidelity. Select technologies, such as the Illumina Infinium II, cannot use whole genome amplified DNA.

Previous studies have reported on the feasibility of collecting buccal cell samples as a simple and cost-effective alternative to venous blood samples as a source of genomic DNA (4-8). As a result, many investigators have started collecting buccal cells, either in place of, or as a supplement to, blood samples when blood collection is either refused by a study participant, or is not feasible. However, there are at least two limitations of buccal cell samples compared with blood samples that would decrease their applicability to large-scale genotyping studies. One, the amount of human DNA that can be extracted from a typical buccal cell collection may not be sufficient, and two, the accuracy of genome-wide scan results from a buccal cell sample has not been established.

To determine the feasibility of using DNA from buccal cells for genome-wide association studies, we conducted a genome scan using the HumanHap300 Bead Chip (v.1.0.0) on the Illumina Infinium II platform with paired buffy coat and buccal cell samples from 23 participants in the American Cancer Society Cancer Prevention Study II (CPS-II) Nutrition Cohort. The HumanHap300 Bead Chip includes 317,502 single nucleotide polymorphisms (SNPs) derived from the Phase I HapMap data (9).[9]

Materials and Methods

From June 1998 to June 2001, blood samples were collected from 39,376 participants in the CPS-II Nutrition Cohort (described elsewhere; ref. 10). After obtaining informed consent, a maximum of 43 mL of nonfasting whole blood was collected from each participant, and samples were separated into aliquots of serum, plasma, RBC, and buffy coat. Samples were frozen in liquid nitrogen vapor phase at approximately −130°C for long-term storage.

In 2001 to 2002, buccal cells were collected from 69,467 CPS-II participants who had not provided a blood sample. For

Received 10/9/06; revised 1/10/07; accepted 2/21/07.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: H.S. Feigelson, Department of Epidemiology and Surveillance Research, American Cancer Society, 1599 Clifton Road, NE, Atlanta, GA. Phone: 404-929-6815; Fax: 404-327-6450. E-mail: heather.feigelson@cancer.org

doi:10.1158/1055-9965.EPI-06-0859

[8] http://cgems.cancer.gov/index.asp
[9] http://www.hapmap.org

Table 1. Results of DNA extraction and paired comparison of genotyping completion rates and concordance for blood and buccal samples

	Paired DNA samples (n = 23 pairs)	
	Buffy coat	Buccal cells
DNA yield median (range; ng/μL)	49.6 (0.55-74.68)	44.4 (14.32-66.04)
Genotyping completion rate, % (95% CI)*	99.63 (99.43-99.84)	99.44 (99.14-99.72)
Concordance, % (95% CI)	99.96 (99.93-99.99)	

*Percent based on 315,314 SNPs using the HumanHap 300 chip; P value for the difference in completion rates by sample type, 0.15.

quality control purposes, we also collected buccal cell samples from 537 individuals who had already given a blood sample. Buccal cell collection kits were mailed to CPS-II participants, and two buccal cell samples were collected using the ''swish-and-spit'' protocol (4). Briefly, participants were mailed two prelabeled collection cups, a small bottle of Scope mouthwash, and a return envelope. Participants were instructed to swish a tablespoon of Scope in their mouth for 60 s and spit the contents into one of the collection cups. Participants were instructed to do this once before bedtime, and then again the following morning as soon as they woke. After both the evening and the morning samples were collected, the participant was instructed to place both collection vials into the self-addressed, postage-paid envelope. Samples were received at the biorepository (Fisher BioServices) via regular mail. The two samples from each individual were then combined into one vial and centrifuged at 2,700 rpm for 15 min. The resulting pellet was resuspended in Tris-EDTA, split into two cryovials, and frozen in liquid nitrogen vapor phase at approximately −130°C for long-term storage.

We randomly selected 23 men who had both a blood and buccal cell sample to be used for this study (N = 46 samples). The age range of these men at the time of sample collection was 70 to 90 years (median, 79 years).

DNA was extracted from both buffy coat and buccal cell samples using the Autopure System (Gentra, Inc.). The volume of each sample was measured and transferred to moveable 1.2-mL tubes in a 96-well format (ScreenMates Sample Storage Tubes, Matrix Technologies). At the time of transfer, 2 μL of each sample was analyzed with the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies) to determine the working concentration (ng/μL) of each sample. Due to an error during DNA extraction, one buffy coat sample had very low volume.

Following the manufacturer's protocol, samples were run on the Illumina Golden Gate QDNA process (see Illumina BeadLab System Manual–Automated Process, Make QDNA/Read QDNA sections), a PicoGreen-based assay that uses 2 μL of DNA and compares it to the reference standard curve of lambda DNA. The resulting concentration data were used to transfer 2,000 ng of each sample to an ABgene 0.8-mL storage plate (ABgene, Inc. USA). The DNA was desiccated and resuspended in 40 μL of Tris-EDTA, for an overall expected concentration of 50 ng/μL.

Before beginning the Infinium II protocol, the concentrated samples were run through the QDNA process (as above) once more to be sure the concentrations were in the range of 50 ng/μL. Four low outlier samples were observed (concentrations between 0.5 and 36 ng/μL), which had previously been reported to have low amounts of DNA. The lowest of these outliers (concentration of 0.5 ng/μL) was the buffy coat sample that had very low volume as a result of lab error during DNA extraction. These samples were left in the process as a test to determine minimum DNA amounts needed for the Infinium II assay, which begins with a whole genome amplification step. Following the manufacturer's instructions for Infinium II, the samples were run using the automated process on the HumanHap300 Bead Chips (v.1.0.0). The system used was equipped with the Illumina Infinium LIMS. All data collected were analyzed using Illumina's Bead Studio software with Genotyping Module (v.2.2.22).

For analysis of the genotype results, completion rates per sample were calculated, and the Wilcoxon signed-rank test was used to compute a P value for the difference in completion rates by sample type (buffy coat versus buccal cell). Concordance was determined by comparison of matched blood and buccal sample results. The 95% confidence intervals (95% CI) were computed, accounting for clustering within subject using a generalized estimating equation.

Results and Discussion

The results of this study are shown in Table 1. The median (range) of total DNA (as measured by PicoGreen) for the blood and buccal cell samples was 49.6 ng/μL (0.55-74.68) and 44.4 ng/μL (14.32-66.04), respectively. Genotype analysis yielded high completion rates for each sample, regardless of DNA yield or sample type (Fig. 1). In fact, the four samples (one blood and three buccal cell) with concentrations between 0.5 and 36 ng/μL performed comparably to those samples with a measured concentration of DNA >50 ng/μL. Of the 317,502 genotypes attempted, 315,314 (99.3%) were successfully called; the majority of assays that failed were monoallelic in our population. Genotype completion rates from these 315,314 genotypes were similar for buffy coat and buccal cell samples (99.63% and 99.44%, respectively; P = 0.15). Completion rates for SNP genotype assays were above 99% for all but four samples; two buffy coat samples and one buccal sample had completion rates between 97% and 99%. One buccal sample had a completion rate of 96.6%. An analysis of fitness for Hardy-Weinberg proportion revealed ~15,000 SNPs were significant (P < 0.05), which represents 4.7% of the attempted SNPs (data not shown).

The analysis of paired samples showed exceptionally high concordance (99.96%; 95% CI, 99.93-99.99). A total of 18 of the 23 pairs were 100% concordant, 4 were 99.9% concordant, and the lowest pair was 99.6% concordant, and concordance did not depend on sample concentration. Based on these data, we infer that the observed error rate seems to be significantly below the error rates reported for other genotyping methods. These data provide clear evidence that buccal cell samples obtained using the swish-and-spit protocol can be used on the Illumina Infinium II platform, with results comparable to those obtained using DNA extracted from buffy coat.

The primary limitation of this study is the small sample size. However, the size of this study (23 pairs) has sufficient power

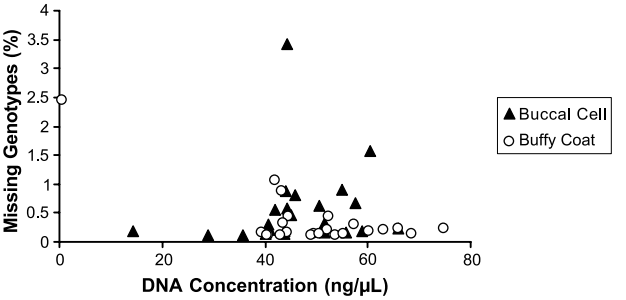

Figure 1. Genotype success rates (displayed as percent missing) for 315,314 SNPs using the HumanHap 300 chip for DNA samples extracted from buccal cell and buffy coat samples.

to detect a difference as low as 0.5% between the genotyping success rates. This study was designed to investigate the performance of high-quality genomic DNA extracted from buccal material using the Infinium II technology, and as such, its conclusions are suitable for this technology. Additional studies are needed to address the performance of other genotyping platforms currently available as well as different sample handling methods and sources of DNA.

Our results have important implications for the conduct of whole genome studies. In studies where buccal samples have been extracted and stored properly, it will be possible to increase sample size and could obviate whole genome amplification before genotyping (11-19). Although the latter can generate large quantities of DNA from nanograms of high-quality DNA, a number of studies have shown that the amplification step is not 100% accurate, and loss of heterozygosity, often clustered in GC-rich loci, can arise (12). We have shown that it is possible to generate high-quality genotype data with DNA extracted from the swish-and-spit protocol under optimal circumstances. It is possible that mishandling of samples or suboptimal storage can undermine the integrity of genomic DNA and, thus, its performance in high-throughput genotype analyses (20). Thus, it is critical that established protocols be carefully followed, and that high-quality genomic DNA be extracted and stored properly if buccal cell materials are to be used for dense whole genome scans.

References

1. Gunderson K, Steemers F, Lee G, Mendoza L, Chee M. Genome-wide scalable SNP genotyping assay using microarray technology. Nat Genet 2005;37:549–54.
2. Thomas D. Are we ready for genome-wide association studies? Cancer Epidemiol Biomarkers Prev 2006;15:595–8.
3. Wang W, Barratt B, Clayton D, Todd J. Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 2005;6:109–18.
4. Feigelson H, Rodriguez C, Robertson A, et al. Determinants of DNA yield and quality from buccal cell samples collected with mouthwash. Cancer Epidemiol Biomarkers Prev 2001;10:1005–8.
5. Garcia-Closas M, Egan KM, Abruzzo J, et al. Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. Cancer Epidemiol Biomarkers Prev 2001;10:687–96.
6. Andrisin TE, Humma LM, Johnson JA. Collection of genomic DNA by the noninvasive mouthwash method for use in pharmacogenetic studies. Pharmacotherapy 2002;22:954–60.
7. King IB, Satia-Abouta J, Thornquist MD, et al. Buccal cell DNA yield, quality, and collection costs: comparison of methods for large-scale studies. Cancer Epidemiol Biomarkers Prev 2002;11:1130–3.
8. Le Marchand L, Lum-Jones A, Saltzman B, Visaya V, Nomura AM, Kolonel LN. Feasibility of collecting buccal cell DNA by mail in a cohort study. Cancer Epidemiol Biomarkers Prev 2001;10:701–3.
9. Altshuler D, Brooks L, Chakravarti A, et al. A haplotype map of the human genome. Nature 2005;437:1299–20.
10. Calle EE, Rodriguez C, Jacobs EJ, et al. The American Cancer Society nutrition cohort: rationale, study design and baseline characteristics. Cancer 2002;94:2490–501.
11. Thomas D, Haile R, Duggan D. Recent developments in genomewide association scans: a workshop summary and review. Am J Hum Genet 2005; 77:337–45.
12. Bergen A, Haque K, 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.
13. Blanco L, Bernad A, Lazaro J, Martin G, Garmendia C, Salas M. Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. J Biol Chem 1989;264:8035–40.
14. Dean F, 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.
15. Lasken R, Egholm M. Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. Trends Biotechnol 2003; 21:531–5.
16. Lovmar L, Fredriksson M, Liljedahl U, Sigurdsson S, Syvanen A. Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA. Nucleic Acids Res 2003;31:e129.
17. Rook M, Delach S, Deyneko G, Worlock A, Wolfe J. 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.
18. Tanabe C, Aoyagi K, Sakiyama T, et al. Evaluation of a whole-genome amplification method based on adaptor-ligation PCR of randomly sheared genomic DNA. Genes Chromosomes Cancer 2003;38:168–76.
19. 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.
20. Bergen A, Qi Y, Haque K, et al. Effects of electron-beam irradiation on whole genome amplification. Cancer Epidemiol Biomarkers Prev 2005;14:1016–9.

Successful Genome-Wide Scan in Paired Blood and Buccal Samples

Heather Spencer Feigelson, Carmen Rodriguez, Robert Welch, et al.

Cancer Epidemiol Biomarkers Prev 2007;16:1023-1025.