ELSEVIER

Clinical Biochemistry, Vol. 32, No. 5, 327–332, 1999

0009-9120/99/$–see front matter

PII S0009-9120(99)00023-5

Rapid and Accurate Determination of $(CAG)_n$ Repeats in the Androgen Receptor Gene Using Polymerase Chain Reaction and Automated Fragment Analysis

BUPINDER S. BHARAJ,[1] EVYENIA J. K. VASSILIKOS,[1] and ELEFTHERIOS P. DIAMANDIS[1,2]

[1]Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada M5G 1X5, and [2]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada M5G 1L5

Objectives: To develop and evaluate a new method for determination of the CAG repeat length in Exon 1 of the androgen receptor gene.
Design and Methods: The method is based on PCR amplification of a DNA region encompassing the repeats and analysis of the length of the PCR product on a sequencing gel. One of the PCR primers was labeled with Cy5.5 fluorescent dye to facilitate detection after laser excitation. We used a fully automated system for electrophoretic separation of the PCR product and accurate sizing of the length of the PCR product using fragment analysis.
Results: The major advantages of the new technique are its simplicity, speed, accuracy, and reproducibility. Analysis of the CAG repeats in genomic DNAs from 18 males indicated that they were all hemizygous with a mean CAG repeat number of 22 (range 20–30 repeats). Among 60 DNAs from females, 16 were homozygous and 44 were heterozygous. The repeat length ranged from 17–30 with a mean of 22. In both males and females, the distribution of CAG repeats was bimodal.
Conclusion: We anticipate that this improved method for CAG repeat analysis will find applications in clinical studies involving prostate and breast cancer patients.

KEY WORDS: gene polymorphisms; CAG repeats; androgen receptor; prostate cancer; fragment analysis.

Introduction

Prostate cancer is the most frequently diagnosed cancer of males and it is the second most common cause of cancer-related deaths in Americans (1). The incidence of this cancer differs significantly among various racial-ethnic groups (2). Androgens play a direct role in normal and malignant growth of prostate cells via the androgen receptor (AR) gene which binds dihydrotestosterone and transactivates the transcription cascade (3,4). It has been established that the first exon of the AR gene contains two polymorphic trinucleotide repeat sequences, CAG and CGG (5,6). The CAG repeats encode for a polyglutamine tract and the CGG repeats for a polyglycine tract. The changes in the CAG repeat lengths have been correlated with reduced expression or functional alterations in androgen receptor action (7,8). The diseases caused by the trinucleotide repeat length variations exhibit a phenomenon called anticipation, which predicts the increase in the severity of the disease with earlier onset of symptoms in successive generations. The severity and the onset of the disease are correlated with the size of the repeats on the mutant alleles, the onset being inversely related with the size of the expansion (9). Among the pathological conditions associated with CAG repeat expansions is the disease neurodegenerative spinal and bulbar muscular atrophy (SBMA) or Kennedy's disease, which is associated with low virilization, oligospermia or azoospermia, testicular atrophy, and infertility (10). The contraction of the CAG repeat lengths has also been associated with higher incidence, aggressiveness and early onset prostate cancer (11–16).

In general, highly polymorphic short DNA repeats (microsatellites) with uniform distribution within the human genome may constitute valuable genetic markers. The detection of the number of CAG repeat lengths could, therefore, allow either diagnosis, risk assessment or evaluation of aggressiveness of genetic diseases, including prostate cancer. The use of microsatellite analysis for possible prediction of an associated disease state has been hampered by some technical drawbacks. PCR amplification followed by electrophoresis and Southern blotting is a labour intensive technique. The use of radioisotopes poses a health hazard. In addition, radioactive analysis or silver staining of polyacrylamide gels produces shadow bands around the allele bands due to the

Correspondence: E. P. Diamandis, MD, PhD, FRCPC, Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, Canada M5G 1X5. E-mail: ediamandis @ mtsinai.on.ca

Manuscript received January 11, 1999; revised and accepted March 12, 1999.

nature of the repeat amplification (see subsequent text), which, in turn, makes it difficult to differentiate between the main and shadow bands when these have similar intensities (17). In addition, manual allele sizing is not very accurate. This is particularly important in microsatellite analysis where alleles can differ by as little as two base pairs. Finally, there is the problem of "stutters" caused by the misincorporation of bases towards the end of the repeats due to "slippage" of the Taq polymerase. This "stuttering" causes confusion in determining the correct peaks for the purpose of molecular sizing. In this article, we report a new method which can accurately determine the size of the CAG repeat in Exon 1 of the AR gene. This method is characterized by simplicity, speed, automation and accuracy and is suitable for CAG repeat analysis in large numbers of samples.

Methods

DNA extraction from whole blood

DNA was extracted from whole blood using the Qiagen QIAamp blood kit, following the manufacturer's recommendations (Qiagen, Chatsworth, CA, USA). Briefly, after lysis of cells, the DNA is entrapped onto a silica membrane, washed and then eluted in a buffer solution. The procedure takes approximately 1 hour to complete. DNA was quantified by absorbance measurements at 260 nm and stored at 4° C until analysis.

DNA extraction from serum

Recently, a number of methods have been evaluated for isolating traces of genomic DNA from human serum, suitable for PCR analysis (18,19). We evaluated two methods: (a) frozen sera were thawed and then boiled at 100° C for 5 min. The samples were then spun at 15,000g and the supernatant was used directly for PCR amplification (18); (b) we used the QIAamp blood kit with serum samples, as described (19).

PCR Amplification

The paired primer sequences (5′-TCC AGA ATC TGT TCC AGA GCG TGC-3′ and 5′ -GCT GTG AAG GTT GCT GTT CCT CAT-3′) flanking the CAG repeat region in Exon 1 of the AR gene were designed based on the AR sequence deposited in Genbank by Marcelli *et al.* (20) (accession # M35844).

The primers were fluorescently labeled with Cy5.5 dye as described elsewhere (21). DNA was amplified using these primers to generate fragments of the N-terminal domain of the AR gene (approximately 279 bp product). PCR amplification was performed in a final volume of 25 μL containing approximately 100 ng of DNA template, 10 mM trimethamine (tris) buffer pH 8.3, 50 mM KCI, 2.5 units of Taq polymerase (Boehringer Mannheim GmbH, Mannheim, Germany), 250 μM deoxynucleoside triphosphates, 1.5 mM $MgCl_2$, 5 μL solution Q (Qiagen) and 10 pmoles of the primers. Only one of the two primers was labeled with fluorescent dye Cy5.5 in order to facilitate the detection of the PCR product. The thermal cycling profile on a Perkin-Elmer 9600 instrument consisted of a 30 s denaturation step at 95° C, a 60 s annealing step at the optimized temperature of 65° C and a 60 s extension step at a temperature of 72° C, for a total of 35 cycles. Each PCR was initiated with a 5 min denaturation step at 95° C and terminated with a 5 min extension step at 72° C. The number of PCR cycles can be decreased or increased if the intensities of the fragment bands are relatively very high or low, respectively. The success of the PCR was verified by running an 8 μL aliquot of the PCR product on a 0.8% agarose gel containing ethidium bromide.

Preparation of molecular weight markers

Fluorescently labeled standard size markers (SSM) were produced by PCR amplification of a segment of the cDNA sequence of the PSA gene deposited in the Genbank, accession # X07730. One of the primers (upper primer) was labeled at its 5′ end with the fluorescent dye Cy5.5. The sequences of the primers were as follows:

Upper primer: 5′-TGC GCA AGT TCA CCC TCA-3′ (common for the two lower primers shown below).

200 bp marker: 5′-TCC GGT AAT GCA CCA CCT T – 3′

400 bp marker: 5′-TTT AAG AAA CAC TCT CGT C – 3′

PCR was performed as previously described (22).

Sample preparation

Two microliters of the PCR product were mixed with 4 μL of a gel loading buffer (Visible Genetics Inc., Toronto, Ontario, Canada) to which we also added 2 μL each of the two molecular weight markers, prepared by PCR as described above. This mixture was denatured at 95° C for 2 min, placed on ice and then loaded onto the sequencing gel (see subsequent text).

Sequencing apparatus–analysis

The MicroGene Blaster™ (Visible Genetics Inc.) automated sequencer is a compact device containing a visible light laser diode (675 nm) as the excitation source. The laser power is 0.5 mW per lane. The machine applies high voltage across the sequencing gel and can complete a run in 30 min, resolving approximately 300 bases of sequence. The instrument accepts 14 × 14 cm polyacrylamide sequencing gels with a 50 μm thickness. The gel is cast between two disposable glass plates (MicroCel™ cassette). For this application (fragment analysis), each gel is capable of analyzing 16 patient samples (16 lanes). The cassette is filled with 6% acrylamide solution

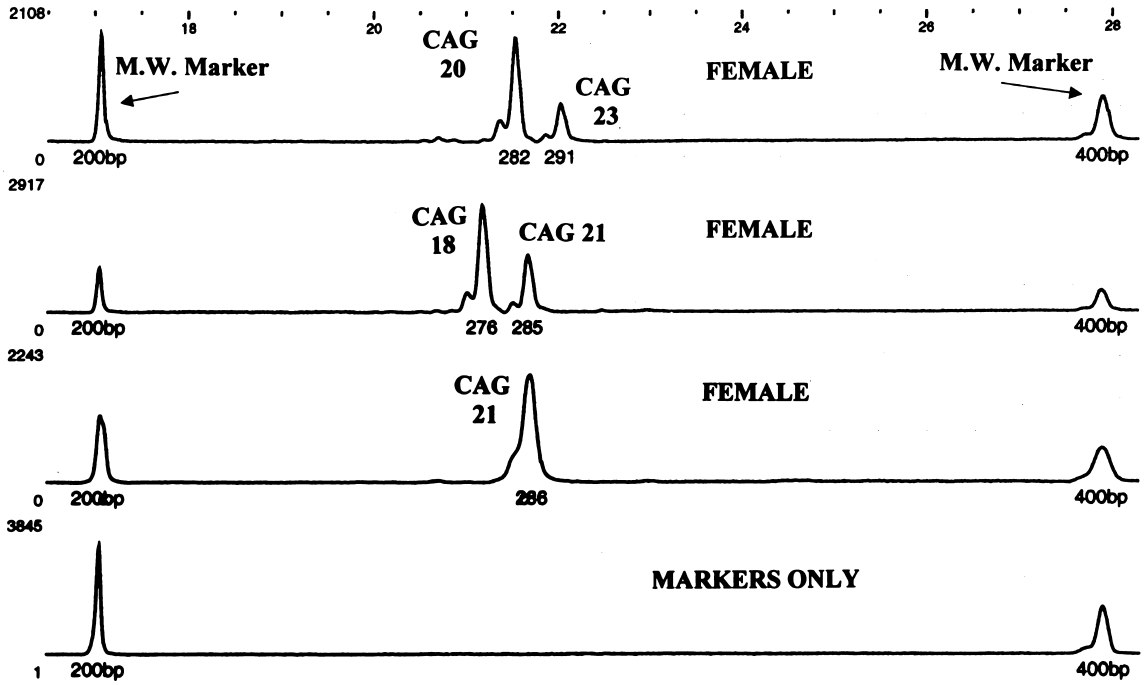

Figure 1 — Analysis of CAG repeats in three DNAs from females. M.W., molecular weight. The numbers below the peaks represent lengths of PCR products in bases. A PCR product of 282 bases represents 20 CAG repeats.

containing a photoinitiator which is activated by UV-light on a special unit (Gel Toaster™ polymerising unit). The process of gel filling and casting takes about 5 min. Once the gel is ready, it is positioned into the sequencer and the buffer chambers filled with 150 mL of 1 × TBE (Tris-Borate-EDTA) buffer, pH 8.3. The gel temperature and voltage are set at 54° C and 1200 V, respectively. A 5 min pre-run is initiated to bring the gel temperature to the set values. Each of the 16 lanes is loaded with 2 μL of denatured mixture of unknown DNA, molecular weight markers and loading buffer. For quality control purposes, every fourth lane is loaded with molecular weight standards only. In total, 12 samples and four control lanes can be loaded on each MicroCel plate. The laser is switched on and the electrophoresis started. Laser intensity can be varied accordingly in order to bring the signals within the dynamic range of the detectors. If the signals are too intense, the sample dilution should be increased appropriately. At the end of the run (30 min), the data can be analyzed by using the Fragment Tool™ of the Gene Objects™ software (Visible Genetics). The unknown PCR product in the three lanes can be aligned with the fourth lane which contains the size standards only. This minimizes any discrepancy that may have arisen during the run and increases accuracy of determining PCR product length. A more detailed description of the Visible Genetics sequencing system has been recently published (21).

Clinical studies

In order to evaluate the capabilities of the new method, we analyzed DNA obtained from 18 blood samples from males and from 60 blood samples from females.

Results

Our method is based on PCR amplification of the region of the AR gene containing the CAG repeats and then analysis of the length of the PCR product on a high resolution gel, capable of resolving length differences of 1 base. A DNA containing 19 CAG repeats will produce a PCR product of 279 bp. For every additional CAG repeat, the length will increase by 3 bases. In Figure 1, we show a representative example of 3 DNAs from females along with a lane containing only the molecular weight sizing standards. Because the AR gene is localized on chromosome X, females can be either heterozygotes (lanes 1 and 2) or homozygotes (lane 3). The lengths of the CAG repeats are 20/23 (lane 1), 18/21 (lane 2) and 21/21 (lane 3). In Figure 2, we show representative examples for three males. Because males have only one chromosome X, they produce a single PCR band in all cases (hemizygous state). The number of CAG repeats are 22 (lane 1), 23 (lane 2) and 18 (lane 3). In order to assess precision, we have repeated the analysis of various DNA samples from 3–10 times. In all cases, the calculated CAG repeat length was the same, and the variation in PCR product length was ±1 base pair, indicating that our method is highly precise.

We have also examined the possibility of amplifying genomic DNA isolated from serum. Using the two alternative extraction methods described (18, 19), we were able to amplify DNA and quantify CAG repeats in approximately 30% of the samples. If the

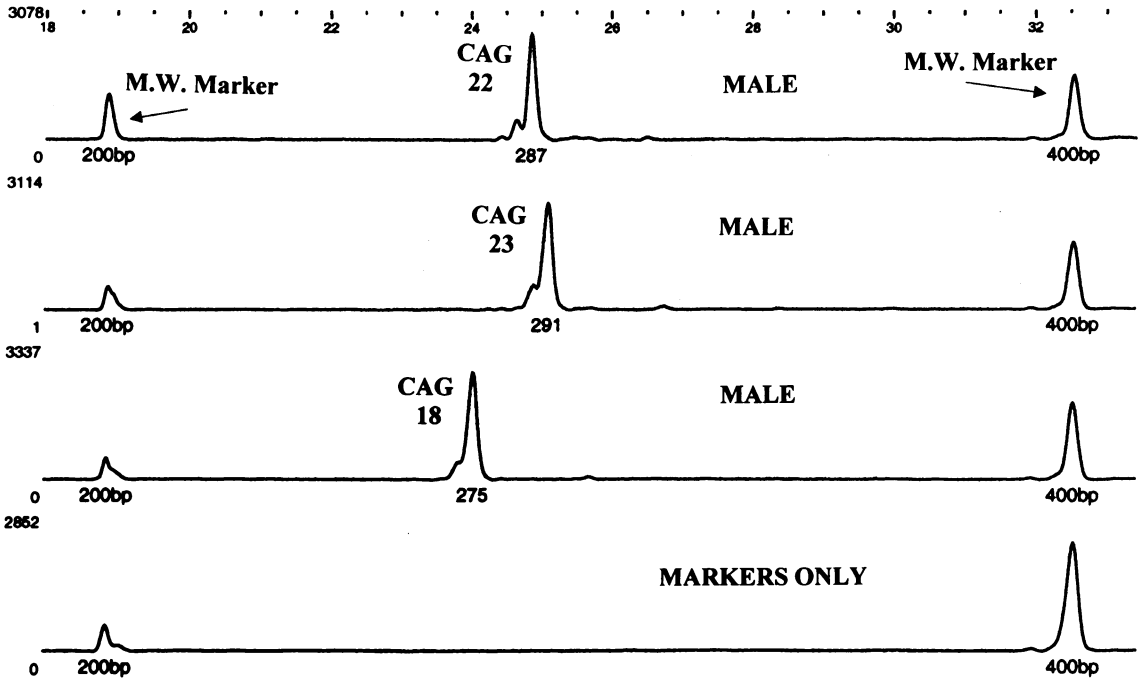

Figure 2 — Analysis of CAG repeats in three DNAs from males. For more details see legend of Figure 1.

DNA was amplifiable, the number of CAG repeats calculated from serum DNA or blood leucocyte DNA was identical in all cases.

Analysis of 18 DNAs isolated from whole blood of normal males (ages 40–50 years) indicated that they were all hemizygous, as expected. CAG repeat length ranged from 20–30 with a mean of 22 and a median of 21. The frequency distribution is shown in Figure 3. Among 60 normal females, 16 were homozygous (27%), and 44 were heterozygous (73%). Among the homozygotes, the range of CAG repeats was between 19–25 with a mean of 22 and a median of 22. Among the heterozygotes, the CAG repeats ranged from 17 to 30 with a mean of 22 and a median of 21. The frequency distribution of the alleles in homozygotes, heterozygotes and the whole female population is shown in Figure 4. In panel 4C, we have considered that each homozygous female possessed two identical alleles (60 women = 120 alleles).

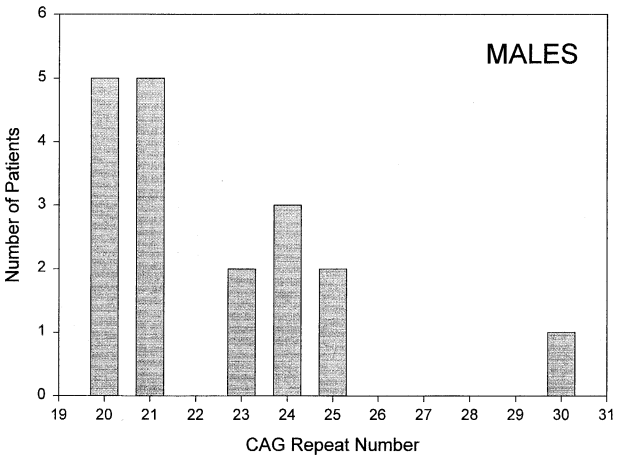

Figure 3 — Frequency distribution of CAG repeats in 18 males.

Discussion

Analysis of polymorphic repeats in genomic DNA is useful in various areas of genetic analysis. The CAG/polyglutamine tract is implicated in the pathogenesis of at least eight different neurodegenerative disorders. Among these diseases is spinal and bulbar muscular atrophy (Kennedy's disease) (9). Other repeats associated with length expansion in various diseases include CGG (fragile X syndrome), CTG (myotonic dystrophy) and GAA (Friedreich's ataxia) (*9*).

Recently, the CAG/polyglutamine tract length polymorphism in the androgen receptor gene has attracted much attention. It is believed that length variation in this repeat alters the transcriptional activity of the androgen receptor and this may affect the transcription of genes that are regulated by androgens. In turn, this may contribute to the biology of prostate cancer that is dependent on androgens (4,23).

Studies of various ethnic groups have established that the length of the CAG repeat is generally shorter in populations that are at high risk for developing prostate cancer (e.g., African-Americans), in comparison to low risk populations (e.g., Asians) (4,13). Indeed, this is in accord with the finding that AR activity or transcription is reduced when the CAG repeat is expanded (7,8,24). Other authors have shown that shorter CAG repeats in the

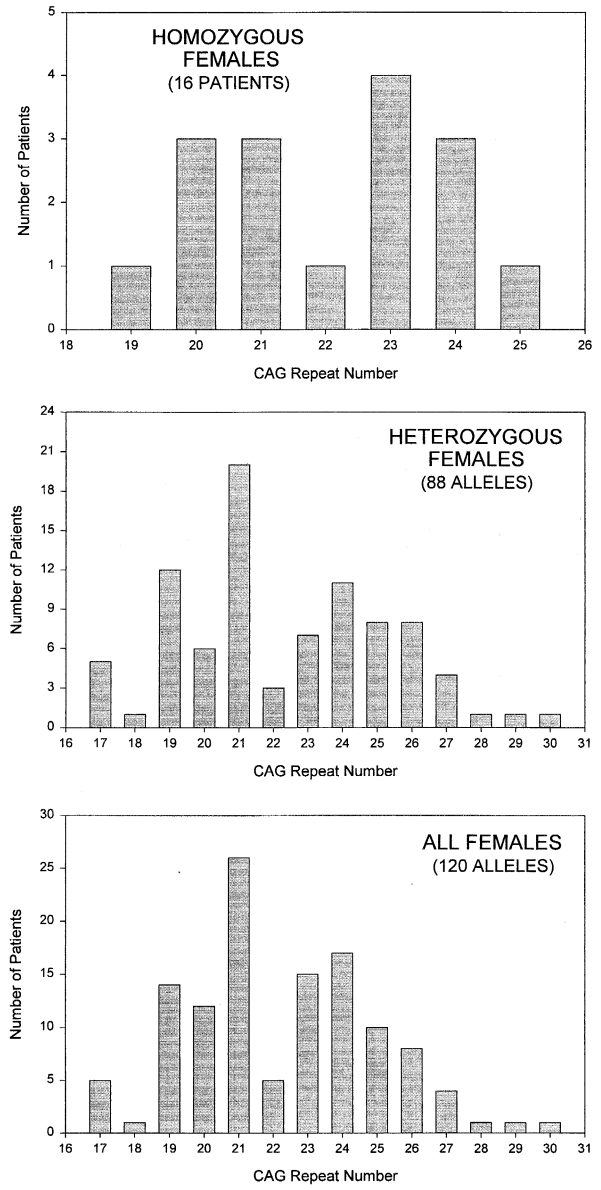

Figure 4 — Frequency distribution of CAG repeats in 60 females of which, 16 were homozygotes.

AR gene is associated with early onset prostate cancer (11) and more aggressive disease (15,16). These data suggest that the determination of CAG repeats may be a useful indicator of risk for developing prostate cancer and/or a marker of disease aggressiveness. Recently, serum IGF-1 levels were also found to be associated with increased risk of developing prostate cancer (25).

In this article, we have attempted to improve on the methodology for determining the length of CAG repeats. Our method capitalizes on the recent introduction of new sequencing instruments, which produce fast results with minimal labor. The advantages of our method are as follows:

1. We have optimized our PCR reaction to avoid polymerase "stuttering" as much as possible. Thus, shadow bands around the major PCR band are small and do not interfere with the CAG repeat sizing (Figures 1 and 2). This "stuttering" is a serious problem in other published PCR-based methods (17).
2. Our method is non-isotopic and it is based on fluorescence detection.
3. We used a sequencing gel to ensure length resolution of 1 base. The sequencing cassette is disposable and can be prepared in ~5 min.
4. The electrophoresis run is completed within 30 min and 12 samples along with 4 sizing standards can be run simultaneously on one gel.
5. Sizing is automatic and accurate due to alignment of the molecular weight standards with the samples, using a special fragment analysis software.

CAG repeat sizing with this method is accurate and highly precise. While we were able to determine the CAG repeat size in all DNAs isolated from whole blood, only ~30% of sera tested contained enough DNA to allow amplification and detection. However, when the serum DNA was amplifiable, the results were identical to those of the corresponding DNA from blood.

In males and females there is a bimodal distribution of CAG repeat lengths with peaks around 20–21 repeats and then around 23–25 repeats. These data are in accordance with those reported for males by other groups (11,13). To our knowledge, this is the first report describing allele frequency of CAG repeats in women. We are now in the process of using this method for clinical studies involving prostate and breast cancer patients.

References

1. Parker S, Tong T, Bolden S, Wingo PA. Cancer Statistics, 1996. *CA Cancer J Clin* 1996; **46**: 5–27.
2. Bernstein L, Ross RK. *Cancer in Los Angeles County. A portrait of incidence and mortality, 1972–1987.* Los Angeles: University of Southern California Press, 1991.
3. Coetzee GA, Ross RK. Prostate cancer and the androgen receptor. *J Natl Cancer Inst* 1994; **86**: 872–3.
4. Ross RK, Pike MC, Coetzee GA, *et al*. Androgen metabolism and prostate cancer: establishing a model of genetic susceptibility. *Cancer Res* 1998; **58**: 4497–504.
5. Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 1992; **12**: 241–53.
6. Sleddens HFBM, Oostra BA, Brinkmann AO, *et al*. Trinucleotide (GGN) repeat polymorphism in the androgen receptor gene. *Hum Mol Genet* 1993; **2**: 273.
7. Choong CS, Kemppainen JA, Zhou ZX, Wilson EM. Reduced androgen receptor gene expression with first exon CAG repeat expansion. *Mol Endocrinol* 1996; **10**: 1527–35.
8. Chamberlain NL, Driver ED, Miesfeld RL. The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Res* 1994; **22**: 3181–6.

9. Koshy BT, Zoghbi HY. The CAG/polyglutamine tract diseases: Gene products and molecular pathogenesis. *Brain Pathol* 1997; **7**: 927–42.
10. Arbizu T, Santamaria J, Gomez JM, Quilez A, Serra JP. A family with adult spinal bulbar muscular atrophy, X-linked inheritance associated with testicular failure. *J Neurol Sci* 1983; **59**: 371–82.
11. Hardy DO, Scher HI, Bogenreider T, *et al.* Androgen receptor CAG repeat lengths in prostate cancer: Correlation with age of onset. *J Clin Endocrinol Metab* 1996; **81**: 4400–5.
12. Stanford JL, Just JJ, Gibbs M, et al. Polymorphic repeats in the androgen receptor gene: Molecular markers of prostate cancer risk. *Cancer Res* 1997; **57**: 1194–8.
13. Irvine RA, Yu MC, Ross RK, Coetzee GA. The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. *Cancer Res* 1995; **55**: 1937–40.
14. Platz EA, Giovannucci, E, Dahl DM, *et al.* The androgen receptor gene GGN microsatellite and prostate cancer risk. *Cancer Epidemiol Biomarkers Prev* 1998; **7:** 379–84.
15. Giovannucci E, Stampfer MJ, Krithivas K, *et al.* The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. *Proc Natl Acad Sci USA* 1997; **94**: 3320–3.
16. Hakimi JM, Schoenberg MP, Rondinelli RH, Piantadosi S, Barrack ER. Androgen receptor variants with short glutamine or glycine repeats may identify unique subpopulations of men with prostate cancer. *Clin Cancer Res* 1997; **3**: 1599–608.
17. Muglia M, Leone O, Annesi G, *et al.* Nonisotopic method for accurate detection of (CAG) repeats causing Huntington disease. *Clin Chem* 1996; **42**: 1601–3.
18. Sartor O, Zheng Q. Determination of CAG repeat length in the androgen-receptor gene using frozen serum. *Urology* 1997; **49**: 301–4.
19. Dixon SC, Horti J, Guo Y, Reed E, Figg WD. Methods for extracting and amplifying genomic DNA isolated from frozen serum. *Nature Biotechnol* 1996; **16**: 91–4.
20. Marcelli M, Tilley WD, Wilson CM, Griffin JE, Wilson JD, McPhaul MJ. Definition of human androgen receptor gene structure permits the identification of mutations that cause androgen resistance: Premature termination of the receptor protein at amino acid residue 588 causes complete androgen resistance. *Mol Endocrinol* 1990; **4**: 1105–16.
21. Bharaj B, Angelopoulou K, Diamandis EP. Rapid sequencing of the p53 gene with a new automated DNA sequencer. *Clin Chem* 1998; **44**: 1397–403.
22. Zarghami N, Diamandis EP. Detection of prostate specific antigen mRNA and protein in breast tumors. *Clin Chem* 1996; **42**: 361–6.
23. Lopez-Otin C, Diamandis EP. Breast and prostate cancer: an analysis of common epidemiological, genetic and biochemical features. *Endocr Rev* 1998; **19**: 365–96.
24. Tut TG, Ghadessy FJ, Trifiro MA, Pinsky L, Yong EL Long polyglutamine tracts in the androgen receptor are associated with reduced trans-activation, impaired sperm production and male infertility. *J Clin Endocrinol Metab* 1997; **82**: 3777–82.
25. Chan JM, Stampfer MJ, Giovannucci E, et al. Plasma insulin-like growth factor-1 and prostate cancer risk: A prospective study. *Science* 1998; **279**: 563–6.