A genome-wide association scan identifies new susceptibility variants for male pattern baldness on chromosome 20p11

Axel M. Hillmer[1,9], Felix F. Brockschmidt[1,9], Sandra Hanneken[2], Sibylle Eigelshoven[2], Michael Steffens[3], Antonia Flaquer[3], Stefan Herms[1], Tim Becker[3], Anne-Katrin Kortüm[2], Dale R. Nyholt[4], Zhen Zhen Zhao[4], Grant W. Montgomery[4], Nicholas G. Martin[4], Thomas W. Mühleisen[1], Margrieta A. Alblas[1], Susanne Moebus[5], Karl-Heinz Jöckel[5], Martina Bröcker-Preuss,[6] Raimund Erbel[7], Roman Reinartz[1], Regina C. Betz[8], Sven Cichon[1,8], Peter Propping[8], Max P. Baur[3], Thomas F. Wienker[3], Roland Kruse[2], Markus M. Nöthen[1,8]

[1]Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany

[2]Department of Dermatology, University of Düsseldorf, Düsseldorf, Germany

[3]Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany

[4]Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia

[5]Institute for Medical Informatics, Biometry and Epidemiology, University Duisburg-Essen, Germany

[6]Clinic of Endocrinology, Central Laboratory Unit Research and Education, University Hospital of Essen, University Duisburg-Essen

[7]Clinic of Cardiology, West-German Heart Center Essen, University Duisburg-Essen, Germany

[8]Institute of Human Genetics, University of Bonn, Bonn, Germany

[9]These authors contributed equally to this work

Supplementary Table 1 Genome-wide association results for AGA (excluding *AR* locus)

SNP	Chr.	Position [bp][a]	MAF cases	MAF controls	*P*	OR (95% CI)
rs1998076	20	21,828,045	0.282 (A)	0.427 (A)	1.30×10^{-7}	1.90 (1.50 - 2.41)
rs6075852	20	21,814,151	0.292 (A)	0.429 (A)	3.50×10^{-7}	1.83 (1.45 - 2.31)
rs2180439	20	21,801,100	0.292 (C)	0.429 (C)	3.85×10^{-7}	1.82 (1.45 - 2.30)
rs6137444	20	21,733,639	0.264 (C)	0.383 (C)	3.11×10^{-6}	1.74 (1.37 - 2.21)
rs201571	20	21,961,514	0.289 (C)	0.411 (C)	4.31×10^{-6}	1.72 (1.36 - 2.17)
rs10992241	9	93,931,599	0.156 (C)	0.262 (C)	5.31×10^{-6}	1.92 (1.44 - 2.54)
rs6137473	20	21,832,693	0.321 (G)	0.445 (G)	5.58×10^{-6}	1.70 (1.35 - 2.13)
rs6047768	20	21,901,649	0.331 (G)	0.457 (G)	6.09×10^{-6}	1.70 (1.35 - 2.13)
rs2207878	20	21,899,847	0.331 (G)	0.457 (G)	6.19×10^{-6}	1.70 (1.35 - 2.13)
rs6035995	20	21,879,683	0.331 (G)	0.455 (G)	6.97×10^{-6}	1.69 (1.35 - 2.13)
rs6113424	20	21,854,780	0.331 (G)	0.455 (G)	7.05×10^{-6}	1.69 (1.35 - 2.12)
rs2024885	20	21,865,393	0.331 (A)	0.455 (A)	7.05×10^{-6}	1.69 (1.35 - 2.12)
rs201543	20	21,894,928	0.331 (A)	0.455 (A)	7.05×10^{-6}	1.69 (1.35 - 2.12)
rs17817969	18	70,473,226	0.223 (T)	0.117 (T)	7.16×10^{-6}	2.17 (1.51 - 3.11)
rs6047769	20	21,906,049	0.333 (A)	0.457 (A)	8.47×10^{-6}	1.68 (1.34 - 2.12)
rs1884592	20	21,838,690	0.331 (C)	0.454 (C)	8.58×10^{-6}	1.68 (1.34 - 2.11)
rs6137476	20	21,845,656	0.331 (A)	0.454 (A)	8.58×10^{-6}	1.68 (1.34 - 2.11)
rs6047731	20	21,847,141	0.331 (G)	0.454 (G)	8.58×10^{-6}	1.68 (1.34 - 2.11)
rs6113491	20	22,005,415	0.359 (C)	0.483 (C)	8.63×10^{-6}	1.66 (1.33 - 2.08)
rs1555264	20	21,909,410	0.332 (G)	0.455 (G)	8.64×10^{-6}	1.68 (1.34 - 2.11)
rs4658627	1	242,577,799	0.172 (A)	0.308 (A)	9.60×10^{-6}	2.14 (1.51 - 3.02)
rs11975187	7	153,863,909	0.490 (A)	0.359 (G)	1.02×10^{-5}	1.86 (1.41 - 2.46)
rs4805229	19	33,692,043	0.380 (T)	0.268 (T)	1.34×10^{-5}	1.67 (1.32 - 2.12)
rs9300398	13	96,690,316	0.363 (T)	0.252 (T)	1.46×10^{-5}	1.69 (1.33 - 2.15)
rs4771987	13	96,699,391	0.429 (G)	0.315 (G)	1.88×10^{-5}	1.63 (1.30 - 2.05)
rs3786343	18	59,798,875	0.458 (T)	0.340 (T)	2.27×10^{-5}	1.64 (1.31 - 2.05)
rs6986303	8	134,547,711	0.309 (A)	0.205 (A)	2.33×10^{-5}	1.74 (1.35 - 2.24)
rs9856948	3	189,240,095	0.419 (C)	0.304 (C)	2.51×10^{-5}	1.65 (1.31 - 2.08)
rs1139488	9	115,193,721	0.306 (C)	0.418 (C)	2.77×10^{-5}	1.63 (1.29 - 2.05)
rs13079866	3	196,641,541	0.061 (G)	0.131 (G)	2.84×10^{-5}	2.33 (1.56 - 3.49)
rs7319347	13	66,977,210	0.072 (C)	0.173 (C)	3.00×10^{-5}	2.70 (1.66 - 4.38)
rs17818946	18	70,519,734	0.170 (G)	0.082 (G)	3.06×10^{-5}	2.30 (1.52 - 3.47)
rs6804475	3	113,152,321	0.246 (A)	0.352 (A)	3.22×10^{-5}	1.66 (1.30 - 2.12)
rs500629	11	113,550,770	0.346 (G)	0.241 (G)	3.62×10^{-5}	1.66 (1.30 - 2.12)
rs10495747	2	24,368,500	0.149 (C)	0.075 (C)	3.78×10^{-5}	2.16 (1.50 - 3.10)
rs1464766	5	111,641,241	0.243 (A)	0.372 (A)	4.14×10^{-5}	1.84 (1.35 - 2.51)
rs2551182	2	24,293,921	0.128 (T)	0.061 (T)	4.75×10^{-5}	2.29 (1.54 - 3.39)
rs300639	9	32,776,553	0.067 (T)	0.021 (T)	4.83×10^{-5}	3.45 (1.85 - 6.42)
rs10001582	4	124,578,020	0.364 (A)	0.478 (A)	4.91×10^{-5}	1.60 (1.28 - 2.01)
rs11584662	1	184,847,447	0.417 (T)	0.467 (G)	4.92×10^{-5}	1.60 (1.28 - 1.99)
rs1009070	11	29,484,237	0.106 (G)	0.038 (G)	5.20×10^{-5}	3.05 (1.78 - 5.24)
rs880013	9	1,425,496	0.474 (T)	0.415 (C)	5.28×10^{-5}	1.56 (1.25 - 1.95)
rs4896028	6	134,580,872	0.309 (G)	0.209 (G)	5.35×10^{-5}	1.69 (1.32 - 2.18)
rs11725633	4	37,868,359	0.368 (T)	0.483 (T)	5.57×10^{-5}	1.60 (1.28 - 2.00)
rs11655206	17	72,501,171	0.497 (T)	0.361 (T)	5.60×10^{-5}	1.74 (1.32 - 2.30)
rs6550859	3	24,385,419	0.390 (G)	0.468 (A)	5.72×10^{-5}	1.77 (1.34 - 2.34)
rs3786939	19	44,888,022	0.301 (A)	0.184 (A)	5.79×10^{-5}	1.91 (1.39 - 2.61)

[a] NCBI build 36

Supplementary Table 2 Replication analysis in 319 German cases and 234 German controls

SNP	Chr.	Position [bp][a]	MAF cases	MAF controls	*P*	OR (95% CI)
rs6113491	20	22,005,415	0.364 (C)	0.447 (A)	8.13×10^{-10}	2.17 (1.7 - 2.77)
rs2180439	20	21,801,100	0.303 (C)	0.485 (C)	1.37×10^{-09}	2.17 (1.7 - 2.78)
rs1998076	20	21,828,045	0.301 (A)	0.479 (A)	3.69×10^{-09}	2.13 (1.66 - 2.73)
rs201571	20	21,961,514	0.313 (C)	0.483 (C)	2.21×10^{-08}	2.05 (1.6 - 2.62)
rs6137444	20	21,733,639	0.277 (C)	0.404 (C)	1.57×10^{-05}	1.76 (1.37 - 2.27)
rs10992241	9	91,971,333	0.289 (C)	0.201 (C)	8.68×10^{-04}	1.62 (1.22 - 2.15)
rs11725633	4	38,014,530	0.422 (T)	0.479 (T)	0.058	1.26 (0.99 - 1.6)
rs4771987	13	96,699,391	0.417 (G)	0.363 (G)	0.075	1.25 (0.98 - 1.6)
rs6804475	3	113,152,321	0.292 (A)	0.340 (A)	0.081	1.25 (0.97 - 1.62)
rs3786939	19	44,888,022	0.202 (A)	0.162 (A)	0.085	1.31 (0.96 - 1.79)
rs880013	9	1,425,496	0.431 (C)	0.479 (C)	0.123	1.21 (0.95 - 1.54)
rs500629	11	113,550,770	0.307 (G)	0.276 (G)	0.249	1.17 (0.9 - 1.52)
rs9300398	13	96,690,316	0.335 (T)	0.303 (T)	0.252	1.16 (0.9 - 1.5)
rs1009070	11	29,484,237	0.062 (G)	0.076 (G)	0.35	1.25 (0.78 - 2)
rs4896028	6	134,580,872	0.268 (G)	0.244 (G)	0.368	1.14 (0.86 - 1.5)
rs10495747	2	24,426,647	0.130 (C)	0.115 (C)	0.458	1.15 (0.8 - 1.65)
rs17817969	18	70,473,226	0.147 (T)	0.132 (T)	0.483	1.13 (0.8 - 1.6)
rs6986303	8	134,547,711	0.288 (A)	0.269 (A)	0.495	1.1 (0.84 - 1.43)
rs10001582	4	124,716,175	0.429 (A)	0.409 (A)	0.51	1.09 (0.85 - 1.38)
rs3786343	18	59,798,875	0.395 (T)	0.376 (T)	0.527	1.08 (0.85 - 1.38)
rs1464766	5	111,641,241	0.324 (A)	0.342 (A)	0.545	1.08 (0.84 - 1.39)
rs11975187	7	153,670,624	0.346 (G)	0.363 (G)	0.559	1.08 (0.84 - 1.38)
rs11584662	1	183,312,481	0.436 (T)	0.453 (T)	0.573	1.07 (0.84 - 1.36)
rs17818946	18	70,519,734	0.108 (G)	0.098 (G)	0.588	1.11 (0.75 - 1.65)
rs6550859	3	24,385,419	0.498 (A)	0.483 (A)	0.608	1.06 (0.84 - 1.35)
rs9856948	3	189,240,103	0.350 (C)	0.340 (C)	0.735	1.04 (0.81 - 1.34)
rs2551182	2	24,352,068	0.099 (T)	0.096 (T)	0.886	1.03 (0.69 - 1.54)
rs7319347	13	66,977,210	0.160 (C)	0.162 (C)	0.91	1.02 (0.74 - 1.41)
rs4658627	1	240,837,217	0.270 (A)	0.274 (A)	0.911	1.02 (0.78 - 1.33)
rs300639	9	32,776,553	0.023 (T)	0.023 (T)	1	1 (0.44 - 2.27)

[a] NCBI build 36

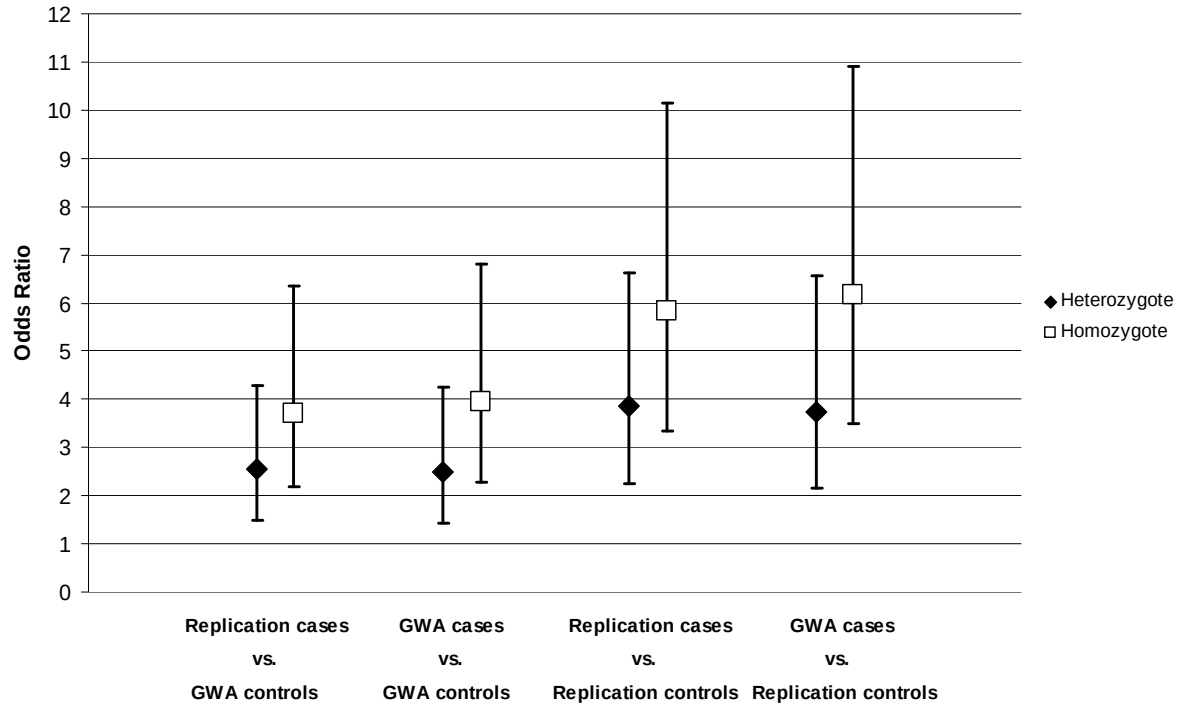
Supplementary Table 3 Replication analysis in Australian AGA sample

		141 cases[a] vs. 150 controls				60 cases[b] vs. 150 controls			
SNP	Minor allele	MAF cases	MAF controls	*P*	OR (95% CI)	MAF cases	MAF controls	*P*	OR (95% CI)
rs6137444	C	0.298	0.343	0.247	1.23 (0.86-1.74)	0.275	0.343	0.171	1.37 (0.86-2.19)
rs2180439	C	0.340	0.420	0.041	1.4 (1-1.96)	0.317	0.420	0.045	1.56 (1-2.44)
rs1998076	A	0.344	0.417	0.064	1.36 (0.97-1.91)	0.317	0.417	0.053	1.54 (0.99-2.41)
rs201571	C	0.309	0.417	0.0069	1.6 (1.14-2.25)	0.292	0.417	0.018	1.73 (1.1-2.73)
rs6113491	C	0.401	0.477	0.066	1.36 (0.98-1.89)	0.392	0.477	0.111	1.41 (0.92-2.17)

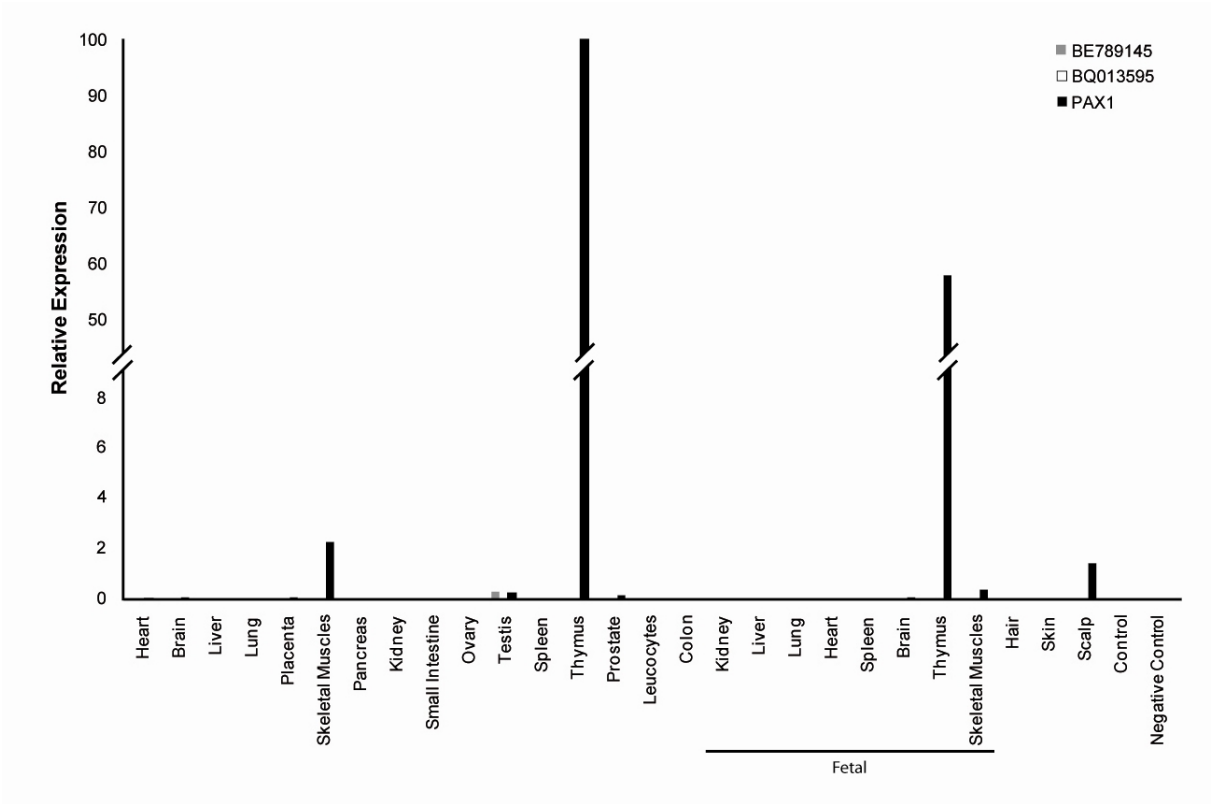
[a] AGA grade ≥ IIa
[b] AGA grade ≥ IV

Supplementary Table 4 Oligonucleotides used in quantitative PCR experiments

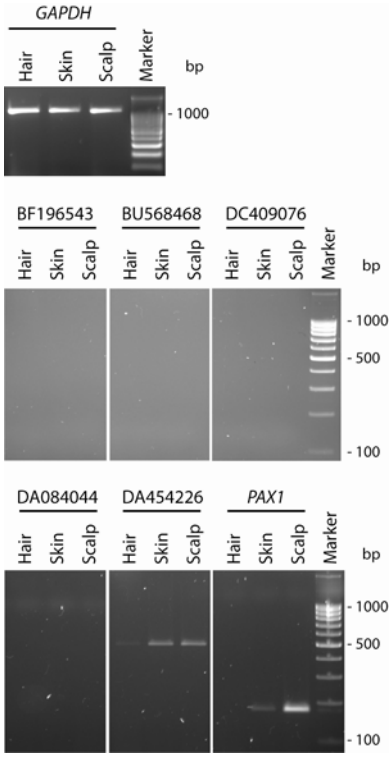
Assay	Name	Sequence (5' → 3')
TaqMan	BQ013595 forward primer	AGACATACTTTTCTATTCACCTTCCAATGATG
	BQ013595 reverse primer	GGTTCTGACCCTTATTATTGGAACTCA
	BQ013595 probe	AAGTCCCAACAAACTCCA
	BE789145 forward primer	ACCTGACATCAACGTACAAAGATTCA
	BE789145 reverse primer	GCGGCTTCTCCTCTGCAT
	BE789145 probe	CATGGCCACAAGAGTTA
	PAX1 forward primer	GGCTAGAGAAACCTGCCTTAGAG
	PAX1 reverse primer	CGCCCACGGCAGAGA
	PAX1 probe	ACACTCAGTCGGCCTCC
Semiquantitative PCR	DC409076_F	GAGCTTGGAAAAGTGGCAAG
	DC409076_R	GAGTCGGTGGGGACAGAATA
	BF196543_F	TATAATGGCCCTAGGCATCG
	BF196543_R	ACCAATCTTCAATGGGCAAG
	DA084044_F	ACCAATCTTCAATGGGCAAG
	DA084044_R	TCCCAGTGAGGTCTGAATCC
	DA454226_F	GTGGGTGTGGTGGAATTAGG
	DA454226_R	ATGAGTCTGCCTGGGTCTTG
	BU568468_F	CATCCTTCCAAGTCAGCACA
	BU568468_R	GGCTCTTGGTCAAACTCAGC
	*PAX1*_F	CGGACGTTTATGGAGCAAAC
	*PAX1*_R	TGGAGGCCGACTGAGTGTAT



Supplementary Figure 1 Genotype odds ratios of rs2180439 with increasing phenotypic divergence between cases and controls. All cases had AGA that was representative of the most severely affected 10% of the distribution for the respective age class. Thereof, cases used in the GWAS were the most severely affecteds and cases used in the replication step were less extreme. Controls of the GWAS were population based where males with extreme AGA had been excluded. Controls of the replication step represented the least-affected 20% of the distribution for their age class. ORs for heterozygotes (C/T) are indicated by filled diamond, ORs for homozygotes (T/T) are indicated by open squares, 95% CIs are indicated by bars.



Supplementary Figure 2 Expression of BQ013595, BE789145, and *PAX1* in human tissues. Expression levels were determined by real time PCR, normalized by cyclophilin and displayed relative to the strongest expression (*PAX1* in adult thymus). *PAX1* expression is detectable in scalp.



Supplementary Figure 3 Semiquantitative expression analysis of EST clusters near the associated LD block on chromosome 20p11. cDNA quantities of human hair, skin, and scalp for PCR reactions were adjusted using house keeping gene *GAPDH*. PCR products were analyzed after agarose gel electrophoresis using ethidium bromide staining. DA454226 shows weak expression in hair, skin, and scalp. *PAX1* shows weak expression in skin and considerable expression in scalp. bp, base pairs.

Supplementary Figure 4 Global distribution of the rs2180439 risk allele. Genotype data of the HGDP-CEPH Diversity Panel[1] was extracted from the HGDP-CEPH Diversity Panel Database Supplement 2 at http://www.cephb.fr/hgdp-cephdb/[2]. Worldwide distribution of the AGA risk allele T is indicated by the proportion of black shading in pie charts. Lower risk allele frequencies were observed in Oceanians and Native Americans.

References for the Supplementary Figure 4

1. Cann, H.M. *et al. Science* **296**, 261-2 (2002).
2. Jakobsson, M. *et al. Nature* **451**, 998-1003 (2008).

Supplementary Methods

German sample

Families and unrelated AGA and non-AGA individuals were recruited through various sources including press reports, and advertisements in magazines, newspapers and placards. Blood samples were drawn from 754 affected males who were <30 years of age having AGA of grades IV-VII, based on the classification of Hamilton[1] and modified by Norwood[2], or <40 years having AGA of grades V-VII. Based on these criteria, the AGA sample was representative of the most severely affected 10% of the distribution for the respective age class[1,2]. Of the affecteds, the 296 unrelated individuals with the strongest hair loss (mean±SD age [in years] 33.3±4.9, Hamilton/Norwood grades: median VI, lower quartile V, upper quartile VII) were selected for GWAS and 319 independent unrelated affecteds (mean±SD age [in years] 33.4±5.1, Hamilton/Norwood grades: median V, lower quartile V, upper quartile VI) were selected for the replication step. Two hundred thirty-four unaffected males were >60 years of age and without AGA (mean±SD age [in years] 67.6±5.9), representing the least affected 20% of the distribution for this age class. Four hundred ninety-one of the 754 affecteds were part of 352 nuclear families comprising of parents with one to four affected sons. As controls, we investigated 383 individuals (mean±SD age [in years] 61.1±5.3) from the Heinz Nixdorf Recall cohort (Risk Factors, Evaluation of Coronary Calcium and Lifestyle)[3] consisting of 182 male and 201 female probands (mean±SD age [in years] 61.1±5.3) who were randomly selected from the general population. Of 129 male controls, hair status information was available: 16 individuals with grade I (Hamilton/Norwood classification), 21 individuals with grade II, one individual with grade IIa, eight individuals with grade III, 21 individuals with grade III vertex, four individuals with grade IIIa, ten individuals with grade IV, one individual with grade IVa, eleven individuals with grade V, 27 individuals with grade VI, and nine individuals with grade VII. Thirty-six males with AGA grade VI and VII, respectively, were excluded from GWAS to increase phenotypic stratification between cases and controls. Females were included in the GWAS control sample as they were considered to show the representative AGA risk allele frequencies for population controls. The study was approved by the Ethics Committees of the Universities of Bonn, Düsseldorf and Essen, and informed consent was obtained from all participating individuals. All participants were of German descent.

Australian sample

As part of a population based twin study[4], hair status was documented in 291 unrelated male individuals for whom DNA was available. Measures of hair loss were obtained in the course of an extensive semi-structured telephone interview with respondent booklet, designed to assess physical, psychological and social manifestations of alcoholism and related disorders, conducted with 6,265 twins born 1964-71 from the volunteer based Australian Twin Registry. All males (45% of the sample) were asked to rate their degree of hair loss, if any, using the Hamilton/Norwood Baldness scale[1,2], which was printed in the respondent booklet. This data collection scheme was validated in a study by Ellis and colleagues[5]. Individuals with hair loss Hamilton/Norwood grade ≥IIa were classified as AGA affecteds and individuals with Hamilton/Norwood grade I were classified as controls. The study was approved by the Human Research Ethics Committee of the Queensland Institute of Medical Research, and informed consent was obtained from all participating individuals.

DNA extraction

EDTA anticoagulated venous blood samples were collected from all individuals. Lymphocyte DNA from German AGA affecteds, unaffecteds, families, and Australian samples was isolated by salting out with saturated NaCl solution[6] and DNA from German population-based controls was isolated using a Chemagic Magnetic Separation Module I (Chemagen, Baesweiler, Germany) according to the manufacturer recommendations. Genomic DNAs were diluted to working concentrations of 50 ng/µl for genome-wide genotyping and 2.5 ng/µl for the replication step.

Genotyping

Genome-wide genotyping of 317,503 SNPs in 150 German AGA affecteds using the Illumina HumanHap300 BeadChip and of 561,466 SNPs in 146 German affecteds and 383 German population based controls using the Illumina HumanHap550 BeadChip was conducted according to the Infinium II protocol from Illumina (Illumina, San Diego, USA). A DNA sample was deemed to have failed if it generated genotypes at fewer than 95% of loci of the respective BeadChip. A SNP was deemed to have failed if fewer than 95% of case or control samples generated a genotype at the locus. SNPs were excluded with a MAF <0.01 in controls and a Hardy-Weinberg-Equilibrium (genotypic χ^2 test with 1-degree of freedom, DF) $P<10^{-7}$ in controls. 531,695 SNPs passed quality criteria and were used for association analysis.

For the replication study, the best SNPs were selected based on the *P* values of the Cochran-Armitage trend test for the autosomes and the pseudoautosomal region and based on the allele frequency difference test for chromosomes X, Y, and the mitochondrion. To maximize efficiency in the replication step, four tagging SNPs of chromosome 20p11 were selected using the tagger algorithm of Haploview software[7] (pairwise tagging parameters: $r^2 \geq 0.8$ and LOD≥3) and CEU HapMap data[8]. The genome-wide best SNP, rs1998076 (Supplementary Table 1), was forced to be included and rs2180439 which was in perfect LD with rs1998076 was added for additional confidence in the most significant SNP. The genome-wide next best 30 SNPs excluding the *AR* locus (chromosome X: 65.5 - 67.2 Mb, build 36) were selected for replication analysis. Visual inspection of Illumina genotype clusters of selected SNPs revealed for rs4976846 a shifted genotype cluster for a subset of samples and the SNP was excluded from the genome-wide data set. Of the 34 selected SNPs four had to be excluded: rs11655206 and rs4805229 could not be included in the two plexes for MassArray genotyping, rs13079866 was a genotyping failure, and rs1139488 showed fuzzy genotype clusters, significant deviation from Hardy-Weinberg-Equilibrium, and 19 Mendel errors in the family data.

Genotyping for the replication step in the German and Australian samples was performed using the MassArray system on a Sequenom Compact MALDI-TOF device (Sequenom Inc., San Diego, CA, USA). Primer sequences and PCR/assay conditions can be obtained from the authors upon request. SNP call rates were >95.0% and samples with call rates <95% were excluded from the analysis.

Statistical analysis

Genome-wide association analysis at single marker level was performed by Cochran-Armitage trend test for the autosomes and the pseudoautosomal region and by the allele frequency difference test for chromosomes X, Y, and the mitochondrion. $P<5\times10^{-7}$ were assumed to be genome-wide significant as suggested by the Wellcome Trust Case Control Consortium[9]. Haplotype and family based analysis of the chromosome 20 locus was conducted with FAMHAP software[10]. 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[11,12]. For each marker (single locus analysis) and each marker combination (haplotype analysis), we used 10^{10} permutation replicates.

As rs2180439 was the most associated marker in the combined German sample, we carried out further statistical analysis based on that marker in the compound data set. We used logistic regression as suggested by Cordell and Clayton[13]. First, we tested for deviation from a multiplicative model on the OR scale by comparing a 2-DF model with parameter β_1 for a multiplicative SNP-effect and parameter β_2 for a dominance effect versus a 1-DF model with just β_1. Secondly, we tested for independent effects of the remaining four SNPs of the region. Therefore, we tested a 3-DF model with parameters β_1, β_2 and β_3 (parameter for multiplicative effect of additional SNP) versus the 2-DF model for rs2180439 (β_1, β_2). If this test was significant, we further tested for additional dominance at the new SNP (parameter β_4) by comparing the 4-DF model to the 3-DF model.

In order to test for interaction with the X chromosome, we considered the initial GWAS case sample (n=296) versus the male controls (n=146). For this data set, the model with β_1, β_3 and β_4 had a *P* value of 5.89×10^{-6}. Introduction of the X-chromosomal SNP rs1041668 improved the model-fit significantly ($P=1.32 \times 10^{-8}$), resulting in a 4-DF model with $P=1.17 \times 10^{-13}$. On top of this model, neither the interaction terms with rs2180439 ($P=0.975$) nor with rs6113491 ($P=0.984$) yielded significant improvement of the model fit.

LD structure of the chromosome 20 locus was displayed by GOLD heatmap[14]. Haplotype block definition of Gabriel and colleagues[15] was used. LD structure and haplotype blocks were displayed using Haploview software v4.0[7].

Expression analysis

To test for tissue specific expression of BQ013595, BE789145 and *PAX1*, we used the Human Multiple-Tissue cDNA (MTC) Panels I, II and the Human Fetal MTC Panel (Takara Bio Europe/Clontech, Saint-Germain-en-Laye, France). In addition, total RNA from hair follicles, skin and scalp was extracted (RNeasy Micro Kit, Qiagen). Quality and quantity of the RNA was analyzed on a NanoDrop ND-1000 spectrophotometer (Peqlab Biotechnologie, Erlangen, Germany) and reverse transcription was performed using SuperScript First Strand Synthesis System for RT-PCR (Invitrogen, Karlsruhe, Germany) with oligo dT primers. cDNAs were adjusted in an initial quantitative real-time PCR experiment using the TaqMan Endogenous Control Human Cyc (Cyclophilin, 4326316E; Applied Biosystems, Darmstadt, Germany) in which 5 µl of the MTC-cDNAs were set as reference. Relative quantifications in real-time experiments were performed in 384-well plates with the adjusted amounts of cDNAs per well in a total volume of 20 µl on the ABI Prism 7900HT Fast Real-Time PCR System (Applied Biosystems, Darmstadt, Germany) using Custom TaqMan Gene Expression Assays for

BQ013595, BE789145, and *PAX1* (Supplementary Table 4). Each sample was normalized by TaqMan Endogenous Control Human Cyc, and Ct values were transformed by $100{,}000{,}000/(2^{Ct})$. Values were calculated relative to the expression of *PAX1* in thymus which was set as 100%.

Semiquantitative RT-PCRs of ESTs DC409076, BF196543, DA084044, DA454226, BU568468, and *PAX1* and the house keeping gene *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase) of human hair, femoral skin, and scalp cDNA samples were performed using 12.5 pmol of each primer pair (Supplementary Table 4), annealing temperatures of 68°C for *GAPDH* and 59°C for ESTs and *PAX1*, 1.5 mM $MgCl_2$ and HotStarTaq (Qiagen, Hilden, Germany) in a PTC-200 thermal cycler (MJ Research, Waltham, MA, USA). PCR products of 20 cycles (*GAPDH*) and 35 cycles (ESTs and *PAX1*), respectively, were analyzed on 2% agarose gels using ethidium bromide staining and UV light.

Acession numbers

AR: NM_000044 (NCBI)

PAX1: NM_006192 (NCBI)

References for the Supplementary Methods

1. Hamilton, J.B. *Ann N Y Acad Sci* **53**, 708-28 (1951).
2. Norwood, O.T. *South Med J* **68**, 1359-65 (1975).
3. Schmermund, A. *et al. Am Heart J* **144**, 212-8 (2002).
4. Nyholt, D.R., Gillespie, N.A., Heath, A.C. & Martin, N.G. *J Invest Dermatol* **121**, 1561-4 (2003).
5. Ellis, J.A., Stebbing, M. & Harrap, S.B. *J Invest Dermatol* **110**, 849-53 (1998).
6. Miller, S.A., Dykes, D.D. & Polesky, H.F. *Nucleic Acids Res* **16**, 1215 (1988).
7. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. *Bioinformatics* **21**, 263-5 (2005).
8. Frazer, K.A. *et al. Nature* **449**, 851-61 (2007).
9. Consortium, W.T.C.C. *Nature* **447**, 661-78 (2007).
10. Becker, T., Cichon, S., Jönson, E. & Knapp, M. *Ann Hum Genet* **69**, 747-56 (2005).
11. Zhao, H. *et al. Am J Hum Genet* **67**, 936-46 (2000).
12. Knapp, M. & Becker, T. *Hum Hered* **56**, 2-9 (2003).
13. Cordell, H.J. & Clayton, D.G. *Am J Hum Genet* **70**, 124-41 (2002).
14. Abecasis, G.R. & Cookson, W.O. *Bioinformatics* **16**, 182-3 (2000).
15. Gabriel, S.B. *et al. Science* **296**, 2225-9 (2002).