Short Technical Report

Automated Fragment Analysis Method for Determining Androgen Receptor CAG Repeat Length

BioTechniques 33:140-143 (July 2002)

D.W. Boorman, Y. Guo, K. Visvanathan¹, K. Helzlsouer¹, and T.G. O'Brien

Lankenau Institute for Medical Research, Wynnewood, PA, and ¹Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

ABSTRACT

Several studies have associated polymorphisms in the androgen receptor gene with the risk of developing hormone-dependent cancers. A highly polymorphic (CAG)_n repeat in exon 1 encodes a polyglutamine tract of varying length. The determination of the number of CAG repeats in the androgen receptor has typically been performed on denaturing polyacrylamide gels with autoradiographic or fluorescent detection of differently sized alleles. Samples run on a capillary electrophoresis-based ABI PRISM[®] 310 Genetic Analyzer gave anomalous results when internal standards supplied by the manufacturer were used. Here we report a modified procedure for androgen receptor allele size determination that can be used on an automated capillary electrophoresis-based DNA sequencer equipped with the appropriate software. The assay is very precise, comparable to DNA sequencing, and is compatible with the latest generation of automated DNA sequencers.

INTRODUCTION

Genetic polymorphisms in genes encoding receptors for steroid hormones are attractive candidates for susceptibility loci for hormone-dependent cancers. The human androgen receptor gene is highly polymorphic, with variable (CAG)_n and (CGN)_n microsatellite repeats in its N-terminal domain (2,5). The length of the polyglutamine tract encoded by the (CAG)_n repeats typically varies from 11 to 31 in Caucasians, although very large alleles (>40 CAG repeats) have been associated with spinal and bulbar muscular atrophy (7). Some reports have associated shorter androgen receptor alleles, which are more transcriptionally active (1), with increased risk for prostate cancer (6,9) and benign prostatic hyperplasia (4). Conversely, short androgen receptor alleles have been shown to protect against breast cancer (3). It seems likely that the effect of androgen receptor genetic polymorphisms on hormone-dependent cancer risk may be tissue-dependent and could also be influenced by common polymorphisms in other susceptibility genes.

The standard assay for determining CAG repeat length involves amplifying by PCR the region containing the (CAG)_n repeat sequence, followed by electrophoresis in denaturing polyacrylamide gels (2,5). Depending on whether the PCR products were radioactively or fluorescently labeled, detection is accomplished by autoradiography or fluorescence detection on an automated DNA sequencer equipped with appropriate software. Allele size calls are typically made by comparison with a stan-

dard curve composed of fragments containing CAG repeats of known length (determined by DNA sequencing). With the shift of automated DNA sequencing and fragment analysis from gel-based systems to capillary electrophoresis-based systems, a method for androgen receptor allele size determination based on the latter technology is needed. Here we describe such a method, along with some caveats regarding its use.

MATERIALS AND METHODS

Subjects

We obtained 482 samples of DNA isolated from the blood of Caucasian men participating in a nested case/control study of genetic risk factors for prostate cancer (K. Visvanathan et al., manuscript in preparation). The Institutional Review Board of Johns Hopkins University approved the study. Laboratory workers were blinded to case/control status.

Sample Preparation for Automated Fragment Analysis

DNA samples were amplified by PCR using forward primer 5'-FAM-TCCAGAATCTGTTCCAGAGCGTGC-3' and reverse primer 5'-GCTGTGAA-GGTTGCTGTTCTCAT-3', where FAM is 6-carboxy-fluorescein. The reaction mixture contained 10 ng DNA, 12.5 µL 2× Master Amp[™] PCR PreMixG (Epicentre Technologies, Madison, WI, USA), 10 pmol each primer, and 0.5 U *Taq* DNA polymerase in a total volume of 25 µL. PCR conditions were 95°C

for 2 min, followed by 25 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 1 min, and 72°C for 3 min.

Preparation of Androgen Receptor-Specific Size Standards for Fragment Analysis

Because of anomalous results from initial experiments using internal standards supplied with the ABI PRISM® 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA), a series of six androgen receptor-specific size standards were prepared using samples whose exact repeat lengths were determined by DNA sequencing. These standards contained 11, 14, 21, 25, 28, and 31 CAG repeats and were synthesized by PCR amplification, using a ROX-labeled forward PCR primer in the reaction mixture, in which ROX is 6-carboxy-x-rhodamine. This allows the unknown samples (labeled with

FAM) to be distinguished from the internal size standards. After separate amplification of each size standard, a custom marker mixture was prepared by combining equal amounts of each size standard into a single solution.

Automated Fragment Analysis and DNA Sequencing

For automated fragment analysis on the ABI PRISM 310 Genetic Analyzer, the samples were prepared by combining 1 µL reaction from unknowns, 12 µL deionized formamide, and 1 µL custom marker mixture described above. This mixture was heated for 2 min at 95°C and then placed on ice until sample loading. Samples were run on a capillary containing POP-4 polymer and data analyzed with GeneScan® analysis software version 2.0.2 (Applied Biosystems). For validation of the GeneScan results, a total of 66 samples

were DNA-sequenced to determine the exact CAG repeat lengths. The androgen receptor allele sizes called by GeneScan were consistently 3.4 ± 0.18 ($\bar{x} \pm SD$) base pairs shorter than the DNA sequence data. This is due to the slightly slower mobility of ROX-labeled versus FAM-labeled fragments of the identical DNA sequence. Therefore, 3 bp were added to each GeneScan size call in reporting the results.

RESULTS AND DISCUSSION

Initial results from an analysis of a subset of 482 DNA samples from a nested core/control study of prostate cancer indicated a lower than expected median CAG repeat size ($n = 17$) compared to the value for North American Caucasian men ($n = 21$ or 22 ; References 4, 6, and 9) previously reported. These results were obtained with a

Molecular Diagnostic Techniques

standard GeneScan protocol using a internal standard matrix supplied by the manufacturer. The suspicion of spurious results was confirmed by DNA sequencing of some of the samples: there was a consistent 15 bp (equivalent to five CAG repeats) smaller allele size determination by GeneScan versus DNA sequencing results. Even though the difference was a consistent 15-bp over a wide range of CAG repeat lengths, this rather large error in accuracy prompted us to

Figure 1. Predicted secondary structure of an androgen receptor DNA sequence containing 19 CAGs. Structure was predicted with the mfold software package (version 3.0) using nearest-neighbor thermodynamics (8). Software is available at <http://Bioinfo.math.rpi.edu/~Zukerm/>. GC base pairs are indicated with gray dots, AT base pairs are indicated with black dots. A $\Delta G = -68.3$ kcal/mol at 25°C was computed by the software.

Table 1. Comparison of DNA Sequencing versus Automated Fragment Analysis Methods for CAG Repeat Length Determination

DNA Sequencing			Automated Fragment Analysis		
Sample No.	bp	No. of CAG	bp	No. of CAG	Difference ^a
1	255	11	251.43	11	3.57
2	255	11	251.35	11	3.65
3	264	14	260.47	14	3.53
4	270	16	266.47	16	3.53
5	279	19	275.59	19	3.41
6	279	19	275.85	19	3.15
7	285	21	281.51	21	3.49
8	285	21	281.46	21	3.54
9	294	24	290.46	24	3.54
10	294	24	290.61	24	3.39
11	300	26	296.73	26	3.27
12	306	28	302.61	28	3.39
13	306	28	302.69	28	3.31
14	315	31	311.76	31	3.24
15	315	31	311.75	31	3.25

A total of 66 DNA samples from a case/control study (total samples = 482) were analyzed by both methods to determine CAG repeat lengths. As described in the text, the number of CAGs from the automated fragment analysis method was adjusted upwards by 1 from the absolute base call determined by GeneScan. Table 1 shows 15 representative samples ranging from 11 to 31 CAG repeats.

^aDifference is base pairs between DNA sequencing result and absolute call by GeneScan.

determine the basis for the anomalous mobility of these fragments in this capillary-based system.

A computer-based analysis of the secondary structure of the PCR-amplified fragment revealed extensive base-

pairing in the CAG repeat region (Figure 1). Because of the very large negative free energy of this structure (-68.3 kcal/mol), it is very unlikely for this secondary structure to melt completely, given the sample preparation

Figure 2. Mobility of androgen receptor gene fragments containing different CAG repeat lengths labeled with either ROX or FAM. Androgen receptor-specific PCR fragments containing repeat lengths of 11, 14, and 21 CAGs were labeled with either ROX (red) or FAM (blue). All six samples were mixed together, run on the ABI PRISM310 Genetic Analyzer, and analyzed by GeneScan software package modified as described in the Materials and Methods section. On the electropherogram, from left to right, each set of peaks corresponds to fragments containing 11, 14, and 21 CAG repeats, respectively. The FAM-labeled fragments migrated 3.55, 3.56, and 3.51 bp quicker than the ROX-labeled fragments for CAG repeat lengths of 11, 14, and 21, respectively.

and running conditions of the ABI PRISM 310 capillary electrophoresis system used.

To improve the accuracy of the androgen receptor allele size calls by GeneScan, we synthesized new fluorescent size standards using the androgen receptor gene sequence containing CAG repeat lengths from 11 to 31. The fluorescent label for these standards was ROX, consistent with the manufacturer-supplied size standards. With this modification, a series of 66 samples from a case/control study were analyzed by automated fragment analysis and sequenced to determine absolute allele size. Table 1 gives the results of these analyses. For each sample, the GeneScan size call was approximately 3 bp smaller than the size determined by DNA sequencing (3.4 ± 0.18 bp, range 2.4–3.8; $n = 66$). The reason for the 3-bp difference is the effect of the fluorescent dye used: the androgen receptor size standards are labeled with ROX, while the unknown samples are labeled with FAM. Figure 2 shows representative electropherograms of identical androgen receptor gene fragments, labeled with either ROX or FAM, showing an approximate and consistent 3-bp difference. When the rest of the samples from the nested case control study ($n = 422$) were assayed by this modified fragment analysis method, the mean CAG repeat length for the whole sample set ($n = 482$) was 21.9 ± 3.1 , which is consistent with previously reported results for North American Caucasians (4,6,9).

In summary, we have developed a modified automated fragment analysis for androgen receptor allele size determination that uses the capillary electrophoresis-based ABI PRISM Genetic Analyzer series and GeneScan software analysis program. This assay avoids the use of radioactivity and polyacrylamide gel-based analytic procedures. With the currently available DNA sequencers (ABI PRISM3100 or equivalent), the assay is relatively inexpensive (\$1.50/sample) and has a high-throughput capability (capable of processing several hundred samples in a 24-h period). These characteristics should facilitate future large-scale studies of the relationship between androgen receptor allele sizes and disease susceptibility.

ACKNOWLEDGMENTS

We thank Dr. Dan Rosson for his critical review of the manuscript and Rosalyn Gilchrist and Loretta Rossino for editorial assistance. This work was supported by National Institutes of Health grant nos. R01 ES09899 to T.G.O. and R03 CA88180 to K.H.

REFERENCES

1. Chamberlain, N.C., E.D. Driver, and R.L. Miesfield. 1994. The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Res.* 22:3181-3186.
2. Edwards, A., H.A. Hammond, L. Jin, C.T. Caskey, and R. Chakraborty. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12:241-253.
3. Giguere, Y., E. Dewailly, J. Brisson, P. Ayotte, N. Laflamme, A. Demers, V.-I. Forest, S. Dodin, et al. 2001. Short polyglutamine tracts in the androgen receptor are protective against breast cancer in the general population. *Cancer Res.* 61:5869-5874.
4. Giovannucci, E., E.A. Platz, M.J. Stampfer, A. Chan, K. Krithivas, I. Kawachi, W.C. Willett, and P.W. Kantoff. 1999. The CAG repeat within the androgen receptor gene and benign prostatic hyperplasia. *Urology* 53:121-125.
5. Irvine, R.A., M.C. Yu, R.K. Ross, and G.A. Coetzer. 1995. The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. *Cancer Res.* 55:1937-1940.
6. Kantoff, P., E. Giovannucci, and M. Brown. 1998. The androgen receptor CAG repeat polymorphism and its relationship to prostate cancer. *Biochim. Biophys. Acta* 1378:C1-C5.
7. La Spada, A.R., E.N. Wilson, D.B. Lubohn, A.E. Harding, and K.H. Fischback. 1991. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 352:77-79.
8. SantaLucia, Jr., J. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA* 95:1460-1465.
9. Stanford, J.L., J.J. Just, M. Gibbs, K.G. Wicklund, C.L. Neal, B.A. Blumenstein, and E.A. Ostrander. 1997. Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk. *Cancer Res.* 57:1194-1198.

Received 17 December 2001; accepted 11 February 2002.

Address correspondence to:

Dr. Thomas G. O'Brien
Lankenau Institute for Medical Research
100 Lancaster Avenue
Wynnewood, PA 19096, USA
e-mail: obrient@mlhs.org