Non-linear association between androgen receptor CAG and GGN repeat lengths and reproductive parameters in fertile European and Inuit men.

Brokken, Leon; Rylander, Lars; Jönsson, Bo A; Spanò, M; Pedersen, H S; Ludwicki, J K; Zviezdai, V; Bizzaro, D; Manicardi, G C; Toft, G; Bonde, J P; Giwercman, Aleksander; Giwercman, Yvonne

Published in:
Molecular and Cellular Endocrinology

DOI:
10.1016/j.mce.2013.03.005

2013

Non-linear association between androgen receptor CAG and GGN repeat lengths and reproductive parameters in fertile European and Inuit men

Brokken L.J.S.[a†], Rylander L.[b], Jönsson B.A.[b], Spanò M.[c], Pedersen H.S.[d], Ludwicki J.K.[e], Zviezdai V.[f], Bizzaro D.[g], Manicardi G.C.[h], Toft G.[i], Bonde J.P.[j], Giwercman A.[k], Lundberg Giwercman Y.[a]

[a] Department of Reproductive Molecular Medicine, Lund University, SE-20502 Malmö, Sweden

[b] Division of Occupational and Environmental Medicine, Lund University, SE-22185 Lund, Sweden

[c] Laboratory of Toxicology, Unit of Radiation Biology and Human Health, ENEA Casaccia Research Centre, Via Anguillarese 301, IT-00123 Rome, Italy

[d] Centre for Arctic Environmental Medicine, Post box 570, DK-3900 Nuuk, Greenland

[e] Department of Environmental Toxicology, National Institute of Public Health-National Institute of Hygiene, Chocimska 24, PL-00791 Warsaw, Poland

[f] Social Medicine and Organization of Public Health, Kharkiv National Medical University, UA-61022 Kharkiv, Ukraine

[g] Department of Life and Environmental Sciences, Polytechnic University of Marche Via Brecce Bianche, IT-60131 Ancona, Italy

[h] Department of Life Sciences, University of Modena and Reggio Emilia, Reggio Emilia, Italy

[i] Danish Ramazzini Center, Department of Occupational Medicine, Aarhus University Hospital, Nørrebrogade 44, DK-8000 Aarhus, Denmark

[j] Department of Occupational and Environmental Medicine, Copenhagen University Hospital, Bispebjerg Bakke 23, DK-2400 Copenhagen, Denmark

[k] Reproductive Medicine Centre, Skåne University Hospital, Malmö, Lund University, SE-20502 Malmö, Sweden

[†] *Corresponding author:* leon.brokken@med.lu.se

Department of Reproductive Molecular Medicine

Clinical Research Centre, Lund University

Jan Waldenströms gata 35

SE-205 02 Malmö

SWEDEN

Tel. +46 40 391104

Fax. +46 40 391222

Abstract

Recently the dogma that there is an inverse linear association between androgen receptor (AR) CAG and GGN polymorphisms and receptor activity has been challenged. We analysed the pattern of association between 21 male reproductive phenotypes and AR CAG/GGN repeat lengths in 557 proven-fertile men.

A linear association was only found between sperm DNA fragmentation index (DFI) and CAG length, and between inhibin B and GGN length. Men with longer CAG then the reference (22-24), had higher oestradiol levels, whereas men with shorter CAG stretches had a higher DFI and a higher proportion of Fas-positive germ cells. Subjects with either short or long CAG had increased seminal levels of prostate-specific antigen and neutral α-glucosidase activity. Compared to men with the median GGN length of 23, those with shorter GGN repeats had higher levels of inhibin B, higher proportions of normal and progressive sperm, and a higher fraction of Fas-positive sperm, while men with longer GGN had higher oestradiol levels.

These data indicate that at least for some markers of male reproductive function the association with CAG or GGN repeat length is curvilinear.

Key words: reproduction, genetic-association studies, trinucleotide repeats, accessory sex organs, cross-sectional studies

Introduction

Male reproductive function is critically dependent on androgen action. Genetic polymorphisms in the gene encoding the androgen receptor (AR) may affect its activity and hence androgen action. The receptor consists of a transactivation domain at the amino-terminal end (encoded by exon 1), a DNA-binding domain (exons 2-3), and a hormone-binding domain (exons 4-8) at the carboxy-terminal end. Exon 1 harbours two polymorphic trinucleotide repeat motifs (CAG and GGN), that encode variable lengths of polyglutamine and polyglycine stretches, respectively (Chang et al., 1988; Lubahn et al., 1988).

The number of AR CAG repeats varies between 10 and 30 in the general population (Lundberg Giwercman et al., 1998; Lundin et al., 2003) whereas subjects with spinal and bulbar muscular atrophy (SBMA), a neuromuscular disorder also characterized by low sperm counts, have an abnormally expanded CAG tract of 40 to 60 repeats (La Spada et al., 1991). Ever since the discovery of an association between an expanded CAG repeat and SBMA, in particular the CAG repeat has been vastly studied in association to prostate cancer risk (Correa-Cerro et al., 1999; Stanford et al., 1997) and male infertility. However, the results of previous studies regarding a possible association between CAG number and sperm counts and/or fertility status have been conflicting, with some studies showing an association between long CAG stretches and impaired spermatogenesis (Casella et al., 2003; Mengual et al., 2003; Patrizio et al., 2001; Tut et al., 1997; von Eckardstein et al., 2001; Yong et al., 1998) whereas others have not (Akinloye et al., 2009; Dadze et al., 2000; Hiort et al., 1999; Lundberg Giwercman et al., 1998; Rajpert-De Meyts et al., 2002; Westerveld et al., 2008).

It has also been shown that men with longer CAG repeats have increased circulating levels of both testosterone (T) and oestradiol (E_2), which illustrates the difficulties in searching for

links between CAG number and phenotypical androgen dependent features, since variations in receptor activity may be blurred by parallel changes in the levels of both T and E2 (Huhtaniemi et al., 2009). However, recent experimental data (Buchanan et al., 2004; Nenonen et al., 2010) as well as analyses of AR genotype in relation to fertility status (Nenonen et al., 2011) and risk of cryptorchidism (Davis-Dao et al., 2012) and hypospadias (Parada-Bustamante et al., 2012) have challenged this view and instead suggested a non-linear relationship between CAG repeat length and receptor function.

The GGN polymorphism in AR occurs in two dominant alleles of 23 and 24 repeats. Only few studies have addressed the functional aspects of AR GGN repeat length. For example, short GGN repeats have been associated with lower semen volume (Lundin et al., 2006), whereas increased GGN repeat length has been observed among boys with penile hypospadias and in men with cryptorchidism (Aschim et al., 2004; Radpour et al., 2007). Both we (Lundin et al., 2003) and others (Rajender et al., 2006; Tut et al., 1997) reported no significant differences in distributions of GGN repeat lengths between infertile and fertile men, while in other studies an association between GGN repeat length and infertility was observed only in combination with certain CAG repeat lengths (Ferlin et al., 2004; Ruhayel et al., 2004).

Thus, despite of the prevailing “dogma” of a negative linear association between CAG repeat length and AR function, there are new data indicating that at least some of the androgenic actions might be linked to the repeat lengths in a non-linear manner. Therefore, we have in this study aimed at analysing the association between AR CAG/GGN repeat lengths and a broader range of parameters of male reproductive function, including sex hormone levels and markers of prostatic, epididymal, and accessory gland function, in fertile men using a univariate regression model allowing for detection of non-linear associations.

Materials and methods

Study population

The study population encompasses proven-fertile men from Greenland, Warsaw, Poland, and Kharkiv, Ukraine that were originally included within a EU-supported study on environment and reproductive function (INUENDO – www.inuendo.dk) using a uniform protocol for data collection (Toft et al., 2005). In all populations, the men were asked to deliver a blood sample for hormone analysis and extraction of leukocyte DNA and an ejaculate for determination of semen parameters and epididymal, prostatic, and accessory gland markers until approximately 200 men had agreed to participate. All included men were 18 years or older at the time of enrolment. A detailed description of study design and data collection procedures has been reported previously (Toft et al., 2005). The participation rates ranged from 29% in Warsaw, 33% in Kharkiv to 79% in Greenland. In total, DNA samples were obtained from 587 men. From 5 samples no genotype could be determined, and from 25 men neither hormone levels nor sperm parameters were obtained, leaving 557 men to be included in the study.

The study was approved by the local ethical committees.

The background characteristics of the study population are presented in Tables 1 and 2.

Genotyping

CAG and GGN repeat lengths were determined by direct sequencing as previously reported (Lundin et al., 2003). DNA was isolated from leukocytes and amplified by polymerase chain reaction (PCR) using the following primers: CAG sense, 5’-TTA GGG CTG GGA AGG GTC TA-3’;

CAG antisense, 5'-TGG GGC CTC TAC GAT GGG CT-3'; GGN sense, 5'-CGG TTC TGG GTC ACC CTC A-3'; GGN antisense, 5'-TCA CCA TGC CGC CAG GGT A-3 (Invitrogen, Stockholm, Sweden) in an Eppendorf Mastercycler (Eppendorf, Hamburg, Germany) with an initial denaturation step at 96˚C followed by 35 cycles of 1 min at 96˚C, 45 s at 61˚C, and 2 min at 72˚C, and a final extension step for 5 min at 72˚C. PCR products were purified and sequenced using an AB3130xl Genetic Analyser (Applied Biosystems, Stockholm, Sweden).

Hormone assays

Details of the hormone measurements have been described earlier (Giwercman et al., 2006). Measurements of follicle stimulating hormone (FSH), luteinising hormone (LH), and E_2 were performed using the UniCel DxI 800 Beckman Access Immunoassay system (Chaska, MN, USA). Total serum T levels were measured using a competitive immunoassay (Access; Beckman Coulter Inc., Fullerton, CA, USA). Sex hormone-binding globulin (SHBG) concentrations were measured using a fluoroimmunoassay (Immulite 2000; Diagnostic Products Corporation, Los Angeles, CA, USA). Inhibin B levels were measured using an immunometric method (Groome et al., 1996). All samples were analysed after the completion of collection at Skåne University Hospital, Malmö and the impact of inter assay variation was minimised by analysing samples from the same centre in the same batch.

Semen analyses

Semen samples were collected by masturbation at the residence or at the hospital. The subjects were asked to abstain from sexual activities for at least two days before collecting the samples, and to note the actual abstinence time. The samples were analysed for sperm concentration, motility, and morphology based on the World Health Organisation manual

for basic semen analysis (World Health Organization, 1999) as described previously (Giwercman et al., 2007). Briefly, sperm concentration was determined using an Improved Neubauer haemocytometer (Glaswarenfabrik Karl Hecht KG, Sondheim, Germany). Sperm motility was determined by counting the proportion of a) fast progressive spermatozoa, b) progressive spermatozoa, c) nonprogressive spermatozoa, and d) immotile spermatozoa. Sperm morphology from all centres was determined centrally at the Fertility Centre, Skåne University Hospital, Malmö on Papanicolaou-stained smears using the WHO 1999 criteria (World Health Organization, 1999).

Sperm DNA fragmentation index (DFI) and high DNA stainability (HDS) were assessed by the sperm chromatin structure assay (SCSA) at the Section of Toxicology and Biomedical Sciences at ENEA Casaccia Research Centre, Italy, as described earlier (Spano et al., 2005) using a FACScan (Becton Dickinson, San Jose, CA, USA). The percentage of abnormal spermatozoa with detectable DNA fragmentation was calculated from the DFI frequency histogram obtained from the ratio between red and total (red + green) fluorescence intensity. HDS was calculated based on the percentage of sperm with high levels of green fluorescence (Evenson et al., 2002).

Strand breaks in sperm DNA were determined using the terminal deoxynucleotidyl transferase-driven dUTP nick end labelling (TUNEL) assay and analysed using flow-cytometry (Epics XL; Beckman Coulter, Fullerton, CA, USA) as described by Stronati et al. (2006).

Determination of percentage of sperm positive for the apoptotic markers Bcl-xL and Fas has been described in detail previously (Stronati et al., 2006).

Sperm Y:X chromosome ratios were determined by two-colour fluorescent *in situ* hybridisation using an Olympus AX 70 epifluorescence microscope equipped with single and double band pass filters as described in detail by Tiido et al. (2006).

Accessory gland markers

Biochemical markers of epididymal function (neutral α-glucosidase, NAG), prostatic function (prostate-specific antigen (PSA) and zinc) and seminal vesicle function (fructose) were assessed as described before (Elzanaty et al., 2002). NAG activity was measured using a commercial assay (Episcreen; Fertipro, Beernem, Belgium) according to the manufacturer's protocol. Seminal PSA concentration was determined using the Prostatus kit (Wallac Oy, Turku, Finland). The concentration of zinc in seminal plasma was determined using a colorimetric assay (Makino et al., 1982). The concentration of fructose in seminal plasma was determined with a spectrophotographical method (Wetterauer and Heite, 1976).

Statistical analyses

In order to normalise the distribution of the residuals, E_2, LH, FSH, inhibin B, PSA, NAG, zinc, semen volume, sperm concentration, total sperm counts, DFI, TUNEL, HDS, abstinence time, and age were transformed by the natural logarithm.

In linear regression models, CAG and GGN lengths were analysed as continuous variables and also trichotomised as short, intermediate, and long (*i.e.,* CAG ≤ 21, 22 ≤ CAG ≤ 24, CAG ≥ 25; GGN ≤ 22, GGN = 23, GGN ≥ 24). The intermediate ranges were chosen as to encompass the median lengths in the three populations.

In order to validate statistically significant associations, we randomly partitioned the data set in two complementary similar-sized pools, while maintaining the relative number of

subjects from each country. We then repeated the linear regression analysis with AR repeat length as categorical variable, and graphically compared the pattern of variation in these two sub-cohorts and the total group of subjects.

Significant associations were further evaluated by modelling with flexible spline terms in generalised additive models (GAMs), adjusted for covariates (gmcv package in R; Wood, 2003). This model avoids the *a priori* choice of a particular response function and allows the data to 'speak for themselves'. All reproductive parameters as well as the covariates (abstinence time, age) were analysed as continuous variables, except time of blood sampling and smoking, which were dichotomised. Heterogeneity between the three centres was examined by interaction terms in the regression models. If the interaction between repeat length category and centre was not statistically significant, the associations were studied for the whole cohort.

In the univariate linear regression models, association of AR repeat lengths with hormone levels was adjusted for potential confounders as centre (Greenland/Warsaw/Kharkiv), age (years), and time of blood sampling (<12:00h: yes/no). Repeat length associations with sperm parameters, and prostatic and epididymal markers were adjusted for potential covariates centre, age, and abstinence time (days). Smoking (yes/no) was added to the model as potential confounder only when the regression coefficients were changed by more than 10%. The log-transformed values are presented in the tables showing linear regression results, whereas back-transformed data are provided in the text. Data expressed in the GAM plots are mean-centred.

P-values < 0.05 were considered as statistically significant. All statistical analyses were performed using the computing environment R v2.15.0 (R Development Core Team, 2012).

Results

Background characteristics of the study population included in the study are shown in Table 1. The mean ± SD CAG and GGN repeat lengths in the combined cohort were 22.4 ± 3.1 and 23.1 ± 1.5, respectively (Table 3). The mean CAG length in the Warsaw population was shorter as compared to the length in Greenland (22.0 ± 3.4 vs. 22.9 ± 2.6; p=0.005), whereas mean GGN length in Warsaw was longer than in Greenland (23.3 ± 1.7 vs. 23.0 ± 0.7; p=0.029).

Associations with CAG repeat length

Linear regression model

For none of the parameters a statistically significant heterogeneity between the centres was seen. With AR CAG length analysed as a continuous variable in a linear regression model, except for a borderline significant negative association with lnDFI [regression coefficient (β), -0.019; 95% confidence interval (CI), -0.037 to -3.6 x 10^{-5}; p = 0.050], no statistically significant association with the reproductive markers was found (data not shown).

Categorical comparison

When analysed as categorical variable with the intermediate lengths as reference, E_2 levels were significantly higher in men with long CAG (mean difference, 8.4 pmol/L; 95% CI, 2.3 to 14 pmol/L; p = 0.007) while they did not differ in men with short CAG (Table 4). No differences in relation to CAG length were observed for the other hormone levels.

DFI and the percentage of Fas-positive cells were both significantly higher in men with short CAG (mean difference, 1.5 percentage points; 95% CI, 0.39 to 2.7 percentage points; p = 0.009, and mean difference, 6.7 percentage points; 95% CI, 0.051 to 13 percentage points; p = 0.048, respectively) as compared to men with median length CAG, while they did not differ in men with long CAG.

Among the epididymal and accessory sex gland markers, both NAG activity and PSA levels were significantly higher in men with either short (mean difference, 0.77 mU/ejaculate; 95% CI, 0.19 to 1.3 mU/ejaculate; p = 0.009, and mean difference, 140 µg/ejaculate; 95% CI 20 to 270 µg/ejaculate; p = 0.023, respectively) or long CAG repeat length (mean difference, 1.4 mU/ejaculate; 95% CI, 0.68 to 2.0 mU/ejaculate; p < 0.001, and mean difference, 136 µg/ejaculate; 95% CI, 1.9 to 270 mU/ejaculate; p = 0.047) as compared to men with intermediate CAG lengths.

The differences between the outcomes in the short or long CAG categories as compared to the reference lengths were still observed when the analysis was repeated in the randomly split data set (Supplemental Figure S1). Although not all the differences reached the level of significance in both pools, they did show similar trends.

Flexible spline regression

The patterns of association were confirmed using flexible spline terms (Fig. 1a). A significant negative linear association was found between CAG repeat length and DFI (R^2 = 0.140; p < 0.05). The u-shaped association of increased NAG activity in men with either short or long CAG repeat length was also found to be significant in the GAM model (R^2 = 0.141; p = 0.010). Although not statistically significant, a clear non-linear relationship between E2 and CAG

with elevated E2 levels only in the men with long CAG repeat lengths and between increased Fas positivity and short CAG could be observed. In the spline analysis, however, a u-shaped association between CAG and PSA levels was not as obvious.

Centre-specific analyses

Besides significantly increased NAG activity in men from Warsaw with long CAG lengths, significantly increased levels of both NAG activity and PSA levels in men with short or long CAG lengths compared to men with intermediate lengths were only observed in the men from Kharkiv. Increased E2 levels in men with long CAG as well as increased DFI in men with short CAG was also only observed in men from Kharkiv. However, except of non-significantly lower PSA levels in Warsaw men with short CAG, all other associations in the individual populations, although not statistically significant, did show similar trends as those found in the total population.

Associations with GGN repeat length

Linear regression model

When analysed as a continuous variable, AR GGN repeat length was only statistically-significantly negatively associated with inhibin B levels (β, -5.5 ng/L; 95% CI, -9.8 to -1.2 ng/L; p = 0.012).

Categorical comparison

When analysed categorically with GGN=23 as the reference, men with short GGN repeat lengths had significantly higher inhibin B levels (mean difference, 29 ng/L; 95% CI, 4.0 to 54

ng/L; p = 0.023), but no difference was seen in the men with long GGN lengths (Table 5). E2 levels were higher in men with long GGN length (mean difference, 6.1 pmol/L; 95% CI, 0.83 to 11 pmol/L; p = 0.023) compared to men with an intermediate length, while no difference was observed in men with short GGN lengths. Men with short GGN had a higher percentage of normal sperms (mean difference, 1.3 percentage points; 95% CI, 0.023 to 2.6 percentage points; p = 0.046) as well as a higher percentage of progressive sperms (mean difference, 7.0 percentage points; 95% CI, 0.24 to 14 percentage points; p = 0.042) as compared to men with a median GGN length. Men with short GGN also had a higher percentage of Fas-positive cells (mean difference, 10 percentage points; 95% CI, 0.27 to 21 percentage points; p = 0.044). Epididymal and accessory sex gland marker levels did not differ with respect to GGN length.

The differences between the outcomes in the short or long GGN categories as compared to the reference lengths were still observed when the analysis was repeated in the randomly split data set (Supplemental Figure S2). Due to the very low number of cases in the short and long GGN repeat categories these differences did not always reach a statistically significant difference.

We further analysed whether there existed an interaction between CAG and GGN repeat length on the effect on the reproductive outcomes, especially since E2 and NAG associated with both CAG and GGN. However, no significant interactions were observed (data not shown).

Flexible spline regression

There was a significant association between GGN repeat length and inhibin B levels when analysed in the spline model (R^2 = 0.085, p = 0.015), confirming the non-linear relationship of elevated inhibin B levels in men with shorter than median length GGN only. The non-linear patterns of association between GGN length and E2, percentage of normal sperm, percentage of progressively motile sperm, and percentage of Fas positive cells were also confirmed, although they did not reach statistical significance in the GAM models.

Centre-specific analyses

Increased levels of inhibin B in men with short GGN length was statistically significant in men from Kharkiv (mean difference; 40 ng/L; 95% CI, 8.7 to 72 ng/L; *p* = 0.013) but not in the other cohorts. Except for non-significantly lower inhibin levels in men with short GGN length in Greenland, the different cohorts showed similar trends as the total population.

Discussion

In this study we have shown heterogeneous, curvilinear, relationships between the polymorphic androgen receptor trinucleotide repeat lengths and a wide range of male reproductive markers.

AR is expressed in Leydig cells, peritubular myoid cells and Sertoli cells of the testis (Suarez-Quian et al., 1999), as well as in the epithelial and stromal cells of the efferent duct, prostate, seminal vesicles, and epididymis (Sar et al., 1990). Genetic polymorphisms in the receptor that affect its transactivation potential are therefore likely to manifest themselves in androgen-dependent markers of reproductive health. It is commonly assumed that the CAG repeat length influences the transactivation capacity of the receptor in an inverse manner, whereas this is less clear for the GGN repeat length. This assumption is mainly supported by reduced receptor activity in relationship with pathological expansions of the CAG repeat length that are seen in Kennedy syndrome (La Spada et al., 1991). Besides spinobulbar muscular atrophy, this syndrome is characterised by hypoandrogenism due to partial androgen insensitivity. However, recent studies have indicated that an intermediate number of CAG repeats confines optimal receptor function *in vitro* (Buchanan et al., 2004; Nenonen et al., 2010) and that both men with short (<20) and long (>23) CAG repeat lengths are at increased risk for infertility (Nenonen et al., 2011). Thus, currently there are at least two models to predict the effect of the repeats on AR activity: the inverse relationship, and the optimal range model.

The main finding of our study is that only two out of twenty one reproductive markers were associated with AR CAG or GGN repeat length in a linear manner, whereas curvilinear associations were found for ten of the reproductive markers. While this does not fully

support our view that the associations between AR repeat lengths and markers for androgen action follow the optimal range model, it does demonstrate that these associations should be studied in a non-linear fashion. The curvilinear relationships between AR repeat lengths and the reproductive outcomes reported in this study suggest a segmented association, where a linear association flattens either below or above a certain threshold length and the effects either disappear or become saturated. Our data further illustrate that androgen action may be differentially affected by AR repeat lengths in different tissues: sperm DNA parameters differed in subjects with short AR repeat lengths; E2 levels in those with longer repeat lengths; accessory sex gland markers exhibited a u-shaped association with both short ánd long CAG repeat lengths. It has indeed been shown that the influence of CAG repeats on the transactivation potential of the receptor is cell type specific, presumably depending on the specific profile of transcriptional cofactors (Buchanan et al., 2004). Interpretation of these associations is further complicated by the intrinsic complexity of the interaction between AR and its ligands. For example, increased T levels may compensate for the weaker transactivating activity of receptors with short CAG/GGN repeats (Crabbe et al., 2007; Huhtaniemi et al., 2009). In a recent study, Davis-Dao et al. (2012) showed an unexpected increased risk of cryptorchidism in cases with short CAG repeats. The authors speculate on an indirect effect of the CAG repeats via alternative nongenomic signalling pathways, where reduced T levels in men with short CAG lengths lead to a suboptimal nongenomic signalling. Interestingly, nongenomic androgen signalling has been reported in studies demonstrating that androgens induce cell proliferation through nongenomic activation of MAPK and PI3K/Akt in prostate cancer cell lines (Kampa et al., 2008; Rahman and Christian, 2007).

Of the two linearly-associated markers, the linear relationship between DFI and CAG repeat length was supported by a significant association using a nonbiased smooth spline regression model, but the association between GGN length and inhibin was found to be curvilinear in the GAM model.

For other reproductive markers a clear curvilinear relationship that flattened at either the shorter or longer repeat lengths was observed. Serum E2 levels were elevated in men with long CAG repeats, while no difference was observed in men with short repeats, compared to men with intermediate lengths. In a recent study on aging European men, significant positive associations of E2 and T with CAG repeat length were reported by Huhtaniemi et al. (2009). In our study with relatively young men (geometric mean, 29.0 ± 5.9 yrs), a positive, but not significant, linear association between CAG and T levels was also observed, as has been reported in several studies (Crabbe et al., 2007; Mifsud et al., 2001a). The spline regression analysis confirmed that increased E2 levels indeed only occur in men with CAG repeats longer than median length.

Increased E2 levels were also observed in men with longer than median GGN repeat length. Only few studies have addressed associations of GGN repeat length with reproductive markers. Ruhayel et al. (2004) reported low prostate marker concentrations in infertile men with 24 GGN repeats as compared to men with the median length of 23 repeats. Lundin et al. (2006) showed that, compared to men with ≥23 GGN repeats, men with short GGN length (< 23) had lower semen volume and semen fructose levels. Currently, the relationship between GGN repeat numbers and receptor function is unclear as *in vitro* studies have been conflicting. Deletion of all GGN repeats leads to a 30% reduction in AR receptor transactivation potential (Gao et al., 1996), indicating the importance of the GGN stretch in

receptor function. One study reported no difference in transactivation between AR receptors with short and long repeats (Ding et al., 2005), whereas others reported an inverse (Werner et al., 2006) or a positive relationship (Brockschmidt et al., 2007). Most recently, a non-linear receptor transactivation potential was shown by Lundin et al. (2007) with highest activity in receptors harbouring 23 repeats. Interestingly, a similar non-linear relationship with risk of prostate cancer has been reported with highest risk in men with 23 repeats compared to men with either shorter or longer lengths (Platz et al., 1998).

Semen characteristics were also found to be associated with AR repeat lengths in a non-linear manner. Men with short CAG repeats had significantly higher DFI as well as higher percentage of Fas-positive sperm, whereas no difference compared to the intermediate length was seen in men with long CAG repeats. A curvilinear association between CAG and Fas positivity was also apparent in the GAM model. Inverse relationships between CAG repeat length and fertility have been observed in several studies (Lundberg Giwercman et al., 2004; Mifsud et al., 2001b; Tut et al., 1997; von Eckardstein et al., 2001). TUNEL-derived DNA strand breaks did not significantly associate with CAG repeat length. Although SCSA-derived DFI is supposed to assess the same biological endpoint as TUNEL-derived sperm DNA breaks, they could mirror different aspects of sperm DNA damage, *i.e.* the former more indirectly evaluating sperm DNA denaturability, and the latter direct accessibility of DNA breaks to the TdT enzyme.

From the epididymal and accessory sex gland markers, the levels of both NAG and PSA showed a clear u-shaped relationship, with increased levels in men with either short or long CAG repeat lengths. This is unexpected when taking into account the reduced transactivation potential of receptors with short and long repeats as reported by Nenonen

et al. (2010). However, although not statistically significant, testosterone levels were higher in men with long CAG repeats, which may compensate for reduced receptor activity in these men (Crabbe et al., 2007; Huhtaniemi et al., 2009). Analysis with flexible splines confirmed the u-shaped association of NAG, whereas increased levels of PSA were only apparent in the men with short CAG. The major strength of this study is that besides analysing the associations between the reproductive markers and AR repeat lengths with the *a priori* assumption of a linear relationship, we compared short and long repeats with the intermediate lengths. Additionally, we performed a smooth curve regression analysis using flexible spline terms. Second, this study analyses a wide variety of androgen-dependent markers of reproductive health in a large population of proven-fertile, healthy men. A disadvantage of the latter is, however, that the effect sizes in this population can be expected to be small. A weaker point of our study is that the population is composed of men from two different ethnic groups. However, analysis of heterogeneity did not reveal an interaction between centre and AR repeat lengths and centre was included as a potential confounder in the regression analyses. It can be argued that whereas the associations with CAG were highly-statistically significant, the associations with GGN repeats were rather weak and on the border of significance. This may be explained by the small number of men with GGN repeat lengths in the outer ranges. However, the observation that most associations were also apparent in smooth spline regression models further supports our observations. Since our study aims to be explorative rather than confirmative, we have not corrected for multiple testing in the statistical analysis. However, we have validated our statistically significant findings by repeating the analysis in two randomly split data sets that overall confirmed the observed associations. Due to the low number of cases in each data

set this did not always reached a level of significance. Further validation to confirm our findings should be performed in an independent cohort.

In summary, we have shown that a range of markers of reproductive function in fertile healthy men are related to AR CAG and GGN repeat lengths in a curvilinear manner. These associations would not have been observed if they had been analysed with an *a priori* expectation of a linear relationship.

Acknowledgements

The INUENDO study was funded by The European Commission's 5th Framework Programme (contract no. QLK4-CT-2001-00202). The Ukrainian part of the study was supported by a grant from INTAS (contract no. 2001 2205). Grants were also received from Aarhus University. The Swedish part of the study has also been funded by the European Commission's 7th Framework Programme (CLEAR study, http://www.inuendo.dk/clear; grant FP7-ENV-2008-1-226217), the Swedish Medical Research Council, the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning. The authors thank Virissa Lenters and Roel Vermeulen for expert assistance with the statistical analyses.

References

Akinloye, O., Gromoll, J., Nieschlag, E., Simoni, M., 2009. Androgen receptor gene CAG and GGN polymorphisms in infertile Nigerian men. J. Endocrinol. Invest. 32, 797-804.

Aschim, E.L., Nordenskjold, A., Giwercman, A., Lundin, K.B., Ruhayel, Y., Haugen, T.B., Grotmol, T., Lundberg Giwercman, Y., 2004. Linkage between cryptorchidism, hypospadias, and GGN repeat length in the androgen receptor gene. J. Clin. Endocrinol. Metab. 89, 5105-5109.

Brockschmidt, F.F., Nothen, M.M., Hillmer, A.M., 2007. The two most common alleles of the coding GGN repeat in the androgen receptor gene cause differences in protein function. J. Mol. Endocrinol. 39, 1-8.

Buchanan, G., Yang, M., Cheong, A., Harris, J.M., Irvine, R.A., Lambert, P.F., Moore, N.L., Raynor, M., Neufing, P.J., Coetzee, G.A., Tilley, W.D., 2004. Structural and functional consequences of glutamine tract variation in the androgen receptor. Hum. Mol. Genet. 13, 1677-1692.

Casella, R., Maduro, M.R., Misfud, A., Lipshultz, L.I., Yong, E.L., Lamb, D.J., 2003. Androgen receptor gene polyglutamine length is associated with testicular histology in infertile patients. J. Urol. 169, 224-227.

Chang, C.S., Kokontis, J., Liao, S.T., 1988. Structural analysis of complementary DNA and amino acid sequences of human and rat androgen receptors. Proc. Natl. Acad. Sci. U. S. A. 85, 7211-7215.

Correa-Cerro, L., Wohr, G., Haussler, J., Berthon, P., Drelon, E., Mangin, P., Fournier, G., Cussenot, O., Kraus, P., Just, W., Paiss, T., Cantu, J.M., Vogel, W., 1999. (CAG)nCAA and GGN repeats in the human androgen receptor gene are not associated with prostate cancer in a French-German population. Eur. J. Hum. Genet. 7, 357-362.

Crabbe, P., Bogaert, V., De Bacquer, D., Goemaere, S., Zmierczak, H., Kaufman, J.M., 2007. Part of the interindividual variation in serum testosterone levels in healthy men reflects differences in androgen sensitivity and feedback set point: contribution of the androgen receptor polyglutamine tract polymorphism. J. Clin. Endocrinol. Metab. 92, 3604-3610.

Dadze, S., Wieland, C., Jakubiczka, S., Funke, K., Schroder, E., Royer-Pokora, B., Willers, R., Wieacker, P.F., 2000. The size of the CAG repeat in exon 1 of the androgen receptor gene shows no significant relationship to impaired spermatogenesis in an infertile Caucasoid sample of German origin. Mol. Hum. Reprod. 6, 207-214.

Davis-Dao, C., Koh, C.J., Hardy, B.E., Chang, A., Kim, S.S., De Filippo, R., Hwang, A., Pike, M.C., Carroll, J.D., Coetzee, G.A., Vandenberg, D., Siegmund, K., Cortessis, V.K., 2012. Shorter androgen receptor CAG repeat lengths associated with cryptorchidism risk among Hispanic white boys. J. Clin. Endocrinol. Metab. 97, E393-399.

Ding, D., Xu, L., Menon, M., Reddy, G.P., Barrack, E.R., 2005. Effect of GGC (glycine) repeat length polymorphism in the human androgen receptor on androgen action. Prostate. 62, 133-139.

Elzanaty, S., Richthoff, J., Malm, J., Giwercman, A., 2002. The impact of epididymal and accessory sex gland function on sperm motility. Hum. Reprod. 17, 2904-2911.

Evenson, D.P., Larson, K.L., Jost, L.K., 2002. Sperm chromatin structure assay: its clinical use for detecting sperm DNA fragmentation in male infertility and comparisons with other techniques. J. Androl. 23, 25-43.

Ferlin, A., Bartoloni, L., Rizzo, G., Roverato, A., Garolla, A., Foresta, C., 2004. Androgen receptor gene CAG and GGC repeat lengths in idiopathic male infertility. Mol. Hum. Reprod. 10, 417-421.

Gao, T., Marcelli, M., McPhaul, M.J., 1996. Transcriptional activation and transient expression of the human androgen receptor. J. Steroid Biochem. Mol. Biol. 59, 9-20.

Giwercman, A., Rignell-Hydbom, A., Toft, G., Rylander, L., Hagmar, L., Lindh, C., Pedersen, H.S., Ludwicki, J.K., Lesovoy, V., Shvets, M., Spano, M., Manicardi, G.C., Bizzaro, D., Bonefeld-Jorgensen, E.C., Bonde, J.P., 2006. Reproductive hormone levels in men exposed to persistent organohalogen pollutants: a study of inuit and three European cohorts. Environ. Health Perspect. 114, 1348-1353.

Giwercman, A., Rylander, L., Rignell-Hydbom, A., Jonsson, B.A., Pedersen, H.S., Ludwicki, J.K., Lesovoy, V., Zvyezday, V., Spano, M., Manicardi, G.C., Bizzaro, D., Bonefeld-Jorgensen, E.C., Toft, G., Bonde, J.P., Giwercman, C., Tiido, T., Lundberg Giwercman, Y., 2007. Androgen receptor gene CAG repeat length as a modifier of the association between persistent organohalogen pollutant exposure markers and semen characteristics. Pharmacogenet. Genomics. 17, 391-401.

Groome, N.P., Illingworth, P.J., O'Brien, M., Pai, R., Rodger, F.E., Mather, J.P., McNeilly, A.S., 1996. Measurement of dimeric inhibin B throughout the human menstrual cycle. J. Clin. Endocrinol. Metab. 81, 1401-1405.

Hiort, O., Horter, T., Schulze, W., Kremke, B., Sinnecker, G.H., 1999. Male infertility and increased risk of diseases in future generations. Lancet. 354, 1907-1908.

Huhtaniemi, I.T., Pye, S.R., Limer, K.L., Thomson, W., O'Neill, T.W., Platt, H., Payne, D., John, S.L., Jiang, M., Boonen, S., Borghs, H., Vanderschueren, D., Adams, J.E., Ward, K.A., Bartfai, G., Casanueva, F., Finn, J.D., Forti, G., Giwercman, A., Han, T.S., Kula, K., Lean, M.E., Pendleton, N., Punab, M., Silman, A.J., Wu, F.C., European Male Ageing Study Group, 2009. Increased estrogen rather than decreased androgen action is associated with longer androgen receptor CAG repeats. J. Clin. Endocrinol. Metab. 94, 277-284.

Kampa, M., Pelekanou, V., Castanas, E., 2008. Membrane-initiated steroid action in breast and prostate cancer. Steroids. 73, 953-960.

La Spada, A.R., Wilson, E.M., Lubahn, D.B., Harding, A.E., Fischbeck, K.H., 1991. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature. 352, 77-79.

Lubahn, D.B., Joseph, D.R., Sullivan, P.M., Willard, H.F., French, F.S., Wilson, E.M., 1988. Cloning of human androgen receptor complementary DNA and localization to the X chromosome. Science. 240, 327-330.

Lundberg Giwercman, Y., Richthoff, J., Lilja, H., Anderberg, C., Abrahamsson, P.A., Giwercman, A., 2004. Androgen receptor CAG repeat length correlates with semen PSA levels in adolescence. Prostate. 59, 227-233.

Lundberg Giwercman, Y., Xu, C., Arver, S., Pousette, A., Reneland, R., 1998. No association between the androgen receptor gene CAG repeat and impaired sperm production in Swedish men. Clin. Genet. 54, 435-436.

Lundin, K.B., Giwercman, A., Dizeyi, N., Lundberg-Giwercman, Y., 2007. Functional in vitro characterisation of the androgen receptor GGN polymorphism. Mol. Cell. Endocrinol. 264, 184-187.

Lundin, K.B., Giwercman, A., Richthoff, J., Abrahamsson, P.A., Lundberg-Giwercman, Y., 2003. No association between mutations in the human androgen receptor GGN repeat and inter-sex conditions. Mol. Hum. Reprod. 9, 375-379.

Lundin, K.B., Giwercman, Y.L., Rylander, L., Hagmar, L., Giwercman, A., 2006. Androgen receptor gene GGN repeat length and reproductive characteristics in young Swedish men. Eur. J. Endocrinol. 155, 347-354.

Makino, T., Saito, M., Horiguchi, D., Kina, K., 1982. A highly sensitive colorimetric determination of serum zinc using water-soluble pyridylazo dye. Clin. Chim. Acta. 120, 127-135.

Mengual, L., Oriola, J., Ascaso, C., Ballesca, J.L., Oliva, R., 2003. An increased CAG repeat length in the androgen receptor gene in azoospermic ICSI candidates. J. Androl. 24, 279-284.

Mifsud, A., Choon, A.T., Fang, D., Yong, E.L., 2001a. Prostate-specific antigen, testosterone, sex-hormone binding globulin and androgen receptor CAG repeat polymorphisms in subfertile and normal men. Mol. Hum. Reprod. 7, 1007-1013.

Mifsud, A., Sim, C.K., Boettger-Tong, H., Moreira, S., Lamb, D.J., Lipshultz, L.I., Yong, E.L., 2001b. Trinucleotide (CAG) repeat polymorphisms in the androgen receptor gene: molecular markers of risk for male infertility. Fertil. Steril. 75, 275-281.

Nenonen, H., Bjork, C., Skjaerpe, P.A., Giwercman, A., Rylander, L., Svartberg, J., Lundberg Giwercman, Y., 2010. CAG repeat number is not inversely associated with androgen receptor activity in vitro. Mol. Hum. Reprod. 16, 153-157.

Nenonen, H.A., Giwercman, A., Hallengren, E., Lundberg Giwercman, Y., 2011. Non-linear association between androgen receptor CAG repeat length and risk of male subfertility--a meta-analysis. Int. J. Androl. 34, 327-332.

Parada-Bustamante, A., Lardone, M.C., Madariaga, M., Johnson, M.C., Codner, E., Cassorla, F., Castro, A., 2012. Androgen receptor CAG and GGN polymorphisms in boys with isolated hypospadias. J. Pediatr. Endocrinol. Metab. 25, 157-162.

Patrizio, P., Leonard, D.G., Chen, K.L., Hernandez-Ayup, S., Trounson, A.O., 2001. Larger trinucleotide repeat size in the androgen receptor gene of infertile men with extremely severe oligozoospermia. J. Androl. 22, 444-448.

Platz, E.A., Giovannucci, E., Dahl, D.M., Krithivas, K., Hennekens, C.H., Brown, M., Stampfer, M.J., Kantoff, P.W., 1998. The androgen receptor gene GGN microsatellite and prostate cancer risk. Cancer Epidemiol. Biomarkers Prev. 7, 379-384.

R Development Core Team, 2012. R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria.

Radpour, R., Rezaee, M., Tavasoly, A., Solati, S., Saleki, A., 2007. Association of long polyglycine tracts (GGN repeats) in exon 1 of the androgen receptor gene with cryptorchidism and penile hypospadias in Iranian patients. J. Androl. 28, 164-169.

Rahman, F., Christian, H.C., 2007. Non-classical actions of testosterone: an update. Trends Endocrinol. Metab. 18, 371-378.

Rajender, S., Rajani, V., Gupta, N.J., Chakravarty, B., Singh, L., Thangaraj, K., 2006. No association of androgen receptor GGN repeat length polymorphism with infertility in Indian men. J. Androl. 27, 785-789.

Rajpert-De Meyts, E., Leffers, H., Petersen, J.H., Andersen, A.G., Carlsen, E., Jorgensen, N., Skakkebaek, N.E., 2002. CAG repeat length in androgen-receptor gene and reproductive variables in fertile and infertile men. Lancet. 359, 44-46.

Ruhayel, Y., Lundin, K., Lundberg-Giwercman, Y., Hallden, C., Willen, M., Giwercman, A., 2004. Androgen receptor gene GGN and CAG polymorphisms among severely oligozoospermic and azoospermic Swedish men. Hum. Reprod. 19, 2076-2083.

Sar, M., Lubahn, D.B., French, F.S., Wilson, E.M., 1990. Immunohistochemical localization of the androgen receptor in rat and human tissues. Endocrinology. 127, 3180-3186.

Spano, M., Toft, G., Hagmar, L., Eleuteri, P., Rescia, M., Rignell-Hydbom, A., Tyrkiel, E., Zvyezday, V., Bonde, J.P., INUENDO, 2005. Exposure to PCB and p, p'-DDE in European and Inuit populations: impact on human sperm chromatin integrity. Hum. Reprod. 20, 3488-3499.

Stanford, J.L., Just, J.J., Gibbs, M., Wicklund, K.G., Neal, C.L., Blumenstein, B.A., Ostrander, E.A., 1997. Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk. Cancer Res. 57, 1194-1198.

Stronati, A., Manicardi, G.C., Cecati, M., Bordicchia, M., Ferrante, L., Spano, M., Toft, G., Bonde, J.P., Jonsson, B.A., Rignell-Hydbom, A., Rylander, L., Giwercman, A., Pedersen, H.S., Bonefeld-Jorgensen, E.C., Ludwicki, J.K., Lesovoy, V., Sakkas, D., Bizzaro, D., 2006. Relationships between sperm DNA fragmentation, sperm apoptotic markers and serum levels of CB-153 and p,p'-DDE in European and Inuit populations. Reproduction. 132, 949-958.

Suarez-Quian, C.A., Martinez-Garcia, F., Nistal, M., Regadera, J., 1999. Androgen receptor distribution in adult human testis. J. Clin. Endocrinol. Metab. 84, 350-358.

Tiido, T., Rignell-Hydbom, A., Jonsson, B.A., Lundberg Giwercman, Y., Pedersen, H.S., Wojtyniak, B., Ludwicki, J.K., Lesovoy, V., Zvyezday, V., Spano, M., Manicardi, G.C., Bizzaro, D., Bonefeld-Jorgensen, E.C., Toft, G., Bonde, J.P., Rylander, L., Hagmar, L., Giwercman, A., Inuendo, 2006. Impact of PCB and p,p'-DDE contaminants on human sperm Y:X chromosome ratio: studies in three European populations and the Inuit population in Greenland. Environ. Health Perspect. 114, 718-724.

Toft, G., Axmon, A., Giwercman, A., Thulstrup, A.M., Rignell-Hydbom, A., Pedersen, H.S., Ludwicki, J.K., Zvyezday, V., Zinchuk, A., Spano, M., Manicardi, G.C., Bonefeld-Jorgensen, E.C., Hagmar, L., Bonde, J.P., Inuendo, 2005. Fertility in four regions spanning large contrasts in serum levels of widespread persistent organochlorines: a cross-sectional study. Environ. Health. 4, 26.

Tut, T.G., Ghadessy, F.J., Trifiro, M.A., Pinsky, L., Yong, E.L., 1997. Long polyglutamine tracts in the androgen receptor are associated with reduced trans-activation, impaired sperm production, and male infertility. J. Clin. Endocrinol. Metab. 82, 3777-3782.

von Eckardstein, S., Syska, A., Gromoll, J., Kamischke, A., Simoni, M., Nieschlag, E., 2001. Inverse correlation between sperm concentration and number of androgen receptor CAG repeats in normal men. J. Clin. Endocrinol. Metab. 86, 2585-2590.

Werner, R., Schutt, J., Hannema, S., Ropke, A., Wieacker, P., Hiort, O., Holterhus, P.M., 2006. Androgen receptor gene mutations in androgen insensitivity syndrome cause distinct patterns of reduced activation of androgen-responsive promoter constructs. J Steroid Biochem Mol Biol. 101, 1-10.

Westerveld, H., Visser, L., Tanck, M., van der Veen, F., Repping, S., 2008. CAG repeat length variation in the androgen receptor gene is not associated with spermatogenic failure. Fertil. Steril. 89, 253-259.

Wetterauer, U., Heite, H., 1976. Eine empfehlenswerte Methode zur gleichzeitigen enzymatischen Bestimmung von Citrat und Fruktose im Seminalplasma. Akt Dermatol. 2, 239-248.

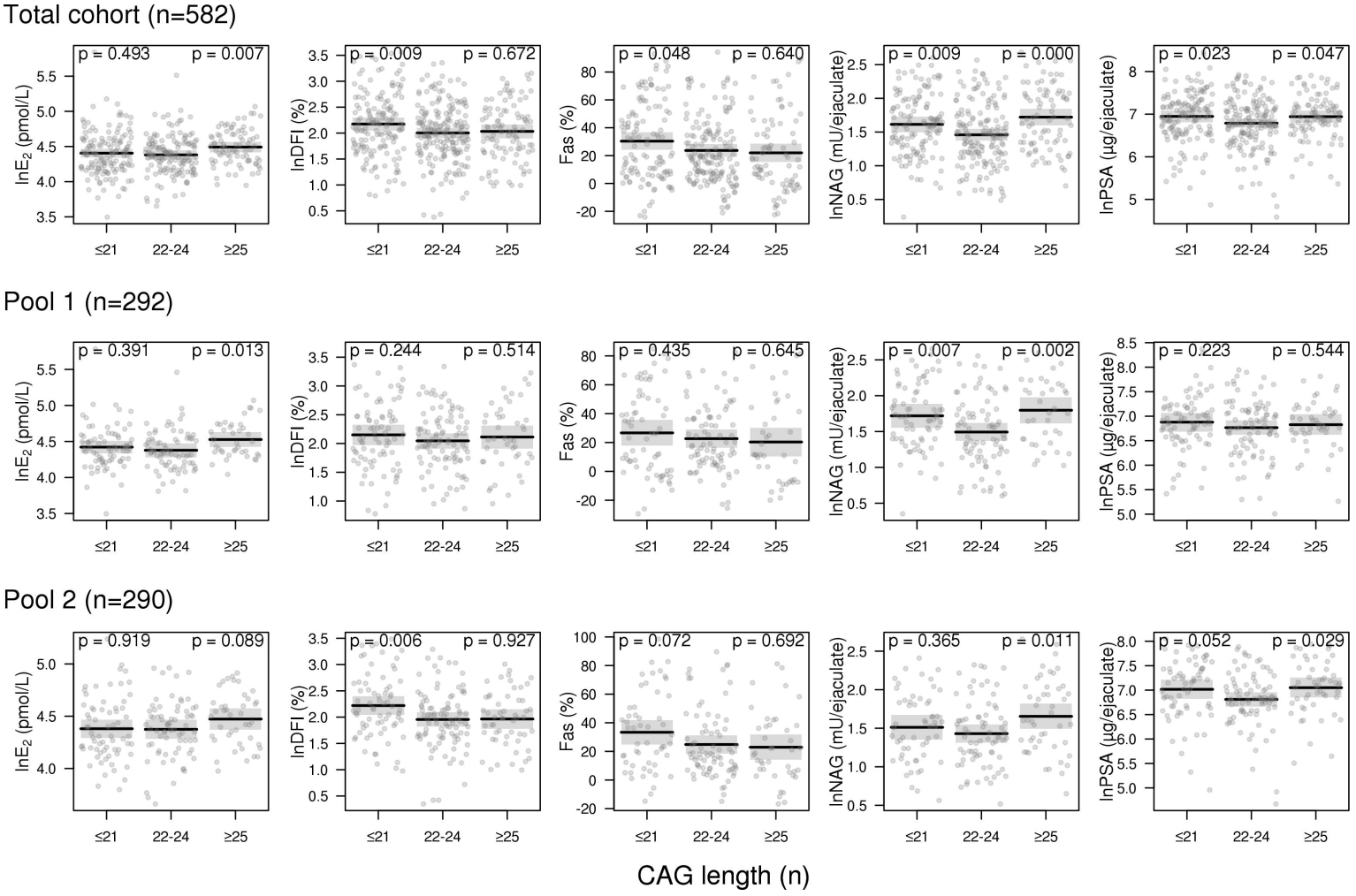
Wood, S.N., 2003. Thin plate regression splines. J R Stat Soc. 65, 95-114.
World Health Organization, 1999. WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction, 3 ed. Cambridge University Press, Cambridge.
Yong, E.L., Ghadessy, F., Wang, Q., Mifsud, A., Ng, S.C., 1998. Androgen receptor transactivation domain and control of spermatogenesis. Rev. Reprod. 3, 141-144.

Figure legends

Figure 1. GAMs of associations between CAG length (upper) and GGN (lower) and markers of reproductive function that were statistically significant using the linear regression models. R^2 coefficients of the penalised spline regression models with P values and 95% CIs (dashed lines) are indicated.

Supplementary Figure S1. Adjusted means (solid line) with 95% confidence interval (gray shading) of reproductive markers stratified according to CAG length category in the total cohort and in two randomly-split, similar-sized, pools. Significance of the difference as compared to the reference category 22 ≤ CAG ≤ 24 is indicated above the respective short and long CAG categories.

Supplementary Figure S2. Adjusted means (solid line) with 95% confidence interval (gray shading) of reproductive markers stratified according to CAG length category in the total cohort and in two randomly-split, similar-sized, pools. Significance of the difference as compared to the reference category GGN = 23 is indicated above the respective short and long CAG categories.


Total cohort (n=582)
p = 0.493
p = 0.007
lnE2 (pmol/L)
p = 0.009
p = 0.672
lnDFI (%)
p = 0.048
p = 0.640
Fas (%)
p = 0.009
p = 0.000
lnNAG (mU/ejaculate)
p = 0.023
p = 0.047
lnPSA (µg/ejaculate)
Pool 1 (n=292)
p = 0.391
p = 0.013
p = 0.244
p = 0.514
p = 0.435
p = 0.645
p = 0.007
p = 0.002
p = 0.223
p = 0.544
Pool 2 (n=290)
p = 0.919
p = 0.089
p = 0.006
p = 0.927
p = 0.072
p = 0.692
p = 0.365
p = 0.011
p = 0.052
p = 0.029
≤21
22-24
≥25
CAG length (n)

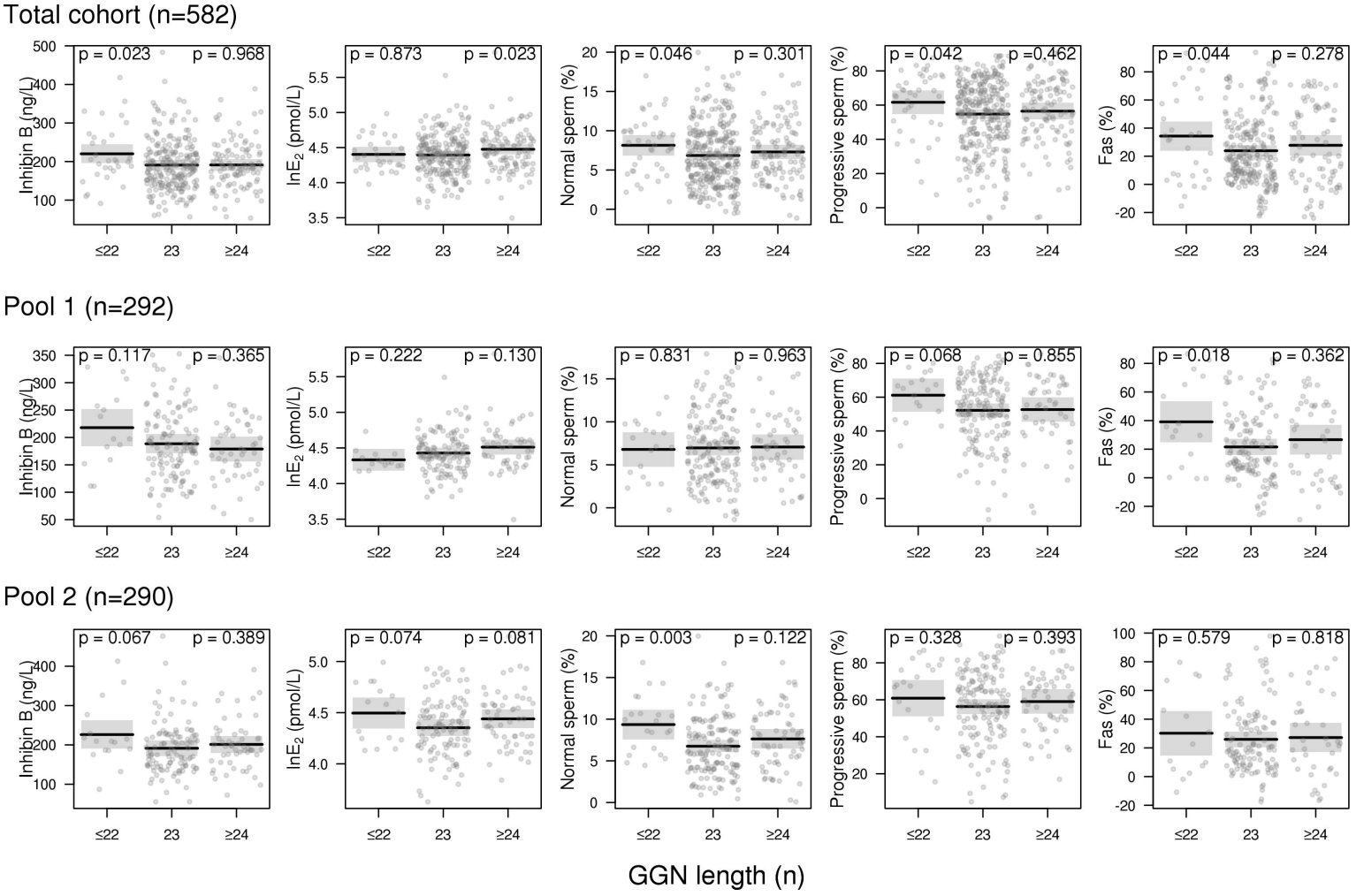

Total cohort (n=582)
Inhibin B (ng/L)
p = 0.023
p = 0.968
InE2 (pmol/L)
p = 0.873
p = 0.023
Normal sperm (%)
p = 0.046
p = 0.301
Progressive sperm (%)
p = 0.042
p =0.462
Fas (%)
p = 0.044
p = 0.278
Pool 1 (n=292)
Inhibin B (ng/L)
p = 0.117
p = 0.365
InE2 (pmol/L)
p = 0.222
p = 0.130
Normal sperm (%)
p = 0.831
p = 0.963
Progressive sperm (%)
p = 0.068
p = 0.855
Fas (%)
p = 0.018
p = 0.362
Pool 2 (n=290)
Inhibin B (ng/L)
p = 0.067
p = 0.389
InE2 (pmol/L)
p = 0.074
p = 0.081
Normal sperm (%)
p = 0.003
p = 0.122
Progressive sperm (%)
p = 0.328
p = 0.393
Fas (%)
p = 0.579
p = 0.818
≤22
23
≥24
GGN length (n)

Table 1. Background characteristics and reproductive hormone levels of the study populations.

	All centres (n=557)			*Greenland (n=213)*			*Warsaw (n=187)*			*Kharkiv (n=157)*		
	n	*GM*[a]	*SD*	*n*	*GM*	*SD*	*n*	*GM*	*SD*	*n*	*GM*	*SD*
Age (y)	551	29.0	5.92	211	30.2	6.89	186	30.1	3.91	154	26.2	5.59
Current smoking[b] (n, %)	557	298	(54)	213	147	(69)	187	53	(28)	157	98	(62)
Abstinence (d)	463	3.30	5.81	186	2.75	3.12	120	4.19	9.73	157	3.37	1.94
Blood sample[b] <12:00 (n, %)	423	375	(89)	105	87	(83)	161	151	(94)	157	137	(87)
Reproductive hormone levels												
Follicle-stimulating hormone (IU/L)	418	3.89	2.65	149	4.47	2.87	113	3.46	2.13	156	3.71	2.67
Luteinizing hormone (IU/L)	419	3.91	1.83	150	4.06	1.90	113	3.79	1.63	156	3.85	1.89
Inhibin B (ng/L)	419	167	67.0	150	172	70.8	113	143	56.0	156	181	66.0
Sex hormone-binding globulin (mmol/L)	418	25.0	9.98	149	27.2	9.45	113	21.8	11.7	156	25.6	8.64
Testosterone (nmol/L)	419	14.7	5.23	150	14.1	4.89	113	12.5	4.26	156	17.1	5.29
Oestradiol (pmol/L)	417	71.9	26.3	149	62.8	16.6	113	71.5	32.2	155	82.1	25.3

[a] GM, geometric mean

[b] data reported as n (%)

Table 2. Semen characteristics and epididymal and accessory sex gland markers of the study populations.

	All centres (n=557)			*Greenland (n=213)*			*Warsaw (n=187)*			*Kharkiv (n=157)*		
	n	*GM*[a]	*SD*	*n*	*GM*	*SD*	*n*	*GM*	*SD*	*n*	*GM*	*SD*
Semen characteristics												
Volume (mL)	477	3.29	1.85	176	3.20	1.75	171	3.40	1.81	130	3.26	2.03
Concentration (million/mL)	529	52.7	67.3	197	51.4	61.1	175	56.2	82.1	157	50.6	53.4
Count (million)	475	174	294	176	160	239	169	192	379	130	169	214
Normal sperm (%)	527	6.01	3.83	196	5.96	3.62	176	5.58	3.82	155	6.63	4.04
Progressive sperm (%) (WHO A+B)	500	53.0	19.2	194	50.4	18.5	156	60.5	16.7	150	49.6	21.1
DNA fractionation index (%DFI)	482	8.98	7.42	196	7.63	5.95	130	9.87	7.72	156	10.2	8.35
High DNA stainability (%HDS)	482	10.4	7.66	196	11.3	8.48	130	8.79	6.08	156	10.8	7.46
DNA fractionation index [TUNEL (%)]	437	5.11	10.1	196	2.91	5.25	121	11.2	11.7	120	6.10	10.9
Fas positivity (%)	401	21.7	29.0	191	17.4	19.3	112	35.3	34.9	98	19.3	29.0
Bcl-xL positivity (%)	247	11.8	28.8	143	11.8	31.0	92	10.4	19.5	12	27.9	39.9
Y chromosome sperm cells (%)	363	50.8	1.77	158	51.2	2.08	109	50.3	1.26	96	50.8	1.55
Epididymal and accessory sex gland markers												
Neutral a-glucosidase (mU/ejaculate)	402	5.21	3.03	155	4.63	2.66	132	6.09	3.24	115	5.12	2.98
Prostate-specific antigen (µg/ejaculate)	454	916	606	165	930	620	163	1058	598	126	744	557
Zinc (mmol/ejaculate)	444	1.48	1.18	165	1.36	1.06	163	1.78	1.29	116	1.29	1.09
Fructose (mmol/ejaculate)	453	12.7	6.86	165	13.7	6.86	163	12.1	7.03	125	12.2	6.63

[a] GM, geometric mean

Table 3. Genotype distributions in the three populations

		All populations *n=557*		*Greenland* *n=213*		*Warsaw* *n=187*		*Kharkiv* *n=157*	
CAG	median (range)	23	(6 - 37)	24	(13 - 29)	22	(6 - 37)	22	(14 - 31)
	mean (SD)	22.4	(3.1)	22.9	(2.6)	22.0	(3.4)	22.4	(3.2)
	n ≤ 21 (%)	201	(36)	48	(22)	84	(46)	69	(44)
	22 ≤ n ≤ 24 (%)	238	(43)	127	(60)	63	(31)	48	(31)
	n ≥ 25 (%)	118	(21)	38	(18)	40	(23)	40	(25)
GGN	median (range)	23	(10 - 30)	23	(20 - 27)	23	(16 - 30)	23	(10 - 27)
	mean (SD)	23.1	(1.5)	23.0	(0.7)	23.3	(1.7)	23.0	(1.9)
	n ≤ 22 (%)	44	(8)	8	(4)	16	(9)	20	(13)
	n = 23 (%)	360	(65)	181	(85)	91	(48)	88	(56)
	n ≥ 24 (%)	153	(27)	24	(11)	80	(43)	49	(31)

Table 4. Adjusted regression coefficients of reproductive markers with median lengths 22 ≤ CAG ≤ 24 as a reference.

	CAG ≤ 23			CAG ≥ 25		
	β	95% CI	p	β	95% CI	p
Reproductive hormone levels[a]						
lnFSH (IU/L)[c]	0.044	(-0.084; 0.17)	0.500	-0.026	(-0.18; 0.12)	0.727
lnLH (IU/L)	-0.042	(-0.15; 0.061)	0.421	-0.035	(-0.15; 0.085)	0.568
Inh B (ng/L)	-7.9	(-23; 7.9)	0.326	1.8	(-16; 20)	0.843
lnSHBG (mmol/L)	-0.053	(-0.14; 0.033)	0.228	-0.099	(-0.12; 0.00045)	0.051
T (nmol/L)	-0.088	(-1.3; 1.1)	0.883	0.71	(-0.65; 2.1)	0.306
lnE2 (pmol/L)	0.024	(-0.046; 0.094)	0.493	0.11	(0.030; 0.19)	**0.007**
Semen characteristics[b]						
lnVolume (mL)[c]	-0.010	(-0.12; 0.10)	0.856	-0.015	(-0.14; 0.11)	0.814
lnConcentration (million/mL)	0.15	(-0.045; 0.35)	0.130	0.16	(-0.064; 0.38)	0.164
lnCount (million)	0.14	(-0.084; 0.37)	0.218	0.11	(-0.14; 0.36)	0.395
Normal sperm (%)	0.51	(-0.32; 1.3)	0.225	0.68	(-0.23; 1.6)	0.143
Progressive sperm (%)	2.1	(-2.3; 6.5)	0.353	2.9	(-1.9; 7.7)	0.235
lnDFI (%)	0.17	(0.043; 0.30)	**0.009**	0.031	(-0.11; 0.17)	0.672
lnHSD (%)	-0.034	(-0.16; 0.087)	0.583	-0.055	(-0.19; 0.079)	0.418
lnTUNEL (%)	0.090	(-0.11; 0.29)	0.367	0.10	(-0.11; 0.32)	0.352
Fas positivity (%)	6.7	(0.051; 13)	**0.048**	-1.7	(-8.9; 5.5)	0.640
Bcl-xL positivity (%)	3.4	(-5.9; 13)	0.471	8.2	(-2.1; 18)	0.119
Y-bearing sperm (%)[c]	0.18	(-0.30; 0.65)	0.469	-0.16	(-0.67; 0.35)	0.530
Epididymal and accessory sex gland markers[b]						
lnNAG (mU/ejaculate)	0.16	(0.038; 0.27)	**0.009**	0.26	(0.13; 0.39)	**0.000**
lnPSA (µg/ejaculate)	0.16	(0.022; 0.30)	**0.023**	0.15	(0.0021; 0.30)	**0.047**
lnZinc (mmol/ejaculate)	0.13	(-0.029; 0.30)	0.107	0.13	(-0.047; 0.31)	0.148
Fructose (mmol/ejaculate)	-0.14	(-1.7; 1.4)	0.865	-1.6	(-3.3; 0.16)	0.075

[a] hormone levels: adjusted for centre, age, and blood sample taken < 12:00; [b] semen characteristics and epididymal and accessory gland markers: adjusted for centre, age, and abstinence time; [c] adjusted for smoking.

Table 5. Adjusted regression coefficients of reproductive markers with median length GGN=23 as a reference.

	GGN ≤ 22			GGN ≥ 24		
	β	95% CI	p	β	95% CI	p
Reproductive hormone levels						
lnFSH (IU/L)[c]	-0.13	(-0.33 ; 0.075)	0.215	0.020	(-0.11 ; 0.15)	0.763
lnLH (IU/L)	-0.094	(-0.26 ; 0.071)	0.262	0.010	(-0.094 ; 0.11)	0.850
Inh B (ng/L)	29	(4.0 ; 54)	**0.023**	0.32	(-15 ; 16)	0.968
lnSHBG (mmol/L)	0.072	(-0.067 ; 0.21)	0.310	0.027	(-0.060 ; 0.12)	0.542
T (nmol/L)	0.97	(-0.90 ; 2.8)	0.309	0.54	(-0.65 ; 1.7)	0.373
lnE2 (pmol/L)	0.0091	(-0.10 ; 0.12)	0.873	0.082	(0.011 ; 0.15)	**0.023**
Semen characteristics[b]						
lnVolume (mL)[c]	0.046	(-0.12 ; 0.22)	0.593	0.029	(-0.087 ; 0.14)	0.625
lnConcentration (million/mL)	0.20	(-0.10 ; 0.51)	0.191	0.039	(-0.17 ; 0.25)	0.709
lnCount (million)	0.20	(-0.14 ; 0.54)	0.254	0.074	(-0.16 ; 0.31)	0.529
Normal sperm (%)	1.3	(0.023 ; 2.6)	**0.046**	0.45	(-0.40 ; 1.3)	0.301
Progressive sperm (%)	6.9	(0.25 ; 13)	**0.042**	1.700	(-2.8 ; 6.2)	0.462
lnDFI (%)	-0.059	(-0.26 ; 0.14)	0.562	-0.13	(-0.27 ; 0.0035)	0.056
lnHSD (%)	-0.13	(-0.31 ; 0.062)	0.189	-0.061	(-0.19 ; 0.065)	0.343
lnTUNEL (%)	0.0029	(-0.30 ; 0.30)	0.985	-0.067	(-0.28 ; 0.15)	0.536
Fas positivity (%)	10	(0.274 ; 20)	**0.044**	3.9	(-3.2 ; 11)	0.278
Bcl-xL positivity (%)	-7.1	(-23 ; 8.8)	0.380	-6.8	(-17 ; 3.5)	0.194
Y chromosome sperm (%)[c]	0.091	(-0.64 ; 0.82)	0.806	-0.067	(-0.57 ; 0.43)	0.793
Epididymal and accessory sex gland markers[b]						
lnNAG (mU/ejaculate)[c]	0.0043	(-0.17 ; 0.181)	0.962	0.031	(-0.091 ; 0.15)	0.616
lnPSA (µg/ejaculate)	-0.024	(-0.23 ; 0.18)	0.818	-0.079	(-0.22 ; 0.062)	0.270
lnZinc (mmol/ejaculate)	-0.13	(-0.37 ; 0.12)	0.302	-0.044	(-0.21 ; 0.12)	0.600
Fructose (mmol/ejaculate)	1.6	(-0.81 ; 3.9)	0.197	-0.45	(-2.1 ; 1.1)	0.578

[a] hormone levels: adjusted for centre, age, and blood sample taken < 12:00; [b]semen characteristics, and epididymal and accessory sex gland markers: adjusted for centre, age, and abstinence time; [c] adjusted for smoking.