Am. J. Hum. Genet. 77:140–148, 2005

Report

Genetic Variation in the Human Androgen Receptor Gene Is the Major Determinant of Common Early-Onset Androgenetic Alopecia

Axel M. Hillmer,[1,2] Sandra Hanneken,[4] Sibylle Ritzmann,[4] Tim Becker,[3] Jan Freudenberg,[2] Felix F. Brockschmidt,[2] Antonia Flaquer,[3] Yun Freudenberg-Hua,[2] Rami Abou Jamra,[2] Christine Metzen,[4] Uwe Heyn,[2] Nadine Schweiger,[4] Regina C. Betz,[2,5] Bettina Blaumeiser,[5] Jochen Hampe,[6] Stefan Schreiber,[6] Thomas G. Schulze,[7] Hans Christian Hennies,[8] Johannes Schumacher,[2] Peter Propping,[2] Thomas Ruzicka,[4] Sven Cichon,[1,5] Thomas F. Wienker,[3] Roland Kruse,[4] and Markus M. Nöthen[1,5]

[1]Department of Genomics, Life & Brain Center, and [2]Institute of Human Genetics, and [3]Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn; [4]Department of Dermatology, University of Düsseldorf, Düsseldorf; [5]Department of Medical Genetics, University of Antwerp, Antwerp; [6]Institute for Clinical Molecular Biology, Christian-Albrechts-Universität zu Kiel, Kiel, Germany; [7]Division of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Mannheim; and [8]Department of Molecular Genetics and Gene Mapping Center, Max Delbrück Center for Molecular Medicine, Berlin

Androgenetic alopecia (AGA), or male-pattern baldness, is the most common form of hair loss. Its pathogenesis is androgen dependent, and genetic predisposition is the major requirement for the phenotype. We demonstrate that genetic variability in the androgen receptor gene (*AR*) is the cardinal prerequisite for the development of early-onset AGA, with an etiological fraction of 0.46. The investigation of a large number of genetic variants covering the *AR* locus suggests that a polyglycine-encoding GGN repeat in exon 1 is a plausible candidate for conferring the functional effect. The X-chromosomal location of *AR* stresses the importance of the maternal line in the inheritance of AGA.

Androgenetic alopecia (AGA [MIM 109200]), or male-pattern baldness, is characterized by a defined pattern of hair loss from the scalp (Hamilton 1951). In whites, the proportion of affected males increases steadily with age, so that a male in his 50s has a 50% chance of having some degree of AGA (Hamilton 1951). Association of AGA with a variety of clinical phenotypes has been suggested, including coronary heart disease (Lotufo et al. 2000), benign prostatic hyperplasia (Hawk et al. 2000), prostate cancer (Oh et al. 1998), and disorders associated with insulin resistance (Matilainen et al. 2000). Androgen dependence is an important characteristic of AGA, and genetic disposition, which is assumed to be polygenic, plays the most substantial role in the development of AGA (Küster and Happle 1984; Ellis et al. 1998; Nyholt et al. 2003). Although the only factor known to influence onset age in patients with AGA is genetic predisposition, no systematic approach has hitherto been undertaken, to our knowledge, to identify the contributing genes.

As part of a genomewide linkage study of AGA, we investigated linkage to markers covering the X chromosome. The sample consisted of 95 families in which at least two brothers had early-onset AGA (391 genotyped individuals, including 201 affected men). We obtained evidence of linkage in chromosomal region Xq12-22 (nonparametric linkage [NPL] score of 2.70 [fig. 1]) (for X-chromosomal marker information, see table 1). This region contains the *AR* gene (MIM 313700), which is an obvious candidate for explaining the development of AGA, and an association with this region has been suggested elsewhere (Ellis et al. 2001), on the basis of results from the investigation of three polymorphic sites. In the present study, we have systematically explored the contribution of the *AR* gene to the development of early-onset AGA.

Received March 21, 2005; accepted for publication April 27, 2005; electronically published May 18, 2005.

Address for correspondence and reprints: Dr. Markus M. Nöthen, Department of Genomics, Life & Brain Center, University of Bonn, Sigmund-Freud-Straße 25, 53127 Bonn, Germany. E-mail: markus.noethen@uni-bonn.de

0002-9297/2005/7701-0015$15.00

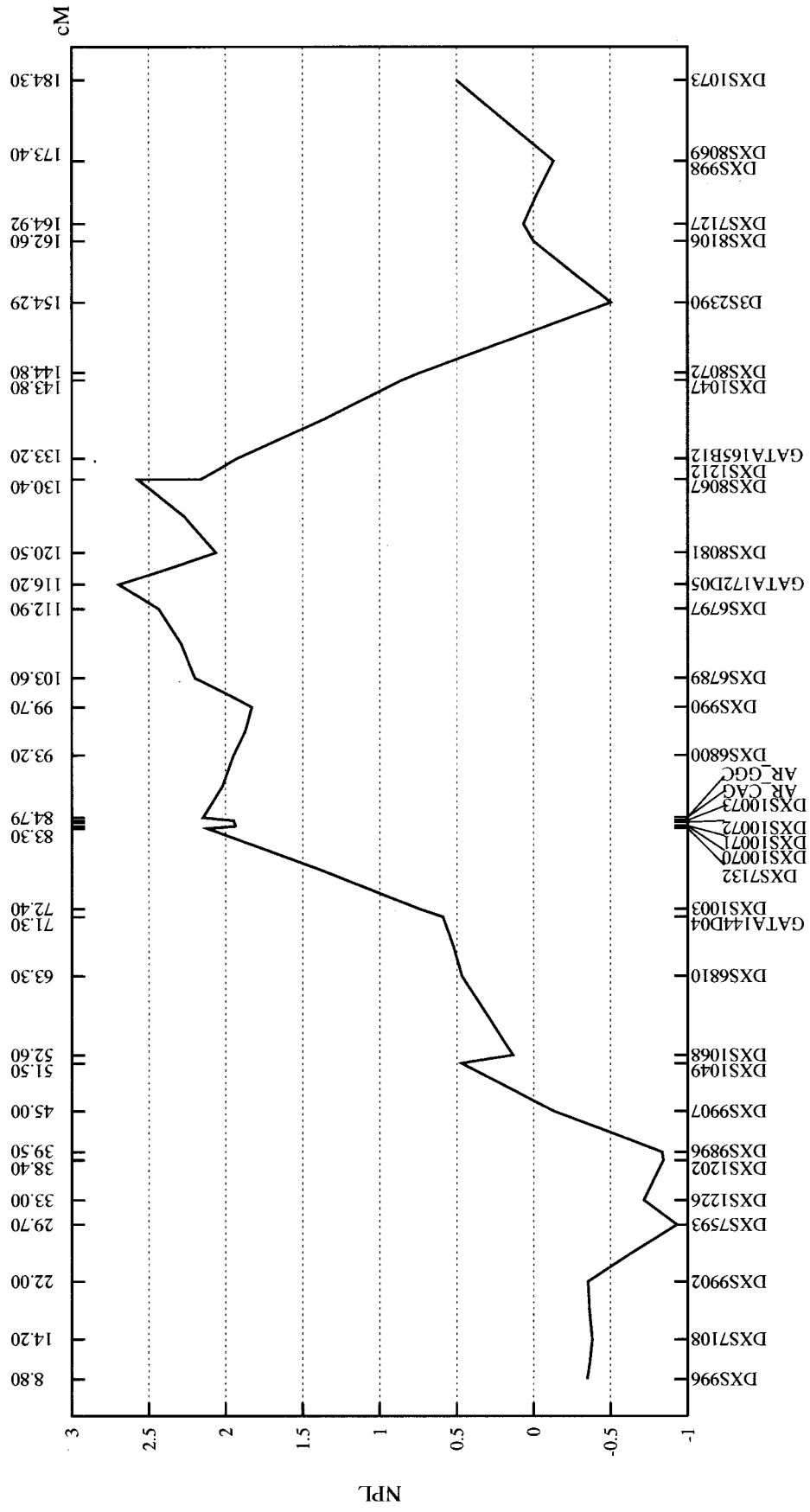

Figure 1 Multipoint NPL analysis of chromosome X, calculated by the Allegro v1.2 software (Gudbjartsson et al. 2000). The *X*-axis is the chromosome location (*top*, cM; *bottom*, STR marker), and the *Y*-axis is the LOD score.

Table 1

STR Markers Used for Linkage Analysis of Chromosome X

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

We tested 39 SNPs, two STRs (a polyglutamine-encoding CAG repeat and a polyglycine-encoding GGN repeat in exon 1 of *AR*), and one biallelic insertion polymorphism (*XARx8insA*), in a range of 4.4 Mb at the *AR* locus, for association with AGA (fig. 2). Our analysis included the three previously studied variants (Ellis et al. 2001): the CAG and GGN repeats and *Stu*I RFLP (*rs6152*). *Taq*Man assays for genotyping of SNPs, listed in table 2, were designed by Applied Biosystems. PCRs were performed using 2.5 ng genomic DNA, Ampli*Taq* Gold DNA polymerase (Applied Biosystems), and an annealing temperature of 60°C for 45 cycles on Biometra T1 (Biometra) or Perkin Elmer GeneAmp 9700 (Applied Biosystems) thermocyclers. Fluorescence was measured with an ABI Prism 7900HT sequence detection system (Applied Biosystems). The fragment lengths of the CAG repeat, the GGN repeat, and *XARx8insA* were determined using the following fluorescence-labeled primers for PCR: CAG-F 5′-TCCAGAATCTGTTCCAGAGCGTGC-3′ and CAG-R 5′-GCTGTGAAGGTTGCTGTTCCTCAT-3′ (La Spada et al. 1991), GGN-F 5′-CCTGGCACACTCTCTTCACA-3′ and GGN-R 5′-GGATAGGGCACTCTGCTCAC-3′, and *XARx8insA*-F 5′-CACGGGAAGTTTAGAGAGCT-3′ and *XARx8insA*-R 5′-TCACCTTCTCGTCACTATTG-3′. Forty nanograms of genomic DNA was used in PCRs with Ampli*Taq* DNA polymerase (Applied Biosystems) in 38 cycles on a PTC-200 (MJ Research). The annealing temperatures were 62°C–60°C (touchdown PCR) for the CAG repeat, 58°C for the GGN repeat, and 63°C–55°C (touchdown PCR) for *XARx8insA*. MasterAmp PCR PreMix G (Epicentre) was used to amplify GGN fragments. Fragment lengths of the amplified products were analyzed on an ABI Prism 377 DNA sequencer (Applied Biosystems). Double-strand sequencing of genomic regions was performed with the ABI Prism BigDye Terminator Cycle Sequencing kit, version 2.0 (Applied Biosystems), and the ABI Prism 3730 DNA analyzer (Applied Biosystems).

For case-control analysis, we compared allele frequencies in 198 males with early-onset AGA (including 95 unrelated and randomly chosen affected individuals from the linkage-analysis families), 188 control individuals, and 157 unaffected individuals. For family-based association analysis, we studied 179 families containing at least one affected male in the youngest generation. The family-based association sample included the 95 families from the linkage analysis and overlapped with the case sample, for 179 individuals. All affected males were aged <40 years (mean [± SD] 32.0 ± 5.2) and had AGA that was representative of the most severely affected 10% of the distribution for the respective age class, on the basis of the classification of Hamilton (1951) (modified by Norwood [1975]). AGA classification, age, and ethnicity were the exclusive criteria used to select individuals for inclusion in the case sample. Unaffected males were aged >60 years (mean 67.9 ± 6.2) and without AGA, representing the least-affected 20% of the distribution for this age class. Families and unrelated individuals both with and without AGA were recruited through various sources, including press reports and advertisements in magazines, newspapers, and placards. Control individuals were male blood donors from the blood transfusion center of the University Hospital Bonn, from whom information was available only on sex, age in years (mean 29.4 ± 8.6), and ethnicity. EDTA anticoagulated venous blood samples were collected from all individuals, and lymphocyte DNA was isolated by salting out with saturated NaCl solution (Miller et al. 1988). All participants were of German descent. The study was approved by the ethics committee of the University of Bonn, and informed consent was obtained from all participants.

Association analysis was conducted using a modification of the FAMHAP software (Becker and Knapp 2004*a*, 2004*b*) for X-chromosomal data. Case-control SNP single-marker analysis was performed using the χ^2 distribution of the 2 × 2 contingency table, and multiallelic markers were evaluated with the permutational version of the χ^2 test. For case-control haplotype analysis, P values were calculated from the χ^2 distribution with $n-1$ df of the respective likelihood-ratio test (n = number of different haplotypes). The family data were analyzed with the permutation-based association test for nuclear families (Zhao et al. 2000; Knapp and Becker 2003). For each marker (single-locus analysis) and each marker combination (haplotype analysis), we used 10^{10} permutation replicates.

Pairwise distances between haplotypes were calculated as allele mismatches. By resampling markers randomly with replacement, 100 bootstrapped data sets were generated as input for the program neighbor contained in the PHYLIP package.

A region of 1 Mb showed strong association with the lowest P value of 2.1×10^{-12} for *rs10521339* in the case-control analysis of affected and unaffected individuals (table 3). As expected, the association is stronger for comparisons between individuals with AGA and individuals without AGA than between individuals with AGA and an unselected control sample (table 3). We also performed a separate analysis of the 103 cases that were not included in the linkage analysis and found that the association was also present in this sample (data not shown). The association is further supported by family-based analysis, for which SNP *rs938059* shows the low-

est P value (4.03×10^{-6}) (table 3). *AR* is the only known gene in the strongly associated 1-Mb–spanning region (fig. 2 and table 3). The X-linked ectodysplasin-A2 receptor (*XEDAR*), which is located 900 kb 5′ of *AR*, is outside this region (fig. 2 and table 3). The significance of the association decreases within the 3′ part of the 180-kb–spanning *AR* gene, and oligophrenin 1 (*OPHN1*), which is located 320 kb 3′ of *AR*, is not within the block of strongest association (fig. 2 and table 3). *rs6152* ($P = 6.66 \times 10^{-10}$ [table 3]) in exon 1 of *AR* corresponds to the *Stu*I RFLP, for which an association has been described by Ellis et al. (2001).

The long range of the associated SNPs implies the presence of a large haplotype block, visible in fig. 2*B*. In principle, this is not unexpected in a location close to the centromere, where recombination events occur at a relatively low frequency (Nagaraja et al. 1997). To test whether the size of the haplotype block stands out even in comparison with X-chromosomal loci with similar low recombination frequencies, we analyzed average pairwise linkage disequilibrium (LD) (measured by $|D'|$) between SNPs retrieved from the HapMap Project database in 1-cM–sized windows. Average pairwise LD was found to be inversely correlated with recombination rate (Spearman's $\rho = -0.598$; $P < .001$). An average LD higher than that at the *AR* locus was displayed by only the six windows covering the centromere; these windows showed distinctively smaller recombination rates (fig. 3*A*). This result might suggest that the predominant *AR* haplotypes are evolutionarily more recent, perhaps indicating positive selective pressure acting at this locus (Bamshad and Wooding 2003). Since androgens mediate a wide range of developmental and physiological responses through the androgen receptor (AR) and are especially important in the male reproductive system (Lee and Chang 2003), it is conceivable that variability in *AR* can have an impact on selection. In accordance with this hypothesis, the haplotype with the highest frequency (0.45 [fig. 3*B*]) (which also confers risk of AGA) in the German population seems to be evolutionarily recent, as indicated by the low sequence identity with the ancestral haplotype (fig. 3*B*).

Sequencing of the transcribed region and of 3.4 kb of conserved sequences in the 5′ region and intron 1 of *AR* in 12 individuals revealed only two additional variants of the associated haplotype (*XARx7_01* and *XARx8insA* [table 3]). The additional variants were noncoding and did not show a stronger association than other tested markers. Of the two repeat polymorphisms in the coding region, the CAG repeat was not associated with AGA (affected/unaffected global P value of .1), whereas the GGN repeat was highly associated (affected/

Table 2

Oligonucleotides for *Taq*Man Assays

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

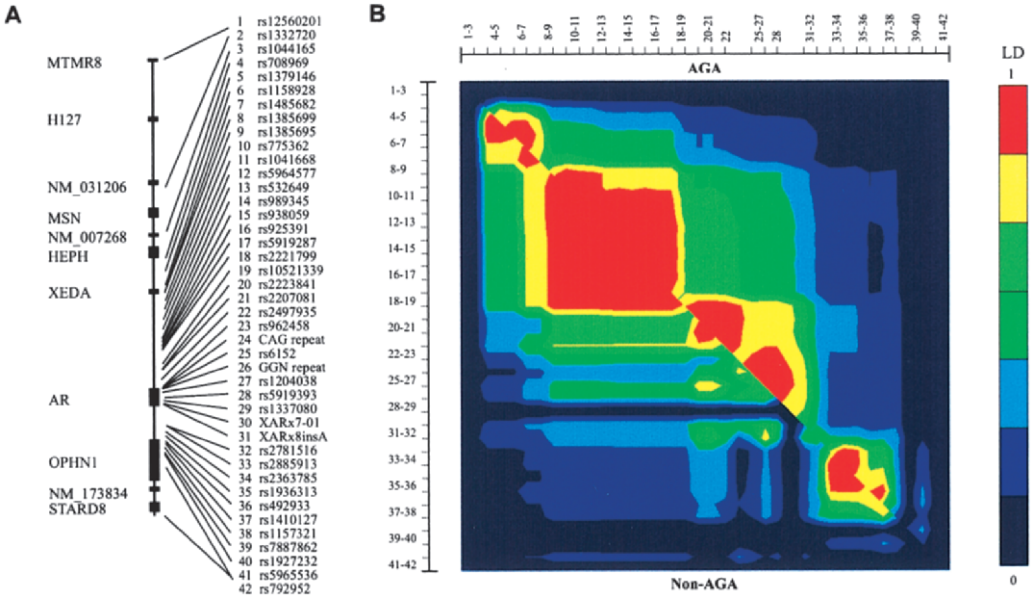

Figure 2 Gene and LD structure of the *AR* locus. *A*, Distribution of known genes (*left*) and typed SNPs and STRs (*right*) at the *AR* locus. The shown genomic region spans 4.4 Mb. Gene content information is based on Ensembl. *B*, LD in 198 individuals with AGA (*upper right diagonal*) and in 157 individuals without AGA (*lower left diagonal*) was measured with χ^2 and was visualized using the GOLD program (Abecasis and Cookson 2000).

Table 3

Association of AGA with Markers at the *AR* Locus

SNP and Allele	Frequency			TDT[a]			Affected/Control Analysis			Affected/Unaffected Analysis			Position[b]
	Affected	Control	Unaffected	Transmitted	Nontransmitted	*P*[c]	*P*	OR[d]	CI	*P*	OR[d]	CI	
rs12560201:						6.68×10^{-1}	8.99×10^{-1}	1.04	.55–1.82	9.38×10^{-1}	.97	.5–1.87	63,335,397
G	.112	.108	.115	23	27								
A	.888	.892	.885										
rs1332720:						1.00	1.00	ND	ND	1.00	ND	ND	64,598,557
C	.995	1.000	1.000	0	1								
T	.005	.000	.000										
rs1044165:						1.00	9.50×10^{-1}	.98	.55–1.89	1.58×10^{-1}	1.53	.85–2.76	65,024,747
A	.124	.122	.178	22	23								
G	.876	.878	.822										
rs708969:						5.35×10^{-1}	1.21×10^{-2}	2.02	1.16–3.53	3.90×10^{-3}	2.27	1.29–4	65,252,916
A	.875	.776	.755	36	29								
T	.125	.224	.245										
rs1379146:						3.25×10^{-1}	2.68×10^{-4}	2.61	1.55–4.45	4.72×10^{-4}	2.60	1.51–4.5	65,389,382
T	.872	.723	.724	39	29								
A	.128	.277	.276										
rs1158928:						1.30×10^{-1}	1.04×10^{-5}	3.23	1.89–5.56	2.08×10^{-5}	3.22	1.85–5.61	65,532,217
A	.883	.700	.701	42	27								
G	.117	.300	.299										
rs1485682:						1.30×10^{-1}	1.97×10^{-5}	3.13	1.82–5.35	3.08×10^{-5}	3.16	1.81–5.48	65,598,612
C	.883	.707	.705	42	27								
T	.117	.293	.295										
rs1385699:						1.22×10^{-1}	1.60×10^{-5}	3.23	1.86–5.59	9.41×10^{-5}	3.00	1.71–5.32	65,608,007
A	.887	.708	.722	44	28								
G	.113	.292	.278										
rs1385695:						2.97×10^{-3}	1.03×10^{-5}	4.68	2.24–9.78	2.40×10^{-8}	6.57	3.17–13.63	65,687,533
A	.949	.799	.739	38	13								
G	.051	.201	.261										
rs775362:						4.10×10^{-3}	2.09×10^{-5}	4.39	2.13–9.21	3.73×10^{-8}	6.40	3.11–13.39	65,762,194
G	.949	.809	.744	37	13								
A	.051	.191	.256										
rs1041668:						1.49×10^{-4}	1.80×10^{-6}	6.42	2.74–15.03	1.58×10^{-9}	9.29	4.03–21.44	65,868,977
G	.036	.194	.258	9	39								
A	.964	.806	.742										
rs5964577:						2.50×10^{-3}	4.24×10^{-6}	4.99	2.32–10.68	1.20×10^{-8}	7.02	3.29–15.08	65,954,512
A	.954	.806	.747	36	12								
T	.046	.194	.253										
rs532649:						5.96×10^{-6}	5.42×10^{-7}	6.38	2.75–14.71	2.33×10^{-10}	9.62	4.17–22.13	66,050,269
A	.960	.807	.735	37	7								
G	.036	.193	.265										
rs989345:						4.65×10^{-6}	1.60×10^{-6}	6.24	2.74–14.66	6.13×10^{-10}	9.56	4.2–22.29	66,084,055
A	.964	.811	.737	36	6								
G	.036	.189	.263										
rs938059:						4.03×10^{-6}	1.10×10^{-6}	6.55	2.82–15.25	9.20×10^{-10}	9.63	4.18–22.17	66,119,748
C	.036	.195	.263	7	40								
A	.964	.805	.737										
rs925391:						4.76×10^{-6}	5.97×10^{-7}	7.28	3.03–18.03	1.98×10^{-10}	11.15	4.65–27.42	66,123,458
G	.969	.811	.737	38	6								
A	.031	.189	.263										
rs5919287:						4.62×10^{-6}	2.57×10^{-7}	6.61	2.87–12.82	1.76×10^{-10}	9.70	4.22–22.45	66,143,635
C	.964	.802	.734	39	7								
T	.036	.198	.266										
rs2221799:						4.62×10^{-6}	3.37×10^{-7}	6.53	2.84–15.18	1.39×10^{-9}	9.21	4.04–21.47	66,159,637
A	.964	.804	.744	39	7								
G	.036	.196	.256										
rs10521339:						4.00×10^{-4}	1.95×10^{-7}	5.65	2.65–12.04	2.10×10^{-12}	9.89	4.62–20.7	66,291,407
A	.954	.778	.677	11	41								
T	.046	.212	.323										
rs2223841:						9.40×10^{-6}	1.20×10^{-6}	5.59	2.62–11.94	1.72×10^{-11}	9.48	4.48–20.05	66,353,192
A	.954	.789	.688	44	11								
G	.046	.211	.312										
rs2207081:						4.33×10^{-6}	7.00×10^{-7}	6.21	2.8–13.78	5.04×10^{-12}	10.72	4.89–23.75	66,356,719
A	.959	.792	.688	44	10								
G	.041	.208	.312										
rs2497935:						1.79×10^{-5}	3.10×10^{-6}	5.00	2.41–10.36	3.17×10^{-11}	8.69	4.23–17.87	66,447,287
A	.949	.790	.684	45	12								
G	.041	.208	.316										
rs962458:						2.89×10^{-3}	2.87×10^{-4}	7.32	2.12–25.3	2.56×10^{-6}	11.10	3.27–37.74	66,528,985
G	.015	.102	.146	2	17								
A	.985	.898	.854										
CAG[e]:						2.71×10^{-1}[f]							66,548,181
18	.042	.085	.110	13	12		NS	ND	ND	NS	ND	ND	
19	.110	.106	.110	30	39		NS	ND	ND	NS	ND	ND	
20	.194	.112	.103	50	19		2.66×10^{-2}	1.91	1.07–3.41	2.20×10^{-2}	2.10	1.10–3.99	

(*continued*)

Table 3 (continued)

SNP and Allele	Frequency			TDT[a]			Affected/Control Analysis			Affected/Unaffected Analysis			Position[b]
	Affected	Control	Unaffected	Transmitted	Nontransmitted	*P*[c]	*P*	OR[d]	CI	*P*	OR[d]	CI	
21	.147	.191	.205	32	31		NS	ND	ND	NS	ND	ND	
22	.136	.122	.096	32	28		NS	ND	ND	NS	ND	ND	
23	.126	.090	.068	26	30		NS	ND	ND	NS	ND	ND	
24	.084	.080	.096	19	36		NS	ND	ND	NS	ND	ND	
25	.047	.064	.075	12	14		NS	ND	ND	NS	ND	ND	
rs6152:						5.88×10^{-5}	3.80×10^{-6}	5.26	2.45–11.3	6.66×10^{-10}	8.21	3.86–17.45	66,548,648
T	.046	.201	.282	11	40								
C	.954	.799	.718										
GGN[e]:						7.79×10^{-5} [f]							66,549,360
23	.651	.495	.421	76	43		2.06×10^{-3}	1.91	1.26–.2.88	2.07×10^{-5}	2.57	1.656–3.98	
24	.182	.367	.461	21	68		5.39×10^{-5}	.38	.24–.62	2.62×10^{-8}	.26	.161–.42	
rs1204038:						9.79×10^{-5}	1.90×10^{-6}	5.45	2.55–11.67	2.83×10^{-9}	7.72	3.62–16.46	66,571,246
A	.046	.208	.271	11	39								
G	.954	.792	.729										
rs5919393:						1.58×10^{-5}	6.00×10^{-7}	5.82	2.73–12.41	1.69×10^{-8}	7.88	3.7–16.77	66,608,378
C	.046	.218	.274	11	38								
T	.954	.782	.726										
rs1337080:						5.73×10^{-3}	6.71×10^{-4}	6.73	1.92–23.14	4.75×10^{-6}	10.78	3.12–36.19	66,661,940
A	.985	.907	.859	15	2								
G	.015	.093	.141										
XARx7_001[g]:						1.00	6.79×10^{-2}	3.95	.81–19.25	1.45×10^{-1}	3.21	.61–16.76	66,725,646
A	.010	.039	.032	3	3								
G	.990	.961	.968										
XARx8insA[h]:						3.29×10^{-3}	5.09×10^{-5}	4.11	1.99–8.47	6.62×10^{-8}	5.94	2.94–12.02	66,727,141
delA	.944	.804	.739	34	13								
A	.056	.196	.261										
rs2781516:						3.28×10^{-2}	1.00×10^{-2}	2.02	1.18–3.48	2.86×10^{-4}	2.69	1.56–4.64	66,885,052
A	.128	.228	.282	21	38								
G	.872	.772	.718										
rs2885913:						4.46×10^{-2}	1.49×10^{-4}	2.31	1.49–3.57	4.29×10^{-5}	2.53	1.62–3.97	66,898,824
G	.255	.441	.465	40	64								
A	.745	.559	.535										
rs2363785:						1.73×10^{-2}	1.82×10^{-4}	2.28	1.47–3.53	2.04×10^{-4}	2.34	1.49–3.68	66,954,319
G	.745	.562	.555	65	37								
T	.255	.438	.445										
rs1936313:						2.48×10^{-2}	7.91×10^{-5}	2.37	1.54–3.64	9.15×10^{-4}	2.15	1.36–3.37	66,983,669
T	.744	.551	.575	65	39								
C	.256	.449	.425										
rs492933:						2.44×10^{-2}	1.15×10^{-2}	1.76	1.13–2.72	2.39×10^{-2}	1.69	1.07–2.65	67,046,865
C	.736	.614	.623	67	40								
T	.264	.386	.377										
rs1410127:						2.83×10^{-2}	8.02×10^{-3}	1.79	1.16–2.76	2.13×10^{-2}	1.71	1.08–2.67	67,063,402
C	.735	.608	.619	66	40								
T	.265	.392	.381										
rs1157321:						7.83×10^{-4}	1.59×10^{-5}	2.89	1.75–4.7	2.73×10^{-5}	2.93	1.75–4.87	67,104,762
G	.848	.659	.656	64	27								
T	.152	.341	.344										
rs7887862:						1.09×10^{-1}	3.31×10^{-1}	1.34	.74–2.41	1.28×10^{-1}	1.65	.86–3.2	67,148,136
C	.152	.118	.098	36	21								
T	.848	.882	.902										
rs1927232:						6.77×10^{-1}	1.77×10^{-1}	1.36	.87–2.11	8.23×10^{-1}	1.06	.66–1.69	67,165,912
C	.719	.653	.708	49	44								
T	.281	.347	.292										
rs5965536:						1.56×10^{-1}	3.57×10^{-1}	1.33	.73–2.41	9.50×10^{-1}	1.02	.57–1.84	67,300,796
C	.154	.121	.151	35	24								
G	.846	.879	.849										
rs792952:						1.39×10^{-1}	8.45×10^{-1}	1.05	.65–1.69	8.56×10^{-1}	1.05	.64–1.71	67,766,144
C	.249	.240	.240	53	36								
G	.751	.760	.760										

Note.—ND = not determined; NS = not significant.

[a] Family transmission/disequilibrium test (Knapp and Becker 2003).

[b] UCSC Genome Bioinformatics Plus Strand Build 35.

[c] Permutation-based value for association.

[d] Odds ratio.

[e] Only alleles with frequencies ≥0.05 in controls.

[f] Global *P* value.

[g] GCTTTGTCTAATGCTCCTTC[G/A]TGGGCATGCTTCCCCTCCCC (intron 6 of *AR*).

[h] ACAAGCAAACAAAAAAAAAA[A/delA]GCAAAAACAAAACAAAAAAT (exon 8 of *AR*).

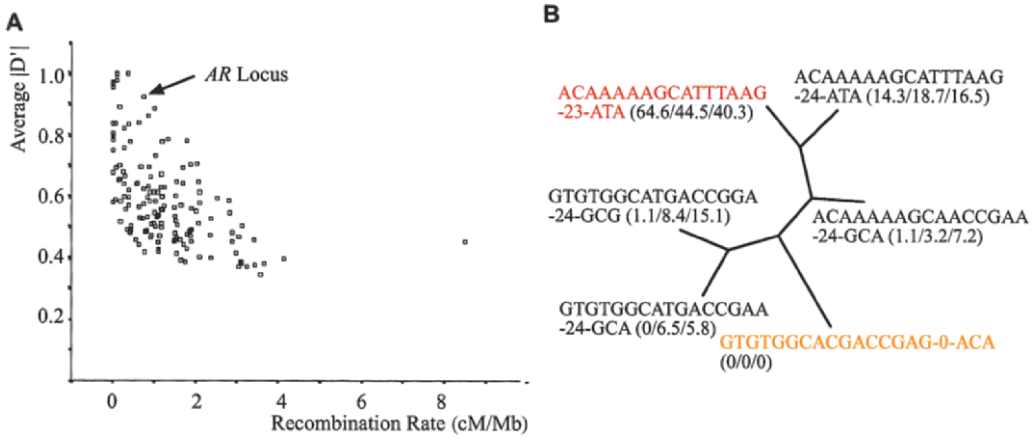

Figure 3 The *AR* locus: LD relative to the X-chromosome and haplotype structure. *A,* Average pairwise LD (measured by $|D'|$) between SNPs retrieved from the HapMap Project database, plotted over X-chromosomal recombination rates in 1-cM–sized windows. An average LD higher than that at the *AR* locus was displayed by only the six windows covering the centromere and showed distinctively smaller recombination rates. *B,* Neighbor-joining tree of frequent haplotype sequences (>5% in the samples from the individuals with AGA, controls, and individuals without AGA) within the most strongly associated haplotype block (genetic markers *rs1385695–XARx8insA*). The CAG repeat, *XARx7_01,* and *XARx8insA* were omitted from tree construction, because the corresponding *Pan troglodytes* allele could not be retrieved from the chimpanzee genomic sequence. The AGA haplotype is shown in red, and the chimpanzee haplotype is shown in orange. The AGA haplotype shows low sequence identity to the ancestral (chimpanzee) haplotype. Haplotype frequencies of the respective samples are indicated in parentheses (affected/control/unaffected [in %]). GGN-23 is indicated as "23"; GGN-24 is indicated as "24."

unaffected global *P* value of .0001). The study by Ellis et al. (2001) also showed a larger effect for the GGN repeat than the CAG repeat. However, the pooling of alleles in their study renders an exact comparison of results difficult. The GGN allele of 23 repeats showed a difference of affected versus unaffected allele frequencies of 0.23, which was in the range of the strongest-associated SNPs (*rs1385695–XARx8insA* [table 3]) but with a lower frequency in controls. This suggests that GGN-23 either is closer to the AGA mutation or is itself the AGA-susceptibility allele. Previously obtained functional data, in which shorter repeat alleles of the GGN repeat were associated with higher protein levels and thereby higher *AR* activity (Ding et al. 2005), support the possibility of a causal role for the repeat, and this would be compatible with current understanding of the involvement of androgens in AGA. Several studies have also suggested an effect of CAG repeat lengths on AR transactivating activity (Mhatre et al. 1993; Chamberlain et al. 1994; Kazemi-Esfarjani et al. 1995; Choong et al. 1996; Nakajima et al. 1996; Beilin et al. 2000; Ding et al. 2004). Since both repeats modulate AR activity and since we observe an association only between the GGN repeat and AGA, it may be that cells of the hair follicle lack cofactors that interact with the CAG-encoded domain.

Interestingly, shorter alleles for the GGN repeat have also been associated with prostate cancer (Hsing et al. 2000; Chang et al. 2002), whereas longer alleles have been associated with endometrial cancer (Sasaki et al. 2005), which would be in accordance with the differing effects of androgens on the endometrium and the prostate (androgens exert an inhibitory effect on endometrial cell proliferation, whereas they have a mitogenic effect in the prostate). However, the association findings with prostate cancer remain controversial, and no effect was shown in a large meta-analysis (Zeegers et al. 2004). The association with endometrial cancer has yet to be confirmed.

It remains possible that an as-yet-undetected variant in either a regulatory region affecting the expression level or an intronic variant affecting the splicing pattern of *AR* might be responsible for AGA susceptibility. The latter seems unlikely since we did not detect alternatively spliced transcripts of *AR* in human hair follicles of seven individuals representing different haplotypes. Previous studies have identified *AR* regulatory elements up to position −737 of the *AR* transcription start site (Faber et al. 1991, 1993; Supakar et al. 1993), as well as exonic enhancers in exon 1 (Faber et al. 1993) and exons 4 and 5 (Grad et al. 2001). Our sequencing analysis of 12 individuals with the associated haplotype revealed no variability in these regulatory elements of *AR*. However, there may be additional regions with regulatory effect that have not yet been fully characterized (Lower et al. 2004). Haplotypes carrying the GGN-24 allele show clearly higher frequencies in individuals without AGA than in those with AGA (fig. 3*B*). Since this effect is strikingly weaker in the ACAAAAAGCATTTAAG-24-ATA haplotype than in the other GGN-24–carrying hap-

lotypes (fig. 3*B*), it is likely that further functionally relevant variability exists that modifies the protective effect of GGN-24–bearing haplotypes.

It is interesting to note that genetic variation in *AR,* which is located on the X chromosome, cannot explain the resemblance of fathers and sons with respect to the development of AGA (Küster and Happle 1984; Ellis et al. 1998), since sons always inherit the X chromosome from their mothers. The fact that family studies of AGA have typically stressed the resemblance of fathers and sons is understandable, given the differences in patterns of hair loss between males and females. Our genetic data, however, stress the relative importance of the maternal line in the inheritance of AGA, since we estimate an etiological fraction of 0.46 that can be attributed to having ⩽23 GGN repeats within *AR*. This suggests that the average phenotypic resemblance should be greater between affected males and their maternal grandfathers than between affected males and their fathers. It is likely that the remaining etiological fraction is due to genetic variation at autosomal loci, which could explain the similarity of the AGA pattern of fathers and sons. Some autosomal candidate genes have been investigated in the past, including the insulin gene (Ellis et al. 1999), the 5α-reductase genes (Ellis et al. 1998), and the *hairless* gene (Hillmer et al. 2001, 2002), but none of these has been associated with AGA. A systematic linkage-based approach should enable the identification of additional loci.

Acknowledgments

We thank all participants for consenting to the study and for providing blood samples. This study was supported by grants from the Deutsche Forschungsgemeinschaft (DFG) (Forschergruppen "Keratinozyten—Proliferation und differenzierte Leistung in der Epidermis" and "Genetische Epidemiologie und Medizinische Genetik komplexer Erkrankungen"), the Bundesministerium für Bildung und Forschung (German Human Genome Project, "Pharmacogenomics"), and the Alfried Krupp von Bohlen und Halbach-Stiftung. R.C.B. is a recipient of a DFG Emmy Noether fellowship. We thank Faten Dahdouh and Carola Müller, for assistance in STR genotyping, and Dr. Christine Schmäl, for help in preparing the manuscript.

Web Resources

The URLs for data presented herein are as follows:

Ensembl, http://www.ensembl.org/ (for *AR* locus information)
HapMap, http://www.hapmap.org/ (for pairwise LD on the X chromosome [genotypes queried in November 2004])
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for AGA and *AR*)
University of California–Santa Cruz (UCSC) Genome Bioinformatics, http://genome.ucsc.edu/ (for X-chromosomal recombination rates and definition of the reference strand for SNP allele calling)

References

Abecasis GR, Cookson WO (2000) GOLD—graphical overview of linkage disequilibrium. Bioinformatics 16:182–183
Bamshad M, Wooding SP (2003) Signatures of natural selection in the human genome. Nat Rev Genet 4:99–111
Becker T, Knapp M (2004*a*) A powerful strategy to account for multiple testing in the context of haplotype analysis. Am J Hum Genet 75:561–570
——— (2004*b*) Maximum-likelihood estimation of haplotype frequencies in nuclear families. Genet Epidemiol 27:21–32
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
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
Chang BL, Zheng SL, Hawkins GA, Isaacs SD, Wiley KE, Turner A, Carpten JD, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB, Xu J (2002) Polymorphic GGC repeats in the androgen receptor gene are associated with hereditary and sporadic prostate cancer risk. Hum Genet 110:122–129
Choong CS, Kemppainen 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
——— (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 5α-reductase genes. J Invest Dermatol 110:849–853
——— (1999) Insulin gene polymorphism and premature male pattern baldness in the general population. Clin Sci (Lond) 96:659–662
——— (2001) Polymorphism of the androgen receptor gene is associated with male pattern baldness. J Invest Dermatol 116:452–455
Faber PW, van Rooij HC, Schipper HJ, Brinkmann AO, Trapman J (1993) Two different, overlapping pathways of transcription initiation are active on the TATA-less human androgen receptor promoter: the role of Sp1. J Biol Chem 268: 9296–9301
Faber PW, van Rooij HC, van der Korput HA, Baarends WM, Brinkmann AO, Grootegoed JA, Trapman J (1991) Characterization of the human androgen receptor transcription unit. J Biol Chem 266:10743–10749
Grad JM, Lyons LS, Robins DM, Burnstein KL (2001) The androgen receptor (AR) amino-terminus imposes androgen-specific regulation of AR gene expression via an exonic enhancer. Endocrinology 142:1107–1116
Gudbjartsson DF, Jonasson K, Frigge M, Kong A (2000) Al-

legro, a new computer program for multipoint linkage analysis. Nat Genet 25:12–13

Hamilton JB (1951) Patterned loss of hair in man: types and incidence. Ann N Y Acad Sci 53:708–728

Hawk E, Breslow RA, Graubard BI (2000) Male pattern baldness and clinical prostate cancer in the epidemiologic follow-up of the first National Health and Nutrition Examination Survey. Cancer Epidemiol Biomarkers Prev 9:523–527

Hillmer AM, Kruse R, Betz RC, Schumacher J, Heyn U, Propping P, Nöthen MM, Cichon S (2001) Variant 1859G→A (Arg620Gln) of the *hairless* gene: absence of association with papular atrichia or androgenetic alopecia. Am J Hum Genet 69:235–237

Hillmer AM, Kruse R, Macciardi F, Heyn U, Betz RC, Ruzicka T, Propping P, Nöthen MM, Cichon S (2002) The hairless gene in androgenetic alopecia: results of a systematic mutation screening and family-based association approach. Br J Dermatol 146:601–608

Hsing AW, Gao YT, Wu G, Wang X, Deng J, Chen YL, Sesterhenn IA, Mostofi FK, Benichou J, Chang C (2000) Polymorphic CAG and GGN repeat lengths in the androgen receptor gene and prostate cancer risk: a population-based case-control study in China. Cancer Res 60:5111–5116

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

Knapp M, Becker T (2003) Family-based association analysis with tightly linked markers. Hum Hered 56:2–9

Küster W, Happle R (1984) The inheritance of common baldness: two B or not two B? J Am Acad Dermatol 11:921–926

La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 352:77–79

Lee HJ, Chang C (2003) Recent advances in androgen receptor action. Cell Mol Life Sci 60:1613–1622

Lotufo PA, Chae CU, Ajani UA, Hennekens CH, Manson JE (2000) Male pattern baldness and coronary heart disease: the Physicians' Health Study. Arch Intern Med 160:165–171

Lower KM, Kumar R, Woollatt E, Villard L, Gecz J, Sutherland GR, Callen DF (2004) Partial androgen insensitivity syndrome and t(X;5): are there upstream regulatory elements of the androgen receptor gene? Horm Res 62:208–214

Matilainen V, Koskela P, Keinanen-Kiukaanniemi S (2000) Early androgenetic alopecia as a marker of insulin resistance. Lancet 356:1165–1166

Mhatre AN, Trifiro MA, Kaufman M, Kazemi-Esfarjani P, Figlewicz D, Rouleau G, Pinsky L (1993) Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. Nat Genet 5: 184–188

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215

Nagaraja R, MacMillan S, Kere J, Jones C, Griffin S, Schmatz M, Terrell J, Shomaker M, Jermak C, Hott C, Masisi M, Mumm S, Srivastava A, Pilia G, Featherstone T, Mazzarella R, Kesterson S, McCauley B, Railey B, Burough F, Nowotny V, D'Urso M, States D, Brownstein B, Schlessinger D (1997) X chromosome map at 75-kb STS resolution, revealing extremes of recombination and GC content. Genome Res 7: 210–222

Nakajima H, Kimura F, Nakagawa T, Furutama D, Shinoda K, Shimizu A, Ohsawa N (1996) Transcriptional activation by the androgen receptor in X-linked spinal and bulbar muscular atrophy. J Neurol Sci 142:12–16

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

Oh BR, Kim SJ, Moon JD, Kim HN, Kwon DD, Won YH, Ryu SB, Park YI (1998) Association of benign prostatic hyperplasia with male pattern baldness. Urology 51:744–748

Sasaki M, Karube A, Karube Y, Watari M, Sakuragi N, Fujimoto S, Dahiya R (2005) GGC and *Stu*I polymorphism on the androgen receptor gene in endometrial cancer patients. Biochem Biophys Res Commun 329:100–104

Supakar PC, Song CS, Jung MH, Slomczynska MA, Kim JM, Vellanoweth RL, Chatterjee B, Roy AK (1993) A novel regulatory element associated with age-dependent expression of the rat androgen receptor gene. J Biol Chem 268:26400–26408

Zeegers MP, Kiemeney LA, Nieder AM, Ostrer H (2004) How strong is the association between CAG and GGN repeat length polymorphisms in the androgen receptor gene and prostate cancer risk? Cancer Epidemiol Biomarkers Prev 13: 1765–1771

Zhao H, Zhang S, Merikangas KR, Trixler M, Wildenauer DB, Sun F, Kidd KK (2000) Transmission/disequilibrium tests using multiple tightly linked markers. Am J Hum Genet 67:936–946