

RESEARCH ARTICLE

Genotyping of Single Nucleotide Polymorphisms in DNA Isolated from Serum Using Sequenom MassARRAY Technology

Tess V. Clendenen^{1*}, Justin Rendleman¹, Wenzhen Ge¹, Karen L. Koenig¹, Isaac Wirgin³, Diane Currie³, Roy E. Shore^{3,4}, Tomas Kirchoff¹, Anne Zeleniuch-Jacquotte^{1,2}

1 Department of Population Health, New York University School of Medicine, New York, New York, United States of America, **2** NYU Cancer Institute, New York University Langone Medical Center, New York, New York, United States of America, **3** Department of Environmental Medicine, New York University School of Medicine, New York, New York, United States of America, **4** Radiation Effects Research Foundation, Hiroshima Japan

* tess.clendenen@nyumc.org



OPEN ACCESS

Citation: Clendenen TV, Rendleman J, Ge W, Koenig KL, Wirgin I, Currie D, et al. (2015) Genotyping of Single Nucleotide Polymorphisms in DNA Isolated from Serum Using Sequenom MassARRAY Technology. *PLoS ONE* 10(8): e0135943. doi:10.1371/journal.pone.0135943

Editor: Claire Wade, University of Sydney, AUSTRALIA

Received: March 22, 2015

Accepted: July 28, 2015

Published: August 14, 2015

Copyright: © 2015 Clendenen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Some relevant data are included within the paper and its Supporting Information. Additional patient-level data cannot be made freely available due to restrictions imposed by the IRB of NYU School of Medicine. An anonymized dataset is available upon request to Anne Zeleniuch-Jacquotte (anne.jacquotte@nyumc.org).

Funding: Support was provided by National Cancer Institute (www.cancer.gov) grants R01 CA081212 (AZJ), R01 CA098661 (AZJ) and P30 CA016087; and Center grant ES000260 from the National Institute of Environmental Health Sciences (<http://>

Abstract

Background

Large epidemiologic studies have the potential to make valuable contributions to the assessment of gene-environment interactions because they prospectively collected detailed exposure data. Some of these studies, however, have only serum or plasma samples as a low quantity source of DNA.

Methods

We examined whether DNA isolated from serum can be used to reliably and accurately genotype single nucleotide polymorphisms (SNPs) using Sequenom multiplex SNP genotyping technology. We genotyped 81 SNPs using samples from 158 participants in the NYU Women's Health Study. Each participant had DNA from serum and at least one paired DNA sample isolated from a high quality source of DNA, i.e. clots and/or cell precipitates, for comparison.

Results

We observed that 60 of the 81 SNPs (74%) had high call frequencies ($\geq 95\%$) using DNA from serum, only slightly lower than the 85% of SNPs with high call frequencies in DNA from clots or cell precipitates. Of the 57 SNPs with high call frequencies for serum, clot, and cell precipitate DNA, 54 (95%) had highly concordant ($>98\%$) genotype calls across all three sample types. High purity was not a critical factor to successful genotyping.

Conclusions

Our results suggest that this multiplex SNP genotyping method can be used reliably on DNA from serum in large-scale epidemiologic studies.

www.niehs.nih.gov/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Single nucleotide polymorphisms (SNPs) have been found to be associated with risk of many complex diseases. For instance, more than 70 susceptibility SNPs have been identified for breast cancer [1–3]. Assessment of gene-environment interactions is a natural next step in research on the etiology of these diseases. Prospective cohort studies have well-annotated data on exposures that are collected prior to diagnosis, thereby limiting the risk of bias. DNA can be isolated from serum or plasma samples collected from these cohorts although it may not be of the high quality/quantity of DNA isolated from other types of biological samples, such as buffy coats or whole blood.

DNA isolated from serum has been used by our group and others to achieve high call frequencies with single SNP Taqman assays (without performing whole genome amplification prior to PCR amplification) in epidemiological studies [4–7]. However, Taqman assays, as with other single-SNP genotyping methods, are time consuming and costly, which limits their usefulness for large-scale studies. We conducted a study to assess the utility of the multiplex Sequenom MassARRAY system (Sequenom Inc., CA) for genotyping SNPs in genomic DNA isolated from serum. Oligonucleotide primers are extended, dependent on the SNP-specific template sequence (PCR product), and MALDI-TOF mass spectrometry is used to differentiate SNP alleles based on the different masses of extension products. We designed this study to compare the frequency of successful genotyping calls and the concordance of calls from paired serum- whole blood DNA collected from subjects in a large prospective cohort study.

Methods

Between 1985 and 1991, the NYU Women's Health Study (NYUWHS) enrolled 14,274 healthy women aged 35–65 years at a breast cancer screening center. Enrollment in the cohort required donation of 30mL of blood. Blood collection was performed without anti-coagulant; tubes were kept covered at room temperature (25°C) for 15 minutes and then at 4°C for 60 minutes. Tubes were then centrifuged and the serum supernatant was collected and divided into several 1 mL aliquots. Serum samples were stored in polypropylene screw capped tubes at -80°C within 2 hours of collection. For blood samples collected in or after 1988, the cellular precipitates were also partitioned off and stored in polypropylene tubes at -80°C. The remaining blood clots from each sample were stored starting in 1989. Clots were stored immediately at -40°C after serum and cell precipitate removal in two sealable plastic-lined aluminum bags. The NYU Medical Center Institutional Review Board approved this study. Participants completed written informed consent.

This study included samples from 158 NYUWHS participants: 48 women with both serum and clots, 48 with both serum and cell precipitates, and 62 women with all three types of samples, *i.e.*, serum, clots, and cell precipitates.

DNA was isolated from 200 μ L of serum, clots (a section of clot was digested in streptokinase for at least 8 hours prior to DNA isolation), and cell precipitates manually using QIAamp Blood Mini kits from (Qiagen, Inc., Valencia, CA) according to manufacturer protocols. A nanodrop spectrometer was used to estimate DNA concentration and purity using the A260/A280 ratio. Three repeat readings were taken for each DNA sample and the mean of the three measurements was used.

We genotyped 81 SNPs with minor allele frequencies ranging from 2% to 40%. SNPs genotyped in this study were selected for a separate study. Among the 81 SNPs, 57 (70%) were tag SNPs in three vitamin D-related gene regions on three chromosomes, and 24 were SNPs in 19 gene regions on 10 other chromosomes. Approximately 10% of the SNPs were in coding regions, 10% in UTRs, 4% in non-coding regions or upstream of gene regions, and 75% were intronic.

Table 1. Spectrophotometer estimated DNA yield and purity, by sample type.

	# of samples	DNA yield, µg/mL, Median (10%, 90%)*	A260/A280		
			<1.8 n (%)	1.8–2.0 n (%)	>2.0 n (%)
DNA Source					
Serum	158	9.7 (1.7, 14.4)	101 (64%)	26 (16%)	31 (20%)
Whole Blood Clots	110	8.5 (2.9, 15.1)	40 (36%)	32 (29%)	38 (35%)
Cell Precipitates	110	18.4 (8.3, 49.8)	27 (25%)	78 (71%)	5 (5%)

Note: A260/A280 between 1.8–2.0 indicates good DNA purity

*Based on an averaged measurement of triplicate readings (CVs were ≤7% for A260/A280 for each sample type)

doi:10.1371/journal.pone.0135943.t001

Genotyping was performed using multiplex Sequenom technology, using 2 µL per reaction of DNA isolated from serum, clots, and cell precipitates and following standard manufacturer protocols (without whole genome amplification). The genotyping design software generated three multiplex reactions: 29 SNPs (plex1), 28 SNPs (plex 2), and 24 SNPs (plex 3). All genotyping was performed with the MassARRAY iPLEX platform (Sequenom Inc, San Diego, California, USA). Laboratory personnel were blinded to the identity and source of the DNA sample.

Call frequencies, expressed as percentages, were calculated for each SNP. Genotype concordance was assessed by examining groups of paired samples (serum/clot, serum/cell precipitate, and clot/cell precipitate) separately. SNPs that could not be called were excluded from concordance calculations. A call frequency across samples ≥95%, a commonly used standard in epidemiologic studies [8,9], and concordance >98% were considered acceptable.

Results

Table 1 shows the characteristics of the DNA samples used for this reliability study. Cell precipitates had a median DNA concentration approximately 2-fold higher than that of clots and serum (18.4 µg/ml vs. 8.5 µg/ml and 9.7 µg/ml, respectively). Of the three sample types, only cell precipitates had a majority of samples with an A260/A280 ratio of 1.8–2.0 (72% of cell precipitate DNA samples vs. 29% of clot and 16% of serum DNA samples). Though both clots and cell precipitates are whole blood sources of DNA, isolation of DNA from cell precipitates yielded more DNA with higher purity than clots. Thus, we consider cell precipitate DNA as the highest quality DNA source when comparing with genotyping data based on DNA from serum.

As shown in Table 2, about 85% of SNPs had call frequencies ≥95% for clot and cell precipitate sources of DNA vs. 74% for DNA from serum. There were 57 SNPs with call frequencies ≥95% for all three DNA sources (clots, cell precipitates and serum). Out of these 57 high call-frequency SNPs, 54 (95%) showed >98% concordance between serum and clot DNA and 55 SNPs (96%) showed >98% concordance between serum and cell precipitate DNA. Concordance was also >98% for comparisons between clots and cell precipitate DNA (65/68 = 96% of

Table 2. Summary of call frequencies and concordance.

Source of DNA	N of samples	N of SNPs with ≥95% calls	N of SNPs in >98% concordance with serum ^a
Serum	158	60/81 = 74%	
Clots	110	69/81 = 85%	54/57 = 95%
Cell precipitates	110	70/81 = 86%	55/57 = 96%

^a Denominator (n = 57) is the number of SNPs with call frequencies ≥95% for all three of the DNA source sample types (clots, cell precipitates, and serum)

doi:10.1371/journal.pone.0135943.t002

SNPs with call frequencies >95% for both clots and cell precipitates). The call frequencies and concordance data for each SNP are shown in the [S1 Table](#).

Discussion

Our study demonstrates that Sequenom MassARRAY technology can be used to genotype SNPs with a multiplex approach using a small volume of serum DNA. For the SNPs we examined, 74% had call frequencies of 95% or higher (QC criteria commonly used in epidemiological studies [8,9]) in DNA from serum. Of the SNPs with high call frequencies, 95% also had >98% concordance with genotypes from clot and cell precipitate DNA, indicating that DNA from low quantity/quality serum sources can be reliably genotyped. The proportion of SNPs with $\geq 95\%$ successful calls for serum DNA (74%) was only slightly lower than that observed for DNA from clots (85%) and cell precipitates (86%), and concordance was very high for SNPs with high call frequencies (95% of these SNPs had concordance >98%).

Only one other study examined the utility of multiplex Sequenom for genotyping DNA isolated from serum. It concluded that the genotyping using Sequenom was not reliable, contrary to genotyping using Taqman methods [4]. In that study, DNA was isolated using Qiagen kits for 50 paired samples from white blood cells and serum. Forty-eight SNPs were genotyped (number of plex reactions not given) and overall call frequencies were described in general terms as poor with high discordance for serum DNA. Call frequencies and concordance were also low for a second high-throughput (SNPlex) method that was used to genotype the same samples in a different laboratory. Genotypes obtained from serum DNA across laboratories differed for many SNPs. DNA yields (from samples from the same cohort) were 10–474 ng/mL, substantially lower than our estimated yields, but the investigators noted that concentrations may have been underestimates based on PCR results and that these measurements were done in samples after ethyl acetate and hexane extraction (which is a non-standard treatment of samples prior to DNA analysis). The different observations of that prior study compared to ours may be due to several factors, besides overall DNA quantities: 1) Difference in the number of SNPs included in each multiplex reaction: Higher plexing could reduce genotyping quality; 2) Serum separation procedures (i.e., the use of separator tubes vs. non-separator tubes in our study), which may impact the quantity or quality of serum DNA; 3) Advancements in Sequenom SNP assay designs since 2008 which have made Sequenom technology more accommodating to lower concentration and/or quality DNA; and 4) Difference in SNPs selection between studies, and hence different pools of SNPs constituting the multiplex reactions.

A DNA sample with an A260/A280 ratio of 1.8–2.0 is considered to have low contamination (e.g. by proteins and carbohydrates). Only 29% of clots and 16% of serum DNA had values within this range. It is possible that different DNA isolation protocols/kits, and, for serum, a volume larger than 200 μL , would have led to better results. We were not able to compare different approaches for DNA isolation because the serum samples collected in the NYUWHS (and in other long-term cohorts) are very precious, and the minimal necessary volume is allocated for each assay. It should be noted though that we achieved successful genotyping calls and high concordance for over 70% of SNPs, despite the low percentage of samples within the 1.8–2.0 range. Further, excluding DNA samples with A260/A280 ratios below 1.8 or above 2.0 did not result in appreciable differences in the percentage of SNPs with acceptable call frequencies (data not shown). Since exclusion of samples (and therefore of participants) results in loss of power and possibly bias in epidemiologic studies, it does not appear that it would be a good strategy to exclude from the statistical analysis of these studies samples falling outside the 1.8–2.0 range for the A260/A280 ratio. High purity of DNA does not appear to be a critical factor to successful genotyping using Sequenom iPLEX.

Genome wide association studies (GWAS) and other large targeted genotyping scans, apply relatively strict criteria for exclusion of samples with low genotype call frequencies across all SNPs [10,11]. Excluding samples in our data set with call frequencies <90% across all SNPs resulted in the removal of 6 clots (5%), 7 cell precipitates (6%), and 47 serum samples (30% of all serum samples in the study, data not shown). While these exclusions increased the number of SNPs with call frequencies $\geq 95\%$ for serum DNA (from 74% to 84%), at the same time they reduced the total number of DNA serum samples by $\sim 30\%$. Concordance was similar for all SNPs whether or not these samples were included. Our findings, therefore, suggest that excluding samples with low call frequencies would not substantially reduce the rate of genotype errors, and would result in the loss of a large number of subjects for studies in which serum is the source of DNA.

With the exception of one, all the SNPs with a high call frequency ($\geq 95\%$) in serum DNA samples were also highly concordant ($\geq 98\%$) with DNA from cell precipitates (considered as a high quality reference in our study comparison). This suggests that applying a stringent call-frequency criterion leads to the exclusion of SNPs with unreliable genotype calls. This result is consistent with the high correlation between miscalls and no calls observed by others [12] and suggests that the multiplex Sequenom platform can be successfully used for genotyping DNA extracted from serum. As is done routinely, SNPs that do not meet QC criteria could be replaced with SNPs in high linkage disequilibrium that might be more easily genotyped with Sequenom technology or alternatively, genotyped using single SNP assays, such as Taqman. Such SNPs could be identified by conducting a pilot study prior to genotyping case-control study samples.

Isolation of DNA from 200 μL of serum (average yield $\sim 425\text{ng}$ of DNA suspended in 50 μL buffer, using the Qiagen blood kit and protocol) provides a sufficient amount of DNA to genotype up to 1,000 SNPs using Sequenom (2 μL DNA for each of 25 reactions, 40-plex each = 1000 SNPs). Assuming our results are broadly applicable to other SNP loci, a similarly conducted genotyping experiment from serum DNA would yield approximately 75% of SNPs having $>95\%$ call frequencies (and also $\geq 98\%$ concordance).

Conclusion

Considering that for many diseases, GWAS and other large targeted genotyping studies have identified susceptibility SNPs which are expected to number at most in the hundreds, and that many SNPs can be tagged by other SNPs in high LD should the assay not work for certain SNPs, our study shows that the iPLEX Sequenom platform appears to be a useful tool for SNP genotyping in epidemiologic studies that have serum samples as the only source of DNA for their participants.

Supporting Information

S1 Table. Call frequency and concordance by SNP, all samples.
(DOCX)

Author Contributions

Conceived and designed the experiments: TVC KLK IW RES TK AZJ. Performed the experiments: JR WG DC. Analyzed the data: TVC KLK AZJ. Contributed reagents/materials/analysis tools: IW TK. Wrote the paper: TVC JR KLK TK AZJ. Interpretation of the data and revision of the manuscript for important intellectual content: TVC JR WG KLK IW DC RES TK AZJ.

References

1. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 45: 353–361, 361e351–352. doi: [10.1038/ng.2563](https://doi.org/10.1038/ng.2563) PMID: [23535729](https://pubmed.ncbi.nlm.nih.gov/23535729/)
2. Zheng W, Zhang B, Cai Q, Sung H, Michailidou K, Shi J, et al. (2013) Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. *Hum Mol Genet* 22: 2539–2550. doi: [10.1093/hmg/ddt089](https://doi.org/10.1093/hmg/ddt089) PMID: [23535825](https://pubmed.ncbi.nlm.nih.gov/23535825/)
3. Ghoussaini M, Fletcher O, Michailidou K, Turnbull C, Schmidt MK, Dicks E, et al. (2012) Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet* 44: 312–318. doi: [10.1038/ng.1049](https://doi.org/10.1038/ng.1049) PMID: [22267197](https://pubmed.ncbi.nlm.nih.gov/22267197/)
4. Hoque A, Goodman P, Ambrosone CB, Figg WD, Price DK, Kopp W, et al. (2008) Extraction of DNA from serum for high-throughput genotyping: findings from pilot studies within the Prostate Cancer Prevention Trial. *Urology* 71: 967–970. doi: [10.1016/j.urology.2007.11.042](https://doi.org/10.1016/j.urology.2007.11.042) PMID: [18267333](https://pubmed.ncbi.nlm.nih.gov/18267333/)
5. Clendenen T, Zeleniuch-Jacquotte A, Wirgin I, Koenig KL, Afanasyeva Y, Lundin E, et al. (2013) Genetic variants in hormone-related genes and risk of breast cancer. *PLoS One* 8: e69367. doi: [10.1371/journal.pone.0069367](https://doi.org/10.1371/journal.pone.0069367) PMID: [23935996](https://pubmed.ncbi.nlm.nih.gov/23935996/)
6. Lundin E, Wirgin I, Lukanova A, Afanasyeva Y, Krogh V, Axelsson T, et al. (2012) Selected polymorphisms in sex hormone-related genes, circulating sex hormones and risk of endometrial cancer. *Cancer Epidemiol* 36: 445–452. doi: [10.1016/j.canep.2012.04.006](https://doi.org/10.1016/j.canep.2012.04.006) PMID: [22633539](https://pubmed.ncbi.nlm.nih.gov/22633539/)
7. Clendenen TV, Arslan AA, Koenig KL, Enquist K, Wirgin I, Agren A, et al. (2008) Vitamin D receptor polymorphisms and risk of epithelial ovarian cancer. *Cancer Lett* 260: 209–215. PMID: [18079052](https://pubmed.ncbi.nlm.nih.gov/18079052/)
8. Spurdle AB, Thompson DJ, Ahmed S, Ferguson K, Healey CS, O'Mara T, et al. (2011) Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nat Genet* 43: 451–454. doi: [10.1038/ng.812](https://doi.org/10.1038/ng.812) PMID: [21499250](https://pubmed.ncbi.nlm.nih.gov/21499250/)
9. Gaudet MM, Kuchenbaecker KB, Vijai J, Klein RJ, Kirchhoff T, McGuffog L, et al. (2013) Identification of a BRCA2-specific modifier locus at 6p24 related to breast cancer risk. *PLoS Genet* 9: e1003173. doi: [10.1371/journal.pgen.1003173](https://doi.org/10.1371/journal.pgen.1003173) PMID: [23544012](https://pubmed.ncbi.nlm.nih.gov/23544012/)
10. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 39: 870–874. PMID: [17529973](https://pubmed.ncbi.nlm.nih.gov/17529973/)
11. Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, et al. (2009) Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat Genet* 41: 1303–1307. doi: [10.1038/ng.485](https://doi.org/10.1038/ng.485) PMID: [19915576](https://pubmed.ncbi.nlm.nih.gov/19915576/)
12. Mead S, Poulter M, Beck J, Uphill J, Jones C, Ang CE, et al. (2008) Successful amplification of degraded DNA for use with high-throughput SNP genotyping platforms. *Hum Mutat* 29: 1452–1458. doi: [10.1002/humu.20782](https://doi.org/10.1002/humu.20782) PMID: [18551557](https://pubmed.ncbi.nlm.nih.gov/18551557/)