

Baldness and the androgen receptor: the *AR* polyglycine repeat polymorphism does not confer susceptibility to androgenetic alopecia

Justine A. Ellis · Katrina J. Scurrah ·
Joanna E. Cobb · Sophie G. Zaloumis ·
Anna E. Duncan · Stephen B. Harrap

Received: 18 August 2006 / Accepted: 18 December 2006 / Published online: 26 January 2007
© Springer-Verlag 2007

Abstract Androgenetic alopecia, or male pattern baldness, is a complex condition with a strong heritable component. In 2001, we published the first significant evidence of a genetic association between baldness and a synonymous coding SNP (rs6152) in the androgen receptor gene, *AR*. Recently, this finding was replicated in three independent studies, confirming an important role for *AR* in the baldness phenotype. In one such replication study, it was claimed that the causative variant underlying the association was likely to be the polyglycine (GGN) repeat polymorphism, one of two apparently functional triplet repeat polymorphisms located in the exon 1 transactivating domain of the gene. Here, we extend our original association finding and present comprehensive evidence from approximately 1,200 fathers and sons drawn from 703 families of the Victorian Family Heart Study, a general population Caucasian cohort, that neither exon 1 triplet repeat polymorphism is causative in this condition. Seventy-eight percent of fathers (531/683) and 30% of sons (157/520) were affected to some degree with AGA. We utilised statistical methods appropriate for the categorical nature of the phenotype and familial structure of the cohort, and determined that whilst SNP rs6152 was strongly associated with baldness ($P < 0.0001$), the GGN triplet repeat was not ($P = 0.13$). In the absence of any other known common functional coding variants, we argue that the causative variant is likely to be in the

non-coding region, and yet to be identified. The identification of functional non-coding variants surrounding *AR* may have significance not only for baldness, but also for the many other complex conditions that have thus far been linked to *AR*.

Introduction

The common heritable loss of scalp hair known as androgenetic alopecia (AGA), or male pattern baldness, requires the presence of androgen (Hamilton 1942) and predisposing genes (Ellis et al. 1998; Kuster and Happle 1984; Nyholt et al. 2003) for expression. In 2001, we published the first significant evidence of genetic association with AGA (Ellis et al. 2001), identifying the gene encoding the androgen receptor, *AR*, through marked association with a synonymous single nucleotide polymorphism (SNP) in exon 1 (rs6152, *StuI* RFLP, E211 G > A) in a Caucasian population. Our finding has been faithfully replicated in three independent studies (Hayes et al. 2005; Hillmer et al. 2005; Levy-Nissenbaum et al. 2005), confirming an important role of *AR* in the heritability of AGA.

Functional variant(s) in or around *AR* have been sought and attention has been focused on the transactivating domain, in which are found two triplet-repeat polymorphisms, a polyglutamine (CAG) repeat lying proximal, and a polyglycine (GGN) repeat lying distal to rs6152. The polyglutamine repeat length appears inversely related to the ability of the AR protein to effect transcriptional control on target genes (e.g. Beilin et al. 2000; Chamberlain et al. 1994; Choong et al. 1996; Ding et al. 2004; Kazemi-Esfarjani et al.

J. A. Ellis (✉) · K. J. Scurrah · J. E. Cobb ·
S. G. Zaloumis · A. E. Duncan · S. B. Harrap
Department of Physiology,
The University of Melbourne,
Melbourne, VIC 3010, Australia
e-mail: justine@unimelb.edu.au

1995), whilst the polyglycine repeat length has been inversely associated with AR protein levels (Ding et al. 2005). No other common functional coding variants have been identified in *AR*.

The recent linkage and association study by Hillmer et al. (2005) replicated our original finding, and reported strong associations with a variety of polymorphisms throughout the ~1 Mb region in and around *AR*. Although not demonstrating the largest allele frequency differences between AGA cases and controls, it was argued that, on functional criteria, the GGN repeat in exon 1 may in fact be causal (Hillmer et al. 2005).

To address this important hypothesis we have extended our original study by genotyping the rs6152 SNP and the CAG and GGN repeats in more families and in two generations, and we have employed more sophisticated statistical modelling that takes into account the categorical nature of the baldness phenotype.

Materials and methods

Study population and phenotyping

We analysed the three *AR* exon 1 polymorphisms in approximately 1,200 males from the parental and offspring generations of the Victorian Family Heart Study (VFHS) for whom data on degree of AGA was available. Our original analyses used 163 men from the VFHS and the description of that study and of the population-based Caucasian VFHS cohort have been published in detail elsewhere (Ellis et al. 2001; Harrap et al. 2000). Briefly, self-reported AGA phenotypes were gathered by way of questionnaires, and the accuracy of this approach has been validated (Ellis et al. 1998; Taylor et al. 2004). Participants were asked to assess and report a category that best reflected their degree (if any) of balding according to diagrams from the Hamilton–Norwood Scale (Hamilton 1951; Norwood 1975) as follows: Type 1—no evidence of hair recession, Type II—minimal frontal hair recession, Type III—cosmetically significant frontal recession, Type III vertex—cosmetically significant frontal recession coupled with vertex hair loss, Types IV through VII—increasing degrees of frontal and vertex loss (Ellis et al. 1998, 2001). The phenotypic characteristics of males included in this analysis are shown in Table 1.

Genotyping

Genotyping of the rs6152 SNP was performed as previously described (Ellis et al. 2001). The CAG and

GGN repeats were amplified in a multiplex reaction using the following primers: CAGF 5'-CCAAGCTCA AGGATGGAA-3', CAGR 5'-GAAGGTTGCTGTT CCTCA-3', GGNF 5'-TGGCACACTCTCTTCACA-3', GGNR 5'-GATAGGGCACTCTGCTCA-3'. Forward primers were fluorescently labelled. Approximately 50 ng of DNA from each participant was amplified in a 5 µl reaction containing 0.5 mM of each primer, 1 × standard PCR buffer (Bioline), 1 × GC rich PCR buffer (Roche), 1.5 mM MgCl₂ (Bioline), 250 mM dNTPs (Applied Biosystems), and 0.2 unit Immolase DNA polymerase (Bioline). Thermal conditions required for the reaction were 95°C for 10 min, followed by 35 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min, followed by a final extension time of 72°C for 20 min. ET400-R size standard (GE/Amersham Biosciences) was added to each PCR product and samples were denatured at 95°C for 5 min. The sizes of the resulting PCR products were determined using the MegaBACE 1000 DNA analysis system and Genetic Profiler genotyping software (GE/Amersham Biosciences). A number of individuals in the VFHS were previously genotyped for the CAG and GGN repeats by direct sequencing (Ellis et al. 2001). The sequencing results of these individuals were compared to the results of the fluorescent multiplex method described above, confirming the validity of this method to accurately determine the number of CAG and GGN repeats.

Statistical approach

Although the Hamilton–Norwood baldness scale is graded according to the degree of baldness, this scale does not represent a continuous biological measurement and should not be treated as such. However, information may be lost if the scale is dichotomised (bald/not bald). Instead, we used a statistical modelling approach that treats the scale as comprising ordinal categories. These models permit estimation of relationships between genotypes and baldness categories taking into account the ordinal nature of the outcome. They also permit adjustments for key covariates such as age, body habitus and smoking status (see below) as well as adjusting for familial relationships between subjects.

The main phenotype analysed was a four-level ordinal categorical baldness variable (1 = none, 2 = frontal, 3 = vertex, 4 = frontal plus vertex) formed by combining categories of the eight-level Hamilton–Norwood scale as shown in Table 1. This simplified categorisation approximates that used by Hayes et al. (2005).

Table 1 Phenotypic and genotypic characteristics for 1,203 men with at least one measured genotype

Variable	<i>N</i> (%)	<i>N</i> parents (%)	<i>N</i> offspring (%)	Combined category
Baldness type (Hamilton–Norwood)				
I	515 (43)	152 (22)	363 (70)	None (1)
II	226 (19)	126 (18)	100 (19)	Frontal only (2)
III	76 (6)	49 (7)	27 (5)	Frontal only (2)
III—Vertex	113 (9)	97 (14)	16 (3)	Vertex (3)
IV	101 (8)	92 (13)	9 (2)	Vertex (3)
V	69 (6)	66 (10)	3 (1)	Frontal and vertex (4)
VI	63 (5)	63 (9)	0 (0)	Frontal and vertex (4)
VII	40 (3)	38 (6)	2 (<1)	Frontal and vertex (4)
rs6152 genotype				
A	145 (12)	89 (13)	56 (11)	
G	1032 (88)	580 (87)	452 (89)	
Number of CAG repeats				
9–19	222 (19)	134 (21)	88 (27)	1
20	137 (12)	82 (12)	55 (16)	2
21	234 (20)	141 (21)	93 (28)	0 (baseline)
22	93 (8)	55 (8)	38 (11)	4
23	145 (12)	76 (11)	69 (15)	5
24	132 (11)	69 (10)	163 (14)	6
25–33	200 (17)	105 (16)	95 (21)	7
Number of GGN repeats				
11–22	98 (8)	46 (7)	52 (10)	1
23	625 (53)	351 (53)	274 (54)	0 (baseline)
24	371 (32)	221 (33)	150 (30)	3
25–29	76 (6)	46 (7)	30 (6)	4

The SNP rs6152 was a binary variable, while the two triplet repeats were considered continuous covariates. Other measured covariates included age, height, weight (all measured as continuous and centred at or close to the mean value) and smoking status (categorised as never smoked (reference category), ex-smoker, current smoker with less than or equal to 20 cigarettes (or equivalent) per day, or current smoker with more than 20 cigarettes per day). Generation (indicating whether each individual was from the parental or offspring generation of the VFHS) was also included as a binary covariate.

The basic model was a proportional odds logistic regression (POLR) model (McCullagh and Nelder 1989). For an ordinal outcome with four categories, the model essentially consists of three standard logistic regression models fitted simultaneously, with the four categories collapsed into two in each logistic model and with the effect size (measured as an odds ratio, OR) estimated for three specific comparisons. In each of these comparisons the lower (less bald) combined category is the reference group, so that an OR > 1 for a covariate means that the odds of being in a higher (more bald) category are increased for those with a greater value for that covariate. Comparison A estimated the OR for none versus any baldness (1 vs. 2, 3, 4), comparison B estimated the OR for none or frontal baldness versus vertex or frontal plus vertex (1, 2 vs. 3,

4) and comparison C estimated the OR for none, frontal or vertex baldness versus frontal plus vertex baldness (1, 2, 3 vs. 4). In the standard POLR model, these three ORs are forced to be equal. This assumption was checked using the Brant test (Brant 1990) for all covariates and may be assumed to have been met for the models described below unless stated otherwise. Where required, partial proportional odds (PPOLR) models were fitted, with non-proportional ORs only for covariates for which the proportional odds assumption was not met.¹

In more precise terms, let the phenotype for individual *i* be y_i , where y_i is one of $\{1, 2, \dots, K\}$, representing the *K* levels of an ordinal categorical outcome (here *K* = 4), and let $P(y_i = k) = p_{ik}$. The cumulative probabilities $P(y_i > k) = 1 - \sum_{c=1}^k p_{ic} = q_{ik}$ are then modelled, capturing the ordinal nature of the outcome;

$$\text{logit}(q_{ik}) = \beta_{0k} + \beta^T X_i$$

where β_{0k} is a “threshold” parameter and $\beta_{01} > \beta_{02} > \dots > \beta_{0K-1}$, β^T is a vector of log-odds ratios (log-ORs) and X_i is a vector of covariates for individual *i*. As stated above, the proportional odds assumption may be

¹ Although this means that the lines will eventually cross at some point, if this occurs outside the range of the data this is unlikely to be a problem. This was not found to be a problem in our data; all predicted probabilities were within the range [0, 1].

relaxed, resulting in a set of estimates $\{\beta_1^T, \beta_2^T, \beta_3^T\}$ for Comparisons A, B and C, respectively, instead of a single estimate β^T .

All models were fitted using the statistical package Stata v9 (StataCorp 2005, <http://www.stata.com>), and the Stata add-on *gologit2* (Williams 2006) was used extensively.

Models including a single genetic marker were initially fitted, then models which included two and then all three markers were investigated. All models were adjusted for the covariates described above. Model selection was carried out using a combination of the likelihood ratio test (where models were nested) and Akaike's information criterion (AIC). Interactions between the covariate generation (father/son) and all other covariates were considered but are not explicitly described below unless their inclusion significantly improved the model.

The POLR and PPOLR models assume independence between individuals. Since our data included many relative pairs and trios and some larger family groups, the *cluster* and *robust* options in Stata were used to adjust the final models for any within-family correlation present. The two generations of VFHS individuals are referred to as "fathers" and "sons" throughout this paper although the precise genetic relationships are not utilised.

Results

A total of 1,123 men were successfully genotyped for all three polymorphisms, and 1,203 men had at least one genotyped polymorphism (Table 1). In total, 703 VFHS nuclear families were represented. Although most "families" in this study consisted of a single individual (46%), 273 families (39%) included 2 men (272 father–son pairs and 1 brother–brother pair), 91 (13%) included 3 men (mostly a father and 2 sons) and 15 (2%) included 4 men. Since the VFHS deliberately over-sampled families with twins (in either the parental or offspring generations) (Harrap et al. 2000), 18 pairs of MZ twins and 9 pairs of DZ twins (all sons) were included and several of the nuclear families were related, resulting in 685 extended families, 6 of which included 5 or 6 family members. The mean age for fathers was 55.2 years (SD 6.4, range 43–72) and for sons was 23.9 years (SD 3.8, range 18–32). The majority of sons (70%) reported no hair loss (Table 1) while 60% of fathers reported baldness of at least Type III. The minimum sample size (for models including all three genetic polymorphisms and all covariates) was 1,118.

The frequency of the rs6152 A allele was 12.3%. The numbers of CAG repeats observed ranged from 8 to 32, while all GGN repeats were between 11 and 29 (Table 1). For the GGN repeat, the majority (85%) of men had either 23 or 24 repeats.

Age was strongly positively associated with baldness in all models. The effect of age differed for fathers and sons ($P < 0.001$) and the proportional odds assumption was not met for sons. The OR for an increase in age of 10 years was 1.76 ($P < 0.001$) for fathers and 3.17 ($P < 0.001$), 20.99 ($P < 0.001$) and 1.24 ($P = 0.90$) for comparisons A, B and C, respectively, for sons indicating that age has most effect on the probabilities of none/frontal baldness compared with any vertex baldness.

The SNP rs6152 was strongly associated with baldness ($P < 0.0001$) after adjustment for covariates. The proportional odds assumption was not met for this covariate and the OR decreased across baldness categories (Table 2), indicating that this SNP has more effect on the probability of being in a moderate category compared with a severe baldness category, than on the probability of having no baldness compared with at least frontal baldness. Predicted probabilities of being in each of the four baldness categories for age and the two rs6152 genotypes (for non-smokers of average height and weight for their generation) are shown in Fig. 1. The predicted proportion of men with at least Type II (at least category 2) baldness clearly increases as age increases, but this increase is more rapid for men with the rs6152 G allele. A much larger proportion of men with the G allele are also predicted to have at least Type V baldness in comparison to men with the A allele (at age 60, 27% compared with 8%).

CAG showed a weak association with baldness (OR = 0.97, $P = 0.09$). When CAG was categorised as shown in Table 1 (using the most common category of 21 repeats as the baseline), no association was observed ($P = 0.7$).

The proportional odds assumption was not met for GGN ($P = 0.04$) and there was not a significant association with baldness ($P = 0.09$, Table 2). When only men with 23 or 24 repeats were included in the analysis and GGN was treated as a binary variable, results were similar (OR = 0.82, $P = 0.14$). Allowing generation by categorical GGN effects for this reduced sample had a weak effect ($P \approx 0.05$). Further investigation revealed that the association appeared to only be present for fathers, with a parallel OR of 0.68 ($P = 0.014$) for 24 repeats compared with 23 repeats, while for sons the OR was 1.22 ($P = 0.38$). However, continuous GGN (allowing the full range of values) was not associated with baldness even when only fathers were considered (OR = 1.07, $P = 0.13$).

Table 2 Results of statistical modelling

Model	Odds ratio (95% confidence interval) (<i>P</i> value)		
	Comparison A 1 vs. (2, 3, 4)	Comparison B (1, 2) vs. (3, 4)	Comparison C (1, 2, 3) vs. 4
rs6152 only (<i>N</i> = 1172)			
rs6152 (A vs. G)	0.54 (0.36, 0.81) <i>P</i> = 0.003	0.28 (0.17, 0.46) <i>P</i> < 0.001	0.24 (0.11, 0.53) <i>P</i> < 0.001
CAG only (<i>N</i> = 1158)			
CAG	0.97 (0.93, 1.01) <i>P</i> = 0.09	0.97 (0.93, 1.01) <i>P</i> = 0.09	0.97 (0.93, 1.01) <i>P</i> = 0.09
GGN only (<i>N</i> = 1165)			
GGN	0.99 (0.91, 1.08) <i>P</i> = 0.82	1.00 (0.92, 1.09) <i>P</i> = 0.99	1.14 (1.01, 1.28) <i>P</i> = 0.03
All 3 polymorphisms (<i>N</i> = 1118)			
rs6152	0.53 (0.34, 0.82) <i>P</i> = 0.004	0.25 (0.15, 0.42) <i>P</i> < 0.001	0.19 (0.08, 0.42) <i>P</i> < 0.001
CAG	0.94 (0.90, 0.98) <i>P</i> = 0.006	0.94 (0.90, 0.98) <i>P</i> = 0.006	0.94 (0.90, 0.98) <i>P</i> = 0.006
GGN	1.03 (0.94, 1.13) <i>P</i> = 0.55	1.06 (0.96, 1.16) <i>P</i> = 0.25	1.21 (1.07, 1.36) <i>P</i> = 0.002
All 3 polymorphisms, GGN = 23 and GGN = 24 only (<i>N</i> = 953)			
rs6152	0.48 (0.29, 0.79) <i>P</i> = 0.003	0.27 (0.15, 0.48) <i>P</i> < 0.001	0.19 (0.08, 0.45) <i>P</i> < 0.001
CAG	0.93 (0.89, 0.98) <i>P</i> = 0.004	0.93 (0.89, 0.98) <i>P</i> = 0.004	0.93 (0.89, 0.98) <i>P</i> = 0.004
GGN	1.16 (0.84, 1.57) <i>P</i> = 0.32	1.16 (0.84, 1.57) <i>P</i> = 0.32	1.16 (0.84, 1.57) <i>P</i> = 0.32

All estimates are adjusted for age, generation, age by generation interaction, height, weight and smoking status

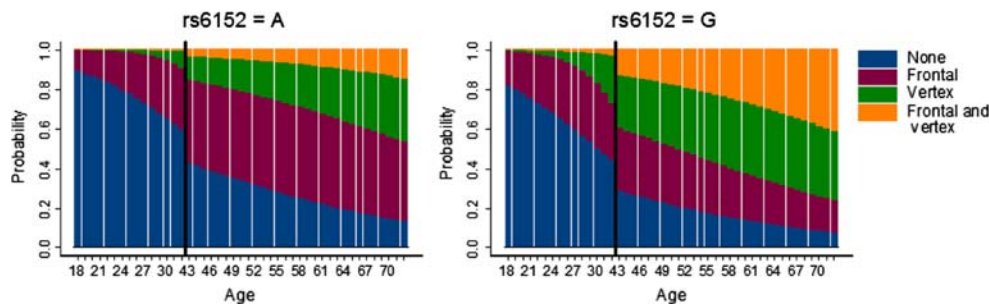


Fig. 1 Predicted probabilities for each baldness category for men with rs6152 A and G genotypes (Non-smoker of average height and weight for their generation assumed). A discontinuity

in each figure (which is due to the lack of men aged between 33 and 42 years in the study) is indicated by a *heavy black line*

Including CAG in a model which included rs6152 was a significant improvement ($P = 0.008$) but including GGN in addition to rs6152 was not ($P = 0.07$). However, when only fathers were considered, including GGN substantially improved the model ($P = 0.005$). Allowing non-proportional odds for GGN was not a significant improvement ($P = 0.19$), but there was some evidence of increasing ORs with increasing baldness type (OR = 1.09, $P = 0.22$; OR = 1.12, $P = 0.05$; OR = 1.23, $P = 0.002$).

Results for models which included all three polymorphisms are shown in Table 2. All three polymorphisms appeared to be associated with baldness and both GGN and CAG were more significant when rs6152 was included in the model, regardless of whether or not the other repeat was also included. A larger number of CAG repeats reduced the odds of baldness, and the rs6152 G allele and a larger number of GGN repeats increased the odds of baldness (Table 2).

Similar trends were evident when the repeats were analysed as categorical variables, but the association between GGN and baldness disappeared when analyses were restricted to individuals with only 23 or 24 GGN repeats (Table 2).

The robust and cluster-adjusted standard errors were very similar to the naïve standard errors (which were estimated assuming independence between individuals), regardless of whether nuclear or extended family was used to define clusters. For example, the nuclear family-based cluster-adjusted confidence intervals for the rs6152 ORs were (0.35, 0.84), (0.17, 0.46) and (0.11, 0.54) for comparisons A, B and C, respectively, compared with (0.36, 0.81), (0.17, 0.46) and (0.11, 0.53) when naïve SEs were used. Results reported in Table 2 used naïve SEs.

Analyses of different categorisations of MPB, including the original eight-level scale (with some covariates excluded due to small numbers), produced

qualitatively similar results, particularly for rs6152 (full results not shown). Different categorisations of CAG and GGN, including as binary variables, also had little effect (results not shown). Exclusion of the covariates height, weight and smoking status from the models also had little effect on the significance or estimated odds ratios of the polymorphisms (full results not shown).

Discussion

We have demonstrated in this large population-based Caucasian cohort that whilst the non-functional rs6152 *AR* SNP is strongly associated with male pattern baldness (AGA), neither the polyglutamine (CAG) nor polyglycine (GGN) repeat polymorphisms appear to be independently responsible for this association, and that the weak associations seen in some analyses are likely to be due to linkage disequilibrium with more relevant variants. Our results contrast those presented by Hillmer et al. in which they showed a strong association between AGA and a shorter polyglycine repeat (GGN-23) both individually, and as part of haplotypes containing polymorphism across the *AR* region (Hillmer et al. 2005). It is possible that these differing results are due to differences in the strength of linkage disequilibrium (LD) between rs6152 and the GGN polymorphism in our Australian-based population compared to the German population studied by Hillmer et al. Exact LD measures between polymorphisms are not quoted by Hillmer et al, but their LD intensity plot indicates strong LD between rs6152 and GGN using the χ^2 statistic. In our Australian population, we also see strong LD using both the D' (Hedrick 1987) and χ^2 ($P < 0.0001$) measures. However in relation to association studies, the most relevant measure of LD may be one that measures the degree of allelic predictiveness of one polymorphism from another. For multi-allelic markers such as GGN, the Uncertainty Coefficient U (calculated using the GOLD program) (Abecasis and Cookson 2000) is one such measure of predictiveness. Using this method, we find little LD (0.18) between rs6152 and GGN.

Interestingly, in haplotype analyses of the variants analysed, Hillmer et al. generally demonstrated that haplotypes carrying the GGN-24 allele were more frequent in individuals without AGA. However one such haplotype in significant frequency in the population was not different between cases and controls (14.3% vs. 16.5%, respectively). The authors explain this anomaly by suggesting the existence of further polymorphisms that modulate the effect of the GGN repeat length (Hillmer et al. 2005). Our results, in

conjunction with Hillmer et al.'s findings, might rather suggest that the polyglycine repeat is not functionally significant in AGA, but that it is in fact other variation in or around *AR* that is contributing to the phenotype.

As has been demonstrated by Hillmer et al., and as is evident from analysis of data from the HapMap database (<http://www.hapmap.org>), strong LD is maintained across the gene-poor region of approximately 1 Mb on the X chromosome that contains only *AR*. SNPs throughout this region have also been shown to be associated with AGA, and some that lie upstream of the *AR* coding region may be more strongly associated than rs6152 or the triplet repeats (Ellis et al. 2005; Hillmer et al. 2005). The long range of LD in this region suggests that functional variants may occur anywhere within exons, introns, promoter or upstream/downstream non-coding sequences. In the absence of obvious coding region functional variants, comparative and physiological genomic methods will likely be needed to identify functionally important non-coding sequence in which to search for variants relevant to AGA. Such analyses might also be relevant to delineating the role of *AR* in the myriad of other complex conditions that have been associated with *AR*, such as prostate cancer.

In conclusion, results of this large two generational association analysis of the exon 1 CAG and GGN polymorphic repeats suggest that these variants are not responsible for susceptibility to AGA conferred by *AR*. Strong LD across the *AR* region indicates that functional variants may exist anywhere in non-coding sequence within, or surrounding, the gene. Further analysis and understanding of the importance of non-coding sequence in this region is required in order to identify variation that may be functionally relevant to the heritability of male pattern baldness.

Acknowledgments We thank Dr Lyle Gurrin for thoughtful and useful suggestions regarding statistical analyses. We also thank Ms Margaret Stebbing, Professor John Hopper, Professor Graham Giles, the general practitioners and research nurses for their contributions to recruitment of VFHS study participants, Dr Zilla Wong for sample management, and Angela Lamantia for DNA extraction. We gratefully acknowledge support from a University of Melbourne Early Career Researcher Grant (KJS). KJS is funded by NHMRC (Project Grant 400255).

References

- Abecasis GR, Cookson WO (2000) GOLD—graphical overview of linkage disequilibrium. *Bioinformatics* 16:182–183
- Beilin J, Ball EM, Favaloro JM, Zajac JD (2000) Effect of the androgen receptor CAG repeat polymorphism on transcriptional activity: specificity in prostate and non-prostate cell lines. *J Mol Endocrinol* 25:85–96

- Brant R (1990) Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics* 46:1171–1178
- Chamberlain NL, Driver ED, Miesfeld RL (1994) The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Res* 22:3181–3186
- Choong CS, Kempainen JA, Zhou ZX, Wilson EM (1996) Reduced androgen receptor gene expression with first exon CAG repeat expansion. *Mol Endocrinol* 10:1527–1535
- Ding D, Xu L, Menon M, Reddy GP, Barrack ER (2004) Effect of a short CAG (glutamine) repeat on human androgen receptor function. *Prostate* 58:23–32
- Ding D, Xu L, Menon M, Reddy GP, Barrack ER (2005) Effect of GGC (glycine) repeat length polymorphism in the human androgen receptor on androgen action. *Prostate* 62:133–139
- Ellis JA, Stebbing M, Harrap SB (1998) Genetic analysis of male pattern baldness and the 5 α -reductase genes. *J Invest Dermatol* 110:849–853
- Ellis JA, Stebbing M, Harrap SB (2001) Polymorphism of the androgen receptor gene is associated with male pattern baldness. *J Invest Dermatol* 116:452–455
- Ellis JA, Scurrah KJ, Harrap SB (2005) Haplotype analysis of the androgen receptor in male pattern baldness using HapMap tag-SNPs (abstract). American society of human genetics 55th annual meeting, Salt Lake City, Utah
- Hamilton JB (1942) Male hormone stimulation is a prerequisite and an incitant in common baldness. *Am J Anat* 71:451–480
- Hamilton JB (1951) Patterned loss of hair in man; types and incidence. *Ann N Y Acad Sci* 53:708–728
- Harrap SB, Stebbing M, Hopper JL, Hoang HN, Giles GG (2000) Familial patterns of covariation for cardiovascular risk factors in adults: the Victorian Family Heart Study. *Am J Epidemiol* 152:704–715
- Hayes VM, Severi G, Eggleton SA, Padilla EJ, Southey MC, Sutherland RL, Hopper JL, Giles GG (2005) The E211 G > A androgen receptor polymorphism is associated with a decreased risk of metastatic prostate cancer and androgenic alopecia. *Cancer Epidemiol Biomarkers Prev* 14:993–996
- Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. *Genetics* 117:331–341
- Hillmer AM, Hanneken S, Ritzmann S, Becker T, Freudenberg J, Brockschmidt FF, Flaquer A, Freudenberg-Hua Y, Jamra RA, Metzen C, Heyn U, Schweiger N, Betz RC, Blaumeiser B, Hampe J, Schreiber S, Schulze TG, Hennies HC, Schumacher J, Propping P, Ruzicka T, Cichon S, Wienker TF, Kruse R, Nothen MM (2005) Genetic variation in the human androgen receptor gene is the major determinant of common early-onset androgenetic alopecia. *Am J Hum Genet* 77:140–148
- Kazemi-Esfarjani P, Trifiro MA, Pinsky L (1995) Evidence for a repressive function of the long polyglutamine tract in the human androgen receptor: possible pathogenetic relevance for the (CAG) $_n$ -expanded neuronopathies. *Hum Mol Genet* 4:523–527
- Kuster W, Happle R (1984) The inheritance of common baldness: two B or not two B? *J Am Acad Dermatol* 11:921–926
- Levy-Nissenbaum E, Bar-Natan M, Frydman M, Pras E (2005) Confirmation of the association between male pattern baldness and the androgen receptor gene. *Eur J Dermatol* 15:339–340
- McCullagh P, Nelder J (1989) Generalised linear models. Chapman and Hall, London
- Norwood OT (1975) Male pattern baldness: classification and incidence. *South Med J* 68:1359–1365
- Nyholt DR, Gillespie NA, Heath AC, Martin NG (2003) Genetic basis of male pattern baldness. *J Invest Dermatol* 121:1561–1564
- Taylor R, Matassa J, Leavy JE, Fritschi L (2004) Validity of self reported male balding patterns in epidemiological studies. *BMC Public Health* 4:60
- Williams R (2006) Generalised ordered logit/partial proportional odds models for ordinal dependent variables. *Stata J* 6:58–85