nature genetics

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]

We carried out a genome-wide association study in 296 individuals with male-pattern baldness (androgenetic alopecia) and 347 controls. We then investigated the 30 best SNPs in an independent replication sample and found highly significant association for five SNPs on chromosome 20p11 (rs2180439 combined $P = 2.7 \times 10^{-15}$). No interaction was detected with the X-chromosomal androgen receptor locus, suggesting that the 20p11 locus has a role in a yet-to-be-identified androgen-independent pathway.

Androgenetic alopecia (AGA, MIM109200) is a common form of hair loss in both men and women. In men, this condition is commonly known as male-pattern baldness. Its origin is genetic[1], with *AR* (encoding the androgen receptor) located on the X chromosome being the only risk gene identified to date[2–4]. Association of AGA with a variety of diseases has been reported[5,6] and underlines the value of understanding its molecular basis.

We conducted a genome-wide association study (GWAS) to identify predisposing genetic factors in AGA. We genotyped 296 males with early-onset AGA (onset <40 years, Hamilton[7] and Norwood[8] hair-loss grades IV–VII: median VI, lower quartile V, upper quartile VII) and 383 population-based controls (201 females and 182 males), all of

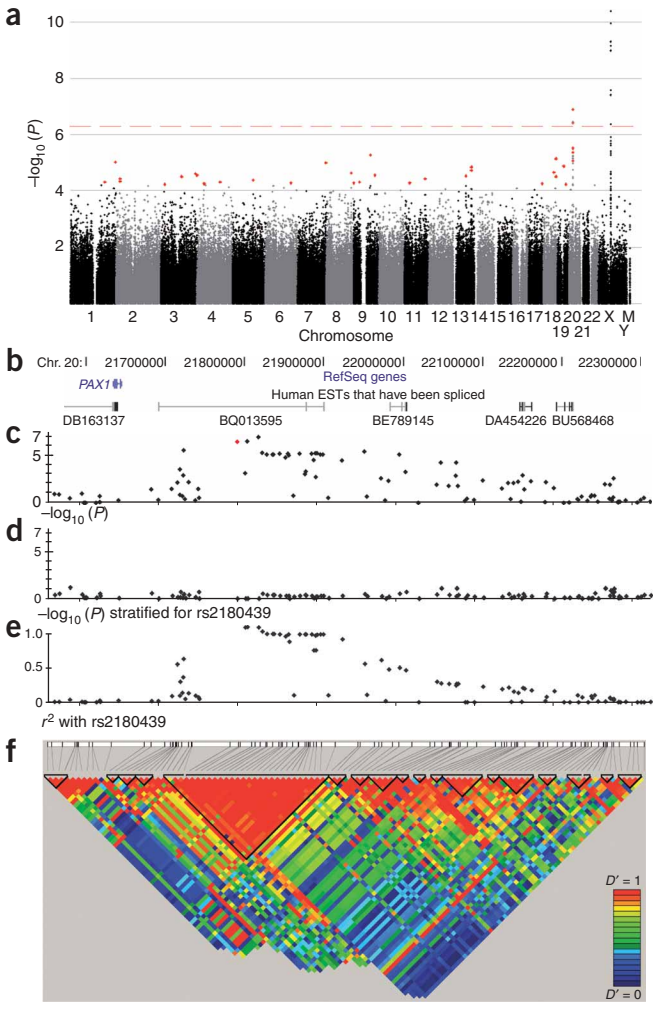

Figure 1 Genome-wide scan for association with AGA and details of the region 21,571,171– 22,321,597 on chromosome 20p11. (**a**) Negative $\log_{10}$ of the Cochran-Armitage trend test *P* for the autosomes and of the allele-frequency difference test *P* of chromosomes X, Y and the mitochondrion for SNPs that passed quality control. Chromosomes are shown in alternating colors, with SNPs selected for the replication step highlighted in red. Genome-wide significance level is indicated by dashed red line. (**b**) Transcript information of the chromosome 20 locus (UCSC genome browser, build 36). (**c**,**d**) Negative $\log_{10}$ association *P* values are shown for the GWAS (with rs2180439 indicated in red) and for an analysis stratified for rs2180439 (**d**). (**e**) The LD measured by r^2 between each SNP genotyped in the region and rs2180439 is shown. (**f**) LD displayed by GOLD heatmap and haplotype blocks indicated by black triangles were analyzed using Haploview software[15].

[1]Department of Genomics, Life & Brain Center, University of Bonn, D-53127 Bonn, Germany. [2]Department of Dermatology, University of Düsseldorf, D-40225 Düsseldorf, Germany. [3]Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn, D-53105 Bonn, Germany. [4]Genetic Epidemiology, Queensland Institute of Medical Research, QLD 4006 Brisbane, Australia. [5]Institute for Medical Informatics, Biometry and Epidemiology, University Duisburg-Essen, D-45122 Essen, Germany. [6]Clinic of Endocrinology, Central Laboratory Unit Research and Education, University Hospital of Essen, University Duisburg-Essen, D-45122 Essen, Germany. [7]Clinic of Cardiology, West-German Heart Center Essen, University Duisburg-Essen, D-45122 Essen, Germany. [8]Institute of Human Genetics, University of Bonn, D-53111 Bonn, Germany. [9]These authors contributed equally to this work. Correspondence should be addressed to A.M.H. (axel.hillmer@uni-bonn.de).

Received 12 March; accepted 20 July; published online 12 October 2008; doi:10.1038/ng.228

Table 1 Association between SNPs in the chromosome 20 locus and AGA in the German sample

SNP (position)[c]	Sample	Cases[d]	Controls[e]	MAF[a] Cases[d]	MAF[a] Controls[e]	Genotypes[b] Cases[d]	Genotypes[b] Controls[e]	*P*	OR (95% CI)[f]
rs6137444	GWAS	296	347	0.264 (C)	0.383 (C)	14/128/154	49/168/130	3.11×10^{-6}	1.74 (1.37–2.21)
(21,733,639 bp)	Replication	319	234	0.277 (C)	0.404 (C)	21/135/163	45/99/90	1.57×10^{-5}	1.76 (1.37–2.27)
	Combined[g]	605	579	0.269 (C)	0.391 (C)	35/255/315	93/267/219	2.20×10^{-10}	
rs2180439	GWAS	296	347	0.292 (C)	0.429 (C)	21/131/144	66/166/115	3.85×10^{-7}	1.82 (1.45–2.30)
(21,801,100 bp)	Replication	319	234	0.303 (C)	0.485 (C)	23/147/149	62/103/69	1.37×10^{-9}	2.17 (1.70–2.78)
	Combined[g]	605	579	0.293 (C)	0.452 (C)	43/268/294	127/269/183	2.67×10^{-15}	
rs1998076	GWAS	296	347	0.282 (A)	0.427 (A)	20/120/144	65/163/115	1.30×10^{-7}	1.90 (1.50–2.41)
(21,828,045 bp)	Replication	319	234	0.301 (A)	0.479 (A)	23/146/150	61/102/71	3.69×10^{-9}	2.13 (1.66–2.73)
	Combined[g]	605	579	0.292 (A)	0.448 (A)	43/267/295	126/267/186	7.73×10^{-15}	
rs201571	GWAS	296	347	0.289 (C)	0.411 (C)	17/137/142	61/163/123	4.31×10^{-6}	1.72 (1.36–2.17)
(21,961,514 bp)	Replication	319	234	0.314 (C)	0.483 (C)	30/140/149	58/110/66	2.21×10^{-8}	2.05 (1.60–2.62)
	Combined[g]	605	579	0.298 (C)	0.44 (C)	46/269/290	119/272/188	1.21×10^{-12}	
rs6113491	GWAS	296	347	0.359 (C)	0.483 (C)	29/154/112	88/159/100	8.63×10^{-6}	1.66 (1.33–2.08)
(22,005,415 bp)	Replication	319	234	0.364 (C)	0.447 (A)	38/156/125	77/105/52	8.13×10^{-10}	2.17 (1.70–2.77)
	Combined[g]	605	579	0.359 (C)	0.488 (A)	66/302/237	165/263/151	1.13×10^{-13}	

[a]Minor alleles are in parentheses. [b]Numbers shown correspond to the following genotypes: homozygotes for the minor allele of cases/heterozygotes/homozygotes for the major allele of cases. [c]NCBI build 36. [d]GWAS cases were the most extremely affected. [e]GWAS controls are population based, the most extreme AGA affected having been excluded; replication controls are males >60 years of age without AGA and representing the least affected 20%. [f]ORs were not calculated for the combined samples since the composition of the two control samples differed as regards the exclusion of AGA cases. [g]Twelve GWAS samples were excluded in the combined analysis because of a call rate <95% for the reduced SNP set.

German descent, with Illumina BeadChips (HumanHap300 and HumanHap550, respectively; **Supplementary Methods** online). For the analysis, 36 male controls with strong AGA were excluded, leaving 347 controls (**Supplementary Methods**). Following quality control, we carried out association analysis on 531,695 SNPs with a minor allele frequency (MAF) of >1% in controls. The average SNP call rate was 99.78%.

We found highly significant association in a 1.7-Mb region flanking the previously implicated *AR* locus (**Fig. 1**). The best SNP (rs1998076, $P = 1.3 \times 10^{-7}$) located outside the *AR* locus was on chromosome 20, a locus that was supported by 17 additional SNPs with $P < 10^{-5}$ (**Supplementary Table 1** online). Three SNPs at this locus reached genome-wide significance on the basis of criteria suggested elsewhere[9]. To replicate the association findings, we selected four tagging SNPs of the chromosome 20 region and the next best 29 SNPs (excluding the *AR* locus, **Supplementary Methods**). In addition, rs2180439, which was in perfect linkage disequilibrium (LD) with rs1998076, was included for the purpose of providing additional confidence in the most significant SNP. Selected markers were analyzed in an independent set of 319 affected (Hamilton and Norwood grades: median V, lower quartile V, upper quartile VI) and 234 unaffected (males >60 years of age without AGA) German individuals. The association with all five SNPs of the chromosome 20 locus was replicated ($P = 8.13 \times 10^{-10}$ for rs6113491; **Supplementary Table 2** online). In addition, rs10992241 on chromosome 9 showed significance. However, the putative risk allele (C) in the replication study differed from that in the GWAS (T), and thus we did not consider rs10992241 to be truly associated. After combining the German samples from the genome-wide and replication analyses, we found that rs2180439 showed the highest significance ($P = 2.67 \times 10^{-15}$, **Table 1**). Haplotype analysis did not significantly improve the association findings (data not shown). A subset of the affected sample consisted of index individuals of 352 nuclear families, and the results of a transmission disequilibrium test (TDT) were in accordance with the case-control findings with significant association for the five chromosome 20 SNPs and the highest significance for rs2180439 (TDT $P = 7.22 \times 10^{-7}$). We also carried out an initial step to investigate the importance of this locus in non-German populations by analyzing an Australian sample ($n = 291$) for the five chromosome 20 SNPs. Again, this locus showed significant association ($P = 0.0069$ for rs201571; **Supplementary Table 3** online). Notably, although rs201571 is highly significant across all three samples, different SNPs show the most significant *P* values in individual samples. This may suggest that the true causative variant has not yet been identified.

In the German sample, the cases used for the GWAS were those who were the most extremely affected and who were thus expected to show slightly stronger effects than the cases used for the replication step. Furthermore, the controls of the GWAS were less extreme in regards to phenotype than the unaffected controls used for the replication step, the latter representing the least-affected 20% of the distribution for this age group. The genotype-based odds ratios (ORs) of rs2180439 in the four possible case-control comparisons showed an increase from the least extreme (replication cases and GWAS controls, OR = 3.72, 95% CI = 2.18–6.34) to the most extreme (GWAS cases and replication controls, OR = 6.16, 95% CI = 3.48–10.92) for the homozygotes (**Supplementary Fig. 1** online). Correlation of effect size and phenotypic difference supported an association between rs2180439 and AGA. However, this relation was less clear for the heterozygotes. We used logistic regression[10] and tested for deviation from a multiplicative model on the OR scale in the combined German dataset. The test showed a tendency toward a dominant model for rs2180439 ($P = 0.023$). We also found evidence that rs2180439 and rs6113491 contribute independently to the risk of AGA ($P = 0.003$), and that rs6113491 showed a dominant effect ($P = 0.019$; **Supplementary Methods**). To observe this effect, however, we needed to use the combined sample, as rs6113491 did not remain significant after stratification of the GWAS data for rs2180439 (**Fig. 1**).

We tested chromosome 20p11 for interaction with the primary associated SNP at the *AR* locus (rs1041668). Introduction of rs1041668 improved the model fit significantly. However, neither the interaction terms with rs2180439 ($P = 0.975$) nor those with rs6113491 ($P = 0.984$) yielded significant improvement of the model fit. Thus, our

data do not provide any evidence for interaction effects between the two regions, although this cannot be ruled out by the present data.

Two ESTs, BQ013595 and BE789145, map to the associated region on 20p11 (**Fig. 1**). Because a gene contributing to the development of AGA would be expected to be expressed in the human scalp, we quantified the expression of BQ013595 and BE789145 in various human tissues, including skin, hair and scalp (**Supplementary Methods** and **Supplementary Table 4** online). In addition, we included the closest reference sequence gene, *PAX1* (paired box 1). BE789145 was not detectable in skin, hair or scalp, and BQ013595 showed very low expression in hair and scalp. *PAX1* was found to be expressed at very low levels in femoral skin and hair, but showed considerable expression in scalp skin (**Supplementary Fig. 2** online). Although *PAX1* is located outside of the associated LD block, the expression data might suggest that *PAX1* confers the AGA-relevant effect at this locus and that a regulatory variant within the associated LD block may modulate its expression. Of note, long-range downstream enhancers have been reported for *PAX6* (ref. 11). The expression analysis of five ESTs outside the associated LD block (**Supplementary Methods** and **Supplementary Table 4**) showed weak expression of DA454226 in, but not restricted to, scalp (**Supplementary Fig. 3** online), compatible with a possible role of DA454226 in hair biology.

The AGA risk allele rs2180439[T] has a frequency of 0.57 in our German population-based control sample of 383 individuals. The etiological fraction for rs2180439 is estimated at 0.32, underlining the importance of this locus in the development of early-onset AGA. The frequency of the risk allele varies worldwide from 0.03 to 0.86 (**Supplementary Fig. 4** online). This may account in part for population-specific differences in AGA prevalence, but it does not explain the lower prevalence reported in Southeast Asian populations[8,12]. It is of note that the chromosome 20 locus did not show linkage to AGA in a recent study[13]. This supports the value of the association approach in detecting common low-penetrance susceptibility variants for complex diseases[14]. The linkage regions are still very promising regions for harboring common disease–associated genes, as the replication step included only the top 30 of the large number of SNPs showing significant association in the GWAS, among them many located in the linkage regions. It will be a matter of future studies using larger GWAS samples and independent replication to identify the true-positive association findings in these regions. It will be also interesting to use these data for exploring the existence of common variants in genes responsible for rare monogenic forms of hair loss.

Our results place the genetic basis of AGA in a genome-wide context. The chromosome 20p11 locus has a strong effect on the development of early-onset AGA, with no obvious genetic connection to the androgen pathway. Discovering the functional consequences will increase our understanding of the molecular and cellular basis of scalp hair loss.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank all probands for participating in the study and providing blood samples. This study was supported by the Deutsche Forschungsgemeinschaft (FOR 423 and Emmy Noether) and the Alfried Krupp von Bohlen und Halbach-Stiftung. The Heinz Nixdorf Recall cohort was established with the support of the Heinz Nixdorf Foundation (G. Schmidt, Chairman).

AUTHOR CONTRIBUTIONS

M.M.N. and R.K. initiated the study; A.M.H. and F.F.B. contributed to the study design; M.M.N. and A.M.H. coordinated the work and prepared the manuscript, with feedback from the other authors; R.K., S. Hanneken, S.E., and A.-K.K. diagnosed the affected and unaffected German individuals and collected the blood samples; S.M., K.-H.J., M.B.-P. and R.E. recruited and characterized German population controls; N.G.M. and D.R.N. recruited and characterized Australian sample; F.F.B., T.W.M., M.A.A., Z.Z.Z., G.W.M. and R.R. prepared DNA and performed genotyping; M.S., A.F., S. Herms, T.B., and D.R.N. performed statistical analysis; F.F.B. performed expression analysis. M.M.N., A.M.H., F.F.B., T.B., N.G.M, D.R.N., T.F.W., M.P.B., S.C., P.P., and R.C.B. analyzed and interpreted data.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Nyholt, D.R., Gillespie, N.A., Heath, A.C. & Martin, N.G. *J. Invest. Dermatol.* **121**, 1561–1564 (2003).
2. Ellis, J.A., Stebbing, M. & Harrap, S.B. *J. Invest. Dermatol.* **116**, 452–455 (2001).
3. Hillmer, A.M. *et al. Am. J. Hum. Genet.* **77**, 140–148 (2005).
4. Ellis, J.A. *et al. Hum. Genet.* **121**, 451–457 (2007).
5. Lotufo, P.A., Chae, C.U., Ajani, U.A., Hennekens, C.H. & Manson, J.E. *Arch. Intern. Med.* **160**, 165–171 (2000).
6. Matilainen, V., Koskela, P. & Keinänen-Kiukaanniemi, S. *Lancet* **356**, 1165–1166 (2000).
7. Hamilton, J.B. *Ann. NY Acad. Sci.* **53**, 708–728 (1951).
8. Norwood, O.T. *South. Med. J.* **68**, 1359–1365 (1975).
9. The Wellcome Trust Case Control Consortium. *Nature* **447**, 661–678 (2007).
10. Cordell, H.J. & Clayton, D.G. *Am. J. Hum. Genet.* **70**, 124–141 (2002).
11. Kleinjan, D.A. *et al. Dev. Biol.* **299**, 563–581 (2006).
12. Paik, J.H., Yoon, J.B., Sim, W.Y., Kim, B.S. & Kim, N.I. *Br. J. Dermatol.* **145**, 95–99 (2001).
13. Hillmer, A.M. *et al. Am. J. Hum. Genet.* **82**, 737–743 (2008).
14. Risch, N. & Merikangas, K. *Science* **273**, 1516–1517 (1996).
15. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. *Bioinformatics* **21**, 263–265 (2005).