Androgen receptor gene GGN repeat length and reproductive characteristics in young Swedish men

K B Lundin, Y L Giwercman, L Rylander, L Hagmar and A Giwercman
Molecular Reproductive Medicine Research Unit, Department of Clinical Sciences, Lund University, Fertility Centre, Malmö University Hospital, SE-205 02 Malmö, Sweden and Department of Occupational and Environmental Medicine, Lund University Hospital, Lund University, SE-221 85 Lund, Sweden

Abstract

Background: The human androgen receptor (AR) gene contains two polymorphisms of CAG and GGN repeats respectively. The GGN repeat function is still largely unknown and to date there are no in vivo data on this segment with respect to the general population.

Methods: We investigated the impact of CAG and GGN repeats on male reproductive function, one by one and in interaction with each other, in 220 adolescent men from the general Swedish population. Physical examination and semen analysis, including accessory sex gland markers and measurement of reproductive hormone levels, were performed. Lifestyle-associated factors, including maternal smoking during pregnancy, were recorded. GGN and CAG repeat lengths were determined by sequencing.

Results: GGN<23 was associated with lower semen volume when compared to GGN=23 (mean difference -0.6 ml, P=0.02) and GGN>23 (mean difference -0.9 ml, P=0.002). Men with GGN<23, exposed to maternal smoking during pregnancy, had higher body mass index compared to men with other GGN lengths, no matter whether their mother had smoked or not during pregnancy (mean difference 4.8 kg/m², P<0.001).

Conclusions: Short GGN repeats seem to be associated with decreased semen volume, possibly due to suboptimal AR activity. Body composition may be influenced by the combination of fetal exposure to maternal smoking and certain AR genotypes.

Introduction

The human androgen receptor (AR), located on chromosome Xq11-12 (1, 2), is a transcription factor regulating the development of male reproductive organs in the fetus and secondary sex characteristics at puberty in response to testosterone (T) and 5α-dihydrotestosterone (DHT) (3). A functioning AR is also crucial for normal spermatogenesis (4). The AR contains two polymorphic regions, the (CAG)nCAA repeat encoding polyglutamine, and the (GGT)3GGG(GGT)2(GGC)n repeat encoding polyglycine, commonly referred to as the CAG and GGN repeats respectively (5). The length of the CAG repeat has been linked to spinal and bulbar muscular atrophy, and has been suggested to play a role in the pathogenesis of prostate cancer, testicular cancer and male infertility (6–10). The CAG stretch has also been suggested to modulate body fat mass and serum concentrations of leptin and insulin in men (11). In vitro experiments have indicated an inverse relationship between the CAG repeat length and AR transcriptional activity (9). In experimental studies, a complete deletion of the GGN repeat resulted in a 30% reduction of the transcriptional activity of the AR, indicating that the GGN repeat is also of functional importance (12). A recently published in vitro study, expressing AR with different lengths of the GGN repeat (19–23 repeats), showed that there was no difference in the transcriptional activity between ARs with short GGN repeats and an AR with the most common length, 23 GGN (13). However, only one concentration of R1881 (10 nM) was tested, and no comparisons were made with the second most common GGN length of 24 repeats, or with extremely short or long GGN repeats. Furthermore, all constructs contained a CAG repeat of 24, which is above the average length.

So far, no data regarding the impact of GGN repeat length in the general male population have been published. Most of the available studies focused on the risk of prostate cancer. Mutations in the first six base triplets of the GGN repeat have been suggested as a possible cause of ambiguous genitalia in 46 XY individuals (14). We have, however, recently shown that these mutations occur quite frequently in the general population, and do not seem to give rise to profound changes in the male phenotype, although
they may be associated with subtle effects on sperm and semen parameters (15).

Two previous studies have indicated that a GGN length of 24 combined with CAG≥21 might be more common in men suffering from infertility (16, 17). Furthermore, it was reported that older men having one of the less common GGN variants (GGN≠23) associated with a short CAG repeat (CAG≤23) presented with higher waist:hip ratios, when compared to men with other GGN/CAG allele combinations (18). It is therefore likely that the GGN segment is involved in metabolic and endocrine regulation and/or male reproductive organ function, possibly in a synergistic fashion with the CAG repeat or other factors. Secondary factors, such as maternal smoking during pregnancy, could also have a negative impact on sperm number and body composition of the offspring (19, 20). However, it is unknown whether certain genotypes might increase the susceptibility of an individual, both during fetal development and as an adult, to the adverse effects of cigarette smoke.

In order to test the hypothesis that the GGN length might have an impact on male reproductive health, data on reproductive markers obtained from 220 Swedish adolescent men were analyzed in relation to GGN repeat number. Body mass index (BMI) was also included as an outcome variable. The association between the GGN length and these parameters was tested alone and in combination with the CAG repeat number. Furthermore, information regarding the subjects’ lifestyle and maternal smoking during pregnancy was included in the analyses, with the purpose of finding possible interactions with the AR genotype in relation to the above-mentioned outcome parameters.

Subjects and methods

Subjects

Three-hundred and five men under medical examination for military service were included in a study of reproductive function in young Swedish men, previously described by Richthoff et al. (21). Their mean (±S.D.) age at the time of the study was 18.1 (±0.4), ranging from 18 to 21 years. The AR gene is located on the X-chromosome. Therefore, in order to exclude a possible impact of ethnic variation, genotyping of the AR CAG and GGN repeats was performed exclusively in men with Swedish mothers (n = 223). Three of the men were excluded due to mutations in the GGN repeat region (22), resulting in 220 subjects studied.

All participants underwent a physical examination. Weight and height were measured and BMI calculated for 218 of the participating men. The participants also delivered a semen sample and were asked to complete a questionnaire pertaining to their birth weight, current smoking and alcohol consumption. Regarding smoking and alcohol consumption, information was given by 219 men. One hundred and fifty-nine of the participants were nonsmokers, 35 were light smokers (1–9 cigarettes/day) and 25 were heavy smokers (>10 cigarettes/day). One hundred and forty-five of the participants reported no regular alcohol intake, whereas 66 consumed 1–49 g alcohol/week, and 8 consumed more than 50 g alcohol/week. Among the 201 men who responded to the question, 42 had mothers who had smoked during pregnancy. All men participated after giving written informed consent according to protocols approved by the ethical review board of Lund University.

Analysis of semen

Of the 220 participants, 219 delivered a semen sample, which was obtained by masturbation. A minimum of 48 h sexual abstinence was suggested. However, in each case, the actual length of the abstinence period was registered. The assessment of volume, sperm concentration and motility was performed as recommended by the World Health Organization’s (WHO) 1999 recommendations (23). For the assessment of the ejaculate volume, the weight of the empty plastic coupe was subtracted from the total weight of coupe and semen. Concentration assessment was done using a modified Neubauer chamber and positive displacement pipettes for proper dilution of the ejaculate. According to the WHO recommendations for the determination of motility, 200 sperms were scored in categories a, b, c, and d, with a corresponding to rapid progressive motility, b to slowly progressive motility, c to nonprogressive motility, and d to immotile spermatozoa. Furthermore, the motility of the spermatozoa was assessed using computer-aided sperm motility analysis (CASA) as described previously (24). The laboratory participates in an external Quality Control Programme on concentration and motility. The program is organized by Nordic Association of Andrology and European Society of Human Reproduction and Embryology.

Since the epididymal as well as accessory sex gland functions are strongly androgen-dependent, measurements of the seminal levels of markers of their function were performed. The activity of neutral-α-glucosidase (NAG) and the concentrations of prostate-specific antigen (PSA), fructose and zinc in the seminal plasma of the participants were analyzed, as previously reported (25). The NAG activity was measured using a commercially available kit (Episcreen: Fertipro, Gent, Belgium) (coefficient of variation (CV) = 15%), while the total PSA concentration was determined with a fluoroimmunometric method (26), the PROSTATUS kit from Wallay Oy’, Finland (CV = 5%). The concentration
of fructose was measured spectrophotographically as described by Wetterau and Heite (27) and run on a Beckman Synchron LX20 instrument. The CV was 5% for control samples. Zinc was analyzed by a colorimetric method (CV = 7%) (28). Due to insufficient amounts of semen, we had to abstain from some assays in a small proportion of ejaculates (Table 1). However, sperm characteristics and hormone concentrations were not significantly different between analyzed and nonanalyzed samples. Semen data for the whole cohort of the military conscripts have been published elsewhere (21, 29).

Hormone analyses

Circulating levels of follicle-stimulating hormone (FSH), luteinizing hormone (LH), sex hormone-binding globulin (SHBG), T, and estradiol (E2) were measured in all 220 participants on an automated fluorescence detection system (Autodelfia, Wallac Oy; Turku, Finland) at the routine clinical chemistry laboratory, Uppsala University Hospital. Intra- and total-assay CV values were below 4 and 7.5% respectively. Free T was calculated as recommended by Vermeulen (30), based on the known concentrations of T, albumin and SHBG. Inhibin B levels were assessed using an immunometric assay (Oxford Bioinnovation; Oxfordshire, UK), as previously described (31), with total-assay variation coefficients below 7%. Prior to analysis, denaturation was performed by heating.

Genetic analyses

Genomic DNA was extracted from peripheral leukocytes and the GGN repeat was amplified in a 50 µl PCR containing approximately 10 ng DNA, 0.3 µmol/l of each of the primers: F-TTAGGGGCTGGGAAGGGTCTA, and R-TGGGGCCTCTACGATGGGCT, 1.5 mmol/l MgCl2, 200 µmol/l of dATP, dCTP and dTTP each, 100 µmol/l dGTP and 7-deaza-dGTP (Roche) respectively. Amplification was performed for 35 cycles in an Eppendorf Mastercycler (Eppendorf; Hamburg, Germany). Each cycle included denaturation at 96 °C for 45 s, primer annealing at 61 °C for 45 s and primer extension at 72 °C for 2 min, with an initial denaturation step at 96 °C for 3 min and a final extension step at 72 °C for 5 min.

PCR-products were purified, directly sequenced with the reverse primers from the PCRs, precipitated, resuspended, and run externally on an eight-capillary...
Beckman Coulter CEQ 2000XL (Beckman Coulter; Bromma, Sweden) sequencing gear.

**Statistical analysis**

The GGN repeat lengths were subdivided into three groups: <23, 23, and >23, based on the knowledge that the GGN=23 allele is present in approximately 52% and the GGN=24 allele in approximately 32% of the Swedish population (15). Alleles with GGN<23 and >24 only comprise approximately 14 and 2% of the population respectively (Fig. 1).

In linear regression models, we evaluated the associations between CAG (as continuous variable) and GGN (as trichotomized variable) repeat lengths and seminal characteristics, hormone parameters, biochemical seminal markers, and BMI. By including interaction terms in the models, we also tested if the CAG length modified the effects of the GGN repeat. Lifestyle-related factors such as current smoking, alcohol consumption and maternal smoking during pregnancy were considered as potential confounders. Other variables were also included as potential confounders (in brackets), in the analyses of reproductive parameters and BMI, such as semen volume and sperm concentration (sexual abstinence), BMI (T), SHBG (BMI) and T (SHBG). These variables were included in the models, one at a time, and were kept in the models if they changed the effect estimates by 15% or more.

Previous findings indicated a functional difference between GGN=23 and 24 (16); consequently, these two alleles were also separately compared with regard to the above-mentioned outcome variables.

Since a recent report claimed that men with both GGN≠23 and CAG≤23 differ from men with all other allele combinations regarding body composition (18), these groups were also compared in relation to the outcome variables defined above.

**Results**

Men with GGN<23 (n=32) had lower semen volume than subjects with GGN=23 (n=115) (mean difference: −0.6 ml, 95% confidence interval (CI) −1.1, −0.08; P=0.02) (Fig. 2, Table 2). The difference was more pronounced if men with GGN<23 were compared to men with GGN>23 (n=72) (mean difference: −0.9 ml, 95% CI −1.4, −0.3; P=0.002). No significant difference in semen volume was seen when comparing GGN=23 with >23. No potential confounding factor fulfilled the criteria to be included in a multivariate model. There were no associations between GGN repeat length and sperm concentration and sperm motility.

Total seminal amount of fructose did not differ between men with GGN<23 (n=31) and men with GGN=23 (n=110). However, when comparing GGN<23 with >23 (n=70), a statistically significant difference was found (mean difference: −14 μmol/ejaculate, 95% CI −27, −1.0; P=0.04). No significant difference was seen when comparing GGN=23 with >23. No potential confounding factor fulfilled the criteria to be included in a multivariate model. There

![Figure 1](https://example.com/f1.png)  
*Figure 1* Distribution of CAG (A) and GGN (B) repeat lengths in 220 Swedish military conscripts.

![Figure 2](https://example.com/f2.png)  
*Figure 2* Box-whisker plot showing seminal volume in relation to androgen receptor GGN repeat length in 219 Swedish military conscripts.

www.eje-online.org
were no associations between GGN repeat length and other seminal markers PSA, zinc, or NAG.

In a univariate analysis comparing men with GGN<23 (n = 32) to men with GGN=23 (n = 116), a statistically significantly lower SHBG concentration was found in the GGN<23 group (mean difference: −4.1 nmol/l, 95% CI −8.0, −0.3 nmol/l; P = 0.04), but this association lost its significance when adjusted for BMI in a multivariate model (mean difference: −2.9 nmol/l, 95% CI −6.7, 0.9; P = 0.13). When comparing GGN<23 (n = 32) with GGN>23 (n = 72) and GGN=23 with GGN>23, no significant differences were found neither before adjustment for BMI nor after. There were no associations between GGN repeat length and any of the reproductive hormones: FSH, LH, total and free T, E2, and inhibin B.

In a univariate analysis comparing men with GGN<23 (n = 31) to men with GGN=23 (n = 116), a statistically significantly higher BMI was found in the GGN<23 group (mean difference: 1.3 kg/m², 95% CI 0.03, 2.6; P = 0.04). No significant differences were seen when comparing GGN<23 with >23 or GGN=23 with >23. After adjustment for total T concentration in a multivariate model, a higher BMI was still associated with GGN<23 (n = 31) when compared to GGN=23 (n = 116) (mean difference: 1.3 kg/m², 95% CI 0.06, 2.6; P = 0.04), but not when compared to GGN>23 or when comparing GGN=23 with >23.

When including maternal smoking during pregnancy in the models, we found that higher BMI in men with GGN<23 repeats was confined to men with mothers who had smoked during pregnancy. These men (n = 7) had significantly higher BMI when compared to the remaining group of men with other GGN lengths or compared to men with mothers who did not smoke during pregnancy (n = 193) (mean difference: 4.8 kg/m², 95% CI 2.3, 7.2; P < 0.001; Table 3). Due to the low number of subjects in this group, we tested the robustness of the statistical analysis by excluding the subject with the highest BMI from the calculations. The association remained, however, statistically significant.

Birth weight did not differ significantly for the group with high BMI (n = 5) when compared to the rest of the subjects (n = 129) (mean difference: 224 g, 95% CI −273, 721; P = 0.4).

Significant differences were not found in any of the tested outcome variables between men with GGN=23 and 24 nor did we find any statistically significant associations between GGN repeat length and any of the other reproductive hormones: FSH, LH, total and free T, E2, and inhibin B.

Table 2 Univariate comparisons of reproductive parameters between GGN<23, 23 and ≥23 in military conscripts.

<table>
<thead>
<tr>
<th>GGN</th>
<th>&lt;23</th>
<th>23</th>
<th>≥23</th>
<th>&lt;23 vs 23</th>
<th>23 vs ≥23</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of CAG</td>
<td>23 (4.0)</td>
<td>22 (2.8)</td>
<td>22 (3.2)</td>
<td>0.22</td>
<td>0.31</td>
</tr>
<tr>
<td>Semen volume (ml)</td>
<td>2.6 (1.1)</td>
<td>3.2 (1.3)</td>
<td>3.5 (1.2)</td>
<td>0.02</td>
<td>0.14</td>
</tr>
<tr>
<td>Sperm concentration (10⁶/ml)</td>
<td>79 (61)</td>
<td>72 (73)</td>
<td>78 (68)</td>
<td>0.66</td>
<td>0.63</td>
</tr>
<tr>
<td>Motility A+B (%)</td>
<td>48 (19)</td>
<td>47 (18)</td>
<td>43 (14)</td>
<td>0.97</td>
<td>0.12</td>
</tr>
<tr>
<td>CASA motility (%)</td>
<td>45 (26)</td>
<td>49 (21)</td>
<td>52 (22)</td>
<td>0.40</td>
<td>0.26</td>
</tr>
</tbody>
</table>
| Zinc (mg/ejaculate) | 4.1 nmol/l, 95% CI 0.04, 2.6; P = 0.04) | No significant differences were found between GGN<23, 23 and ≥23 in military conscripts.

Table 3 BMI (kg/m²) (mean (s.d.)) for men with GGN<23, 23 and ≥23, in relation to maternal smoking during pregnancy.

<table>
<thead>
<tr>
<th>Maternal smoking during pregnancy</th>
<th>&lt;23</th>
<th>23</th>
<th>≥23</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>22.6 (2.1), n=22</td>
<td>22.6 (3.8), n=81</td>
<td>22.7 (2.7), n=55</td>
</tr>
<tr>
<td>Yes</td>
<td>27.2 (3.5) , n=7</td>
<td>21.5 (1.5), n=24</td>
<td>21.0 (3.1), n=11</td>
</tr>
</tbody>
</table>

P ≤ 0.002 when comparing the GGN<23/+ maternal smoking group with all other groups.
associations between the tested outcome variables and CAG repeat length.

When applying interaction models in the statistical analysis, no modifying effect of CAG length on the association between the GGN number and any of the outcome variables was seen. Moreover, when comparing $GGN \neq 23$ in combination with $CAG \leq 23$ vs the remaining group, no statistically significant differences in any of the outcome variables were found.

**Discussion**

The main finding of this study was that in Swedish adolescent males, short $GGN$ repeat lengths ($< 23$) were associated with significantly lower semen volume when compared to men with $GGN \geq 23$. However, no association between $GGN$ length and sperm number and motility, or reproductive hormone concentrations, was found. Since the seminal fluid predominantly consists of prostate- and seminal vesicle fluid, it is conceivable that $GGN$ length-dependent differences in the AR function are of larger physiological importance in highly androgen-dependent organs, like the prostate and the seminal vesicles, than in less androgen-dependent tissues. In accordance with this assumption, we found a significantly lower excretion of fructose – a marker of seminal vesicle function – in men with short $GGN$ repeats when compared to those with 24 or more $GGN$. Although PSA and zinc followed the same trend, statistical significance was not reached. The inconclusive findings regarding the seminal markers could partly be due to low statistical power related to high interindividual variation for excretion of these markers.

Several reproductive end points were tested as a part of this study and we cannot rule out the possibility of mass significance. The 95% CIs are quite large and the $P$-values sometimes close to borderline significance, which is a common problem with unevenly distributed study populations. However, previous studies regarding AR $GGN$ repeat and reproductive parameters indicate that our findings might reflect a true biological phenomenon. Recently, we have reported that cryptorchidism and penile hypospadias are associated with longer $GGN$ repeats compared to controls (32). In similarity to accessory sex gland function, both the development of male external genitalia and testicular descent are DHT-dependent events. Additionally, a low semen volume has been seen in men with point mutations in the $GGN$ region and with very long $GGN$ stretches (15). Furthermore, in a study on infertile men, $GGN$ length of 24 was associated with significantly lower levels of excretion of prostatic products (16). Thus, both the present and the previous studies on infertile men indicate that variation in $GGN$ repeat length may be associated with accessory sex gland function. Notably, in the present study, the lower seminal volume was seen in men with $GGN < 23$, whereas in the infertility patients, a decreased secretion was found in men with $GGN = 24$ compared to those with $GGN = 23$. It was recently reported that infertile men have subnormal Leydig cell function (33). Whether the discrepancy between infertile men and conscripts as considers the relative functional status of the three $GGN$ repeat length subgroups is due to a possible difference in AR function in response to different androgen concentrations, remains to be tested in vivo. An alternative explanation is that even normal adolescent men with $GGN = 23$ and 24 differ from each other with regards to reproductive parameters, but such difference was not disclosed due to relatively limited statistical power of our study. A significant reduction of the risk of type-II error would require substantial increase in the subject number, which is not feasible.

Additional support for the hypothesis that minor changes in $GGN$ repeat length can have a biological effect is presented in a recent publication on early-onset androgenetic alopecia (AGA) (34). The authors found that a $GGN$ repeat length of 23 was associated with early-onset AGA, whereas a $GGN$ repeat length of 24 seemed to have a protective effect. The mentioned in vivo findings clearly indicate a functional effect of varying $GGN$ repeat lengths, at least on DHT-dependent organs/events. Furthermore, it seems plausible that the $GGN$ repeat should not be considered as a continuous variable with linearity between repeat length and AR activity.

How minor changes in the $GGN$ repeat length could influence the function of the AR, especially in a tissue/organ-specific way, which previous studies as well as the present one indicate, remains unclear. One theory is that changes in $GGN$ length could affect the receptor activity in a ligand-specific way, through the interaction between the ligand-dependent activation function 2 region in the ligand-binding domain and the ligand-independent activation function 5 region, harboring the GGN repeat, in the N-terminal domain (35). The binding of DHT to an AR harboring a short $GGN$ repeat might result in impaired interaction when compared to the binding of T, resulting in larger effects in the DHT-dependent tissues than in the T-dependent tissues. Furthermore, different $GGN$ lengths could give rise to conformational differences in mRNA (36, 37). Thus, recognition sites for cofactors and splicing factors, etc. might be affected in response to different repeat lengths in a tissue- or cell-specific manner.

Recently published data showed an inverse relationship between T levels in men and BMI as well as waist:hip ratio, demonstrating the role of androgens in the regulation of body composition (38, 39). A recent animal study also demonstrated the correlation between male adiposity and the androgen–AR system (40). Inactivation of the androgen–AR system caused late-onset obesity in male mice. We found an association between short $GGN$ repeats and higher BMI.
On the other hand, our study showed that a short GGN repeat was a predictor of increased BMI independent of the CAG repeat length. This discrepancy might be due to the fact that we focused on adolescent men, whereas Gustafson et al. studied older men. Since free T levels are decreasing by age, the relative importance of different GGN and CAG repeat lengths in relation to androgen-dependent modulation of body composition may differ between young and old men.

Further statistical analysis revealed that the higher BMI of men with short GGN repeats was confined to men whose mothers smoked during pregnancy. The finding of a possible interaction between fetal exposure to cigarette smoking and an AR polymorphism in relation to body composition is novel. Although this is an interesting observation, indicating a gene–lifestyle interaction, we cannot exclude the possibility of this being a chance of finding, as the group of men with GGN<23 and smoking mothers consisted of seven individuals only. A rough estimation of the robustness of the statistical analysis was done by excluding the highest BMI value in this group. Although the statistical significance was persisting, caution should be taken when interpreting these findings.

In the present work, no statistically significant association between CAG number and semen parameters was found, probably due to the limited statistical power. In a previous study on a larger number of conscripts including those who did not have Swedish mothers, we found a negative association between CAG repeat length and total sperm count as well as sperm motility (41).

In conclusion, our study showed that the short GGN repeats were associated with a decrease in semen volume, possibly due to a decreased AR activity. Nevertheless, the AR GGN repeat length seemed to have rather subtle effects on reproductive function. The GNN number in combination with fetal exposure to maternal smoking also seemed to have an impact on BMI. We speculate that the GNN=23 allele has a superior function compared to all other lengths and therefore, has evolved as the most common GNN allele among Caucasians. However, reporter gene assays on varying GGN lengths, in response to different hormones, clarifying the importance of the GGN repeat alone and in combination with the CAG repeat, are warranted.

Acknowledgements
We wish to acknowledge the help of Jonas Richthoff for sample collection and Camilla Anderberg for excellent technical assistance. The study was supported by grants from Swedish Governmental Founding for Clinical Research, Swedish research Council (Grant No K2005–72X-14 545-03A), the Gunnar Nilssons Cancer Foundation, and the Foundation for Urological Research and Malmö University Hospital Foundation for Cancer Research.

References
354 K B Lundin and others

European Journal of Endocrinology (2006) 155


29 Richthoff J, Rylander L, Jönsson B, Åkesson H, Hagmar L, Nilsson-Ehle P, Stridsberg M & Giwercman A. Serum levels of 2.2',4.4', 5.5'-hexachlorobiphenyl (CB-153) in relation to markers of reproductive function in young males from the general Swedish population. Environmental Health Perspectives 2003 111 409–413.


36 Yeap BB, Wilce JA & Leedman PJ. The androgen receptor mRNA. BioEssays 2004 26 672–682.


Received 3 May 2006
Accepted 5 May 2006