Published in final edited form as:

Urology. 2008 May ; 71(5): 967–970.

Extraction of DNA from Serum for High-throughput Genotyping: Findings from Pilot Studies within the Prostate Cancer Prevention Trial

Ashraful Hoque^a, Phyllis Goodman^b, Christine B. Ambrosone^c, William D. Figg^d, Doug Price^d, William Koop^d, Xifeng Wu^a, Jeffrey Conroy^c, Teresa A. Lehman^e, and Regina M. Santella^f

^aThe University of Texas M. D. Anderson Cancer Center, Houston, Texas ^bFred Hutchinson Cancer Research Center, Seattle, Washington ^cRoswell Park Cancer Institute, Buffalo, New York ^dNational Cancer Institute, Bethesda, Maryland ^eBioServe Biotechnologies, Ltd., Laurel MD ^fColumbia University, New York, New York

Abstract

OBJECTIVES—DNA extraction from blood and genotyping for candidate single nucleotide polymorphisms (SNP) is now an important part of almost all molecular epidemiologic studies. However, in many studies, the amount of blood sample is limited or only serum is available. We conducted several pilot studies to identify methods for DNA extraction and high-throughput SNP genotyping of both white blood cell (WBC) and serum DNA that can be done centrally and reliably for large numbers of samples.

METHODS—We used biospecimens from The Prostate Cancer Prevention Trial (PCPT), a phase III, double-blind, placebo-controlled trial that tested the efficacy of finasteride for the primary prevention of prostate cancer. DNA was extracted from WBCs, from serum, and also from serum after organic solvent extraction for analysis of hormones. We also conducted blinded high-throughput genotyping in 3 laboratories to assess feasibility and reliability of results with differing methodologies using DNA from WBCs and from serum.

RESULTS—Genotyping of DNA extracted from WBCs resulted in highly reliable, reproducible results across laboratories using different genotyping platforms. However, genotyping with DNA extracted from serum did not provide reliable data using high-throughput multiplex approaches such as Sequenom (hME and iPLEX) and Applied Biosystems SNPLex, but was successful using Taqman.

CONCLUSIONS—Based upon the results of these pilot studies, we conclude that DNA obtained from serum must be used judiciously, and that genotyping using multiplex methods is not suitable for serum DNA.

Keywords

Prostate cancer prevention; molecular epidemiology; genotyping; serum DNA extraction; multiplex methods

Corresponding Author's Contact Information: Regina M. Santella PhD, School of Public Health, Columbia University, 630 W 168th St, Room 19-418, New York NY 10032; Phone 212 305-1996; Fax 212 305-5328, Email: rps1@columbia.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

INTRODUCTION

DNA extraction from blood and genotyping for single nucleotide polymorphisms (SNPs) in candidate genes/pathways is now a routine component of many cancer epidemiologic studies. However, in many studies the amount of DNA is limited or only serum is available. To investigate several alternate methods of large scale DNA extraction and genotyping as well as the reliability of serum DNA for genotyping, we carried out several pilot studies testing DNA extraction methods and multiplex genotyping. These studies were carried out in order to select methods for application in an National Cancer Institute-funded Program Project grant investigating the biologic mechanisms underlying the results of The Prostate Cancer Prevention Trial (PCPT), a phase III, double-blind, placebo-controlled trial involving 18,882 men to test the efficacy of finasteride for the primary prevention of prostate cancer. The primary results showed a 24.8% overall reduction in prostate cancer risk and an increased risk of high-grade disease (Gleason score 7 or higher).¹ The Program Project is evaluating potential genetic and environmental risk factors for prostate cancer, and testing whether risk factors modify the effects of finasteride treatment, including effects on the risk of high-grade disease. The factors evaluated include numerous dietary and lifestyle factors, as well as SNPs in genes involved in steroid metabolism, oxidative stress, inflammation, and DNA repair. Serum levels of androgens and estrogens as well as insulin-like growth factors are also under study. Data and samples are available from all 18,882 participants for use by other investigators. Information on the complete biospecimen resource can be found at <https://swog.org/Members/download/bulletinboard/Article163.pdf>.

MATERIALS AND METHODS

As part of the study, the PCPT required collection of approximately 15 ml of blood from participants at the first visit and at each subsequent annual visit. Bloods were collected into tubes containing a gel, to separate serum from clot after centrifugation and kept at room temperature for 30–60 min before centrifugation. Serum was removed and the sample frozen as quickly as possible after separation. Samples were generally sent by overnight courier to Esoterix (Calabasas, CA) within 24 hrs of collection, but some were stored for up to 7 days at -20°C before shipment. Upon receipt, the sample was thawed to remove an aliquot for analysis of PSA. Random samples were also later analyzed for dihydrotestosterone. The remaining serum was refrozen in 0.5 ml aliquots (at least 4 for each specimen). A total of 119,523 serum specimens were obtained during the PCPT and are now housed at -70°C at the University of Colorado.

In the midst of the study, it was decided to collect an additional blood sample with anticoagulant specifically for isolation of DNA for genotyping studies. Three acid washed 7 ml EDTA tubes were used to collect approximately 20 ml of blood from 8,550 men and were kept protected from light. Samples were shipped overnight chilled to NCI Frederick (Frederick, MD) for processing. Plasma was separated into 5×1.5 ml aliquots, WBC's into 2 aliquots and the RBC's into a 4.5 ml aliquot. All samples are currently stored at -70°C at NCI Frederick.

RESULTS

White Blood Cell DNA Extraction and Genotyping

To determine the best methods for DNA extraction and high throughput genotyping, a pilot project was carried out to test three different facilities for DNA extraction and three for genotyping. Since bloods collected for the WBC bank were processed and stored at NCI-Frederick, this site was one of those selected for DNA extraction. Two commercial companies were also selected. These three sites represented a range of DNA preparation methods and costs. Frederick used classical phenol extraction, while the two companies used salting out and

commercial DNA extraction cartridges. NCI-Frederick pulled one of the two frozen vials of WBCs for 200 subjects, thawed them and divided the material into three equal portions, keeping one for their own DNA extraction and shipping the others to the companies. Unfortunately, after receipt of the samples, one company decided to terminate their DNA extraction and genotyping service and the samples were returned to Frederick. Thus, only two facilities for DNA extraction were compared.

DNA extraction pilot results including means, range and values for 1st and 3rd quartiles are summarized in Table 1. Total DNA yields, as assessed by absorbance at 260nm (A260) and volume, were higher in samples from Frederick as was the A260/280 ratio, a common measure of DNA purity. Many of the aberrant A260/280 values were from samples with very low yields and may not be accurate.

For the first genotyping pilot, the commercial company used Sequenom's MassARRAY matrix-assisted laser desorption/ionization/time-of-flight mass spectrometer (MALDI-TOF) for single SNP analysis. This method takes advantage of the ability to mass spectroscopy to detect very small differences in molecular weight after single nucleotide extension of probes that sit one nucleotide short of the SNP. The company tested the 200 DNAs prepared on site as well as those made at NCI-Fredrick, but after relabeling. This method allowed the blinded 100% duplication of the genotyping. Four SNPs were selected for genotyping, *MnSOD* T141C (rs#4880), *CYP3A4* -392, *AR* Stu (rs#6152) and *CAT* -262C/T. There were no discordant data for SNPs in *AR* and *CYP3A4*, however, 2 samples were discordant in *CAT* and 3 in *MnSOD*. The two samples discordant for *CAT* were also discordant for *MnSOD*, suggesting a sample mixup.

Serum DNA Extraction and Genotyping

Prior to the WBC collection in PCPT, genotyping for the CAG repeat in *AR* was being carried out using DNA extracted from serum using the QiaAMP UltraSens Virus kit according to the manufacturer's instructions (Qiagen, Valencia, CA).² Past results using this kit on over 2000 samples for (CAG)_n analysis in the *AR* (androgen receptor) gene showed a 92% rate of extracting amplifiable DNA from serum. This gene is X-linked and hence present in only one copy in men. Thus, it may not be representative of autosomal loci which will occur in two copies.

To determine if serum DNA could be used for genotyping the over 60 SNPs to be analyzed in the PCPT Program Project grant, we originally proposed to use whole genome amplification (WGA) of DNA isolated from serum for genotyping those subjects without a white blood cell sample. However, because of concerns with the reliability of WGA of serum DNA, a pilot study was carried out to determine if DNA could be extracted from serum that had been used for the analysis of hormone levels.³ For this assay, the serum was extracted with ethyl acetate and hexane. It was believed that this process would leave the DNA intact and could provide a convenient source of DNA. After solvent extraction of hormones, 40 serum samples were used for DNA extraction. A total of 200 µl of serum was applied to Qiagen DNA Mini extraction columns, the extracted DNA was dissolved in 200 µl of buffer, and yields were determined by pico green. Yields ranged from 10–474 ng/ml equivalent of serum. However, PCR data indicated that these concentrations were an underestimation. Samples were then used for genotyping of two polymorphisms with variant allele frequencies of ~35% (*XPD* codon 23 and *XRCC1* codon 399) using single nucleotide extension with fluorescence polarization detection of incorporated nucleotides. With a single assay, 39 of 40 samples were successfully genotyped for *XPD* and 35 for *XRCC1* using 6 µl of the 200 µl DNA sample for each assay. The variant frequency was 31% for both SNPs. These results indicate that it is possible to isolate DNA from sera after extraction of hormones that can be used for single SNP analysis.

Comparison of Multiplex Genotyping Methods on WBC and Serum DNA

Genotyping methodologies have significantly changed over the past few years, with multiplexing methods becoming commonly used. We therefore tested competing technologies and laboratories to determine which methodology to use on PCPT samples. In addition, the NCI-Frederick facility had recently switched to using a Qiagen BioRobot M48 for DNA extraction. This system was thus used to isolate WBC and serum DNA. WBC DNA quality and quantity was excellent but there were low yields of serum DNA using the BioRobot method.

We explored various options for genotyping, identifying 3 laboratories with systems that could multiplex genotype WBC DNA and possibly serum DNA. Two of these laboratories (#1 and 2) used the SNPlex™ Genotyping System in a 48 multiplex (Applied Biosystems, Foster City, CA) and one laboratory (#3) the homogeneous MassEXTEND (hME) and iPLEX assays, a multiplex MALDI-TOF method capable of genotyping several SNPs simultaneously (Sequenom, San Diego, CA).

SNPlex™ uses Oligonucleotide Ligation Assay/PCR technology for allelic discrimination and ligation product amplification. Genotype information is then encoded into a universal set of dye-labeled, mobility modified fragments, called Zipchute™ Mobility Modifiers, for rapid detection by capillary electrophoresis. The hME and iPLEX assays are based on the annealing of an oligonucleotide primer adjacent to the SNP of interest in PCR amplified DNA. The addition of a DNA polymerase along with a mixture of terminator nucleotides allows extension of the primer through the polymorphic site, generating allele-specific extension products that were resolved by a MALDI-TOF MS. DNA was extracted from serum and WBC of 50 non-informative PCPT participants by NCI Frederick with the Qiagen robot and shipped to each genotyping facility. The purpose of this pilot was to test agreement between (1) the laboratories and (2) between the serum and WBC DNA genotyping in 48 SNPs. The comparison of data across laboratories showed that the two genotyping methods each did not call the second allele for one SNP, rs#1800871 for hME and rs#3730193 for SNPlex. The call rates for laboratories 1 (99.9%) and 2 (99.8%) were higher than for 3 (96.6%)(Table 2). Among the 45 SNPs (minus the 2 SNPs where data were not available for both methods and 1 SNP where data was not available for one site) on 48 subjects with data from all three sites, there were 150 disagreements for the 2250 SNP/participant combinations. Data from laboratory 1 differed from that of the other two laboratories in 5 cases, laboratory 2 differed from the other 2 laboratories in 78 cases and laboratory 3 differed in 66 cases. All three laboratories had problems with multiplex genotyping of the serum DNA. This may be due to the autosomal nature of these loci, i.e. they are present in two copies. Since the DNA had been isolated by a different method than that used previously (Qiagen kits vs. BioRobot), it was unclear whether the problem was the DNA extraction method or multiplex genotyping. Testing with single SNP TaqMan genotyping resulted in good quality data (data not shown) indicating that the robotically isolated serum DNA was appropriate for single SNP genotyping.

A second pilot using Qiagen kits for serum DNA extraction using additional serum from the same cohort of participants was conducted. Comparison of the multiplex genotyping data from the serum DNA to the white blood cell DNA genotyping data indicated poor call rates and very high discordance within each lab for results from serum DNA. For laboratory 1, of the 2,400 SNP/participant combinations, the serum did not yield results in 269 cases. Of the remaining 2,131, there was a disagreement in the genotyping between the 2 sources of DNA in 911 (43%) samples. For laboratory 2, of the 671 SNP/participant combinations with data, 190 (28%) did not agree. Within a participant, the agreement between the serum and WBC DNA ranged from 100% (n=13) to less than 50% (n=8). Serum DNA genotyping data from laboratory 3 were not available for comparison due to poor quality of the data. Single SNP analysis with TaqMan genotyping was run on several (n=10) of these serum samples in laboratory 2 with satisfactory

results. The conclusion was that regardless of the method of isolation, serum DNA does not work well for genotyping with the two multiplex methods tested. Laboratory 2 also tried one small experiment using BioTrove (Woburn, MA), a system that does individual TaqMan assays in nanowells, on a few serum DNAs with remaining material. Unfortunately, this experiment also gave poor results but these results could have been due to the low amount of DNA (data not shown).

CONCLUSIONS

These pilot studies demonstrated that NCI-Frederick's robotic system for isolation of DNA from WBCs and serum provides DNA of a quality suitable for single SNP genotyping. Multiplex genotyping of WBC DNA by both iPLEX or SNPLEX indicated that both methods are cost effective and reliable. We also demonstrated that serum DNA, including that extracted from serum that had been used for hormone analysis, could be used for single SNP genotyping with Taqman. Unfortunately, neither of the multiplexing methods tested provided reliable data with serum DNA. This is probably due to the limited amount and poor quality of the DNA extracted from serum. Future pilot studies should investigate the use of alternate multiplexing methods such as Illumina. These results demonstrate the challenges and limitations of using the new technologies in large-scale trials, especially in PCPT where only serum is available for some subjects. We are now in the process of carefully considering what genotyping to carry out with the limited serum DNA, that can be used only for individual SNP assays. In the ongoing selenium and vitamin E cancer prevention trial (SELECT) we will not have these challenges since white blood cells were collected from all subjects.^{4,5}

Acknowledgements

National Cancer Institute grant P01 CA108964 and CA37429 and NIEHS ES09089. The authors gratefully acknowledge the expert contributions of Drs. Demetrius Alabanes, Juergen Reichardt, Cathy Tangen and Scott Lucia.

References

1. Thompson IM, Goodman PJ, Tangen CM, et al. The influence of finasteride on the development of prostate cancer. *N Engl J Med* 2003;349:215–224. [PubMed: 12824459]
2. Dixon SC, Horti J, Guo Y, et al. Methods for extracting and amplifying genomic DNA isolated from frozen serum. *Nat Biotech* 1998;16:91–94.
3. Hsing AW, Stanczyk FZ, Bélanger A, et al. Reproducibility of Serum Sex Steroid Assays in Men by RIA and Mass Spectrometry. *Cancer Epidemiol Biomarkers Prev* 2007;16:1004–1008. [PubMed: 17507629]
4. Hoque A, Alabanes D, Lippman SM, et al. Molecular epidemiologic studies within the Selenium and Vitamin E Cancer Prevention Trial (SELECT). *Cancer Causes and Control* 2001;12:627–33. [PubMed: 11552710]
5. Kristal AR, King IB, Alabanes D, et al. Centralized blood processing for the selenium and vitamin E cancer prevention trial: effects of delayed processing on carotenoids, tocopherols, insulin-like growth factor-I, insulin-like growth factor binding protein 3, steroid hormones and lymphocyte viability. *Cancer Epidemiol Biomarkers Prev* 2005;14:727–30. [PubMed: 15767358]

Table 1

DNA Quality and Yields

Source	Mean A260/280Ratio	Range	Q1	Q3	DNA Yield	Range	Q1	Q3
NCI-Frederick	1.92	0.36–13.12	1.90	1.99	40.75	0.21–174	23.95	62.45
Commercial Laboratory	1.79	1.55–2.19	1.70	1.85	17.98	2.66–107	1.05	26.45

Notes: The A260/A280 ratio indicates DNA purity and should be >1.8–2.0. Q1 and Q3 represent the 1st and 3rd quartiles, respectively.

Table 2**Multiplex Genotyping of 48 SNPs With White Blood Cell DNA**

Site of Genotyping	Number of SNPs Called	Call Rate (%)	Number Discordant Compared with Other Sites
Laboratory 1 *	47	99.9	5
Laboratory 2 *	46	99.8	78
Laboratory 3 **	47	96.6	66

* SNPlex

** Sequenom hME