Genome-Wide Association Study Reveals a Novel Locus for Male Pattern Baldness

J Brent Richards, Xin Yuan, Frank Geller, Dawn Waterworth, Veronique Bataille, Daniel Glass, Kijoung Song, Gerard Waeber, Peter Vollenweider, Katja K. Aben, Lambertus A. Kiemeney, Bragi Walters, Nicole Soranzo, Unnur Thorsteinsdottir, Augustine Kong, Thorunn Rafnar, Panos Deloukas, Patrick Sulem, Hreinn Stefansson, Kari Stefansson, Tim D Spector*, Vincent Mooser*

Supplementary Methods

Samples and Phenotype

CoLaus. Members of this cohort were randomly selected from permanent residents of Lausanne, Switzerland between 2003 and 2006.[1,2] Only individuals with 4 European grandparents were eligible for inclusion in the GWA study. Men were considered to have androgenic alopecia if they experienced Hamilton grade V-VII alopecia between the ages of 35 and 65, while controls were men displaying no hair loss (Hamilton grade I) between ages 45 and 55 or grade II between the ages of 56 and 75. Age of onset was assessed by questionnaire. Participants were recruited and assessed by nurses who had been specifically trained to assess for androgenic alopecia.

TwinsUK. The TwinsUK cohort is a population-based sample of Britons,[3] unselected for any disease or trait, which is representative of the UK singleton population.[4] Standardised photographs were assessed in men for the presence of androgenic alopecia by a dermatologist, defined as Hamilton grade V-VII and controls were those without androgenic alopecia (Hamilton grade I). Since photographs were not available for a small subset of the male TwinsUK cohort (n = 78), questionnaire data was used in this subset, where men were asked if they suffered from male pattern baldness. Exclusion of these 78 males did not change the significance of the results or the magnitude of the odds ratio (OR = 1.67 [1.14 – 2.47], $p = 9.0 \times 10^{-3}$). Women from TwinsUK were asked if they had experienced considerable hair loss. Controls were defined as women not reporting considerable hair loss after 50 years of age and cases were those reporting considerable hair loss prior to 50 years of age. The mean (±SD) age of men and women in this cohort was 52.7 ± 14.2 years and 50.7 ± 13.3 years, respectively.

Icelandic study population. Individuals with hair loss were identified from questionnaire data asking "Have you experienced considerable thinning or loss of your hair? (Not including hair-thinning side-effects

of medications or treatments)." The age of onset was specified by the question "How old were you when you first experienced considerable thinning or loss of your hair? (Not including hair thinning side effects of medications or treatments)." Affected individuals were defined as individuals with considerable hair loss before the age of 50 years (N=933, 536 males, 397 females, mean age 44.4 ± 14.3 years). Unaffected persons were defined as individuals older than 50 years who reported no hair loss (N=679, 198 males, 481 females, mean age 61.0 ± 8.0 years).

Dutch study population. The 463 individuals in this study are a subset of the 1,013 prostate cancer patients presented previously,[5] who had answered questions regarding androgenic alopecia. Patients completed a questionnaire and matched their hair pattern at 20 years, 40 years and present age to a Hamilton grading schema. Affected individuals were those reporting Hamilton grade IV-VII by age 40 and controls were those with Hamilton grade I at age 50 years or older, or Hamilton grade II-III at age 60 years or older. The mean age of cases and controls was 66.7 (±6.4) years and 67.7 (±5.9) years, respectively.

All studies were approved by institutional ethics review committees at the relevant organisations, and all participants provided written informed consent. The CoLaus study was sponsored in part by GlaxoSmithKline and the participants consented for the use of their data and samples by GlaxoSmithKline and affiliates.

Genotyping

CoLaus. Subjects were removed from the sample if their sex was inconsistent with genetic data from the X chromosome; the genotype call-rate was <90%; duplicate samples produced inconsistent genotypes. SNPs were removed from analysis if the SNP call-rate was < 95%; displayed Hardy-Weinberg disequilibrium ($p<1.0 \times 10^{-7}$); were monomorphic across all samples; had a minor allele frequency (MAF) < 1%. After these exclusions, 370,102 SNPs remained. Analysis was performed using the PLINK software package.[6] The quantile-quantile (Q-Q) plot is displayed in **Figure S1**, and genomic inflation factor λ was = 1.016, suggesting that there was little evidence of cryptic relatedness in the sample .

Figure S1 also demonstrates a marked excess of "significantly" associated SNPs. Genotypes were obtained with the BRLMM algorithm as described elsewhere.[2]

TwinsUK. Although a portion of the female TwinsUK cohort has been genotyped using the Illumina HumanHap 300k duo chip[3], males have not, therefore rs1160312 was genotyped in men using TaqMan Assay by Design (Applied Biosystems) and analyzed on a 7900HT machine. All genotyping data were scored blind to case-control status and were double scored by a second operator to minimize error. Only subjects of self-defined European ancestry were included in this replication cohort. Since women from this cohort have undergone genome-wide genotyping,[3] the rs1160312 SNP was imputed using IMPUTE[7] and all genotypes with a maximum posterior probability <90% were excluded. The mean maximum posterior probability for this SNP in the imputed data in women was 99.7%.

Icelandic cohort. The SNP rs1160312 is not on the Illumina chip and was genotyped using single track assay by deCODE Genetics in Reykjavik, Iceland applying the Centaurus (Nanogen) platform .[8] The quality of the Centaurus SNP assay was evaluated by genotyping the CEU HapMap samples and comparing the results with the HapMap publicly released data. All Dutch individuals and 916 affected individuals from Iceland were genotyped for rs1160312. In a subset of 459 affected individuals additional data for rs201571 and rs6036025 were available from genome-wide SNP genotyping using genotyping systems and specialized software from Illumina (Human Hap300 and Human Hap300-duo+ Bead Arrays, Illumina, confirming the perfect linkage disequilibrium (r^2=1) between the haplotype consisting of allele T at rs201571 and allele G at rs6036025 with allele A at rs1160312 observed in the HapMap CEU data. Thus, additional genotypes for 17 cases and 679 controls, all with a call rate greater than 98% for all SNPs and both genotypes for rs201571 and rs6036025 from genome-wide studies were used in the analysis.

Dutch Cohort. All Dutch individuals were genotyped for rs1160312 using single track assay by deCODE Genetics in Reykjavik, Iceland applying the Centaurus (Nanogen) platform .[8]

Statistical Methods

The GWA study was performed in the CoLaus study using the PLINK software package (version 1.0).[6] The effects of population stratification were reduced in the design stage by including only individuals with

4 grandparents who were of European origin. Quantile-quantile plots were inspected for evidence of generalized inflation of test statistics. The genomic inflation factor, λ, was calculated using the mean of the χ^2 tests generated on all SNPs that were tested and was found to be 1.016. The relationship between the rs1160312 SNP and androgenic alopecia was assessed, using an additive model, in the TwinsUK cohort by employing logistic regression analysis, accounting for the non-independence of twin pairs using the regression cluster option in Stata 10.0 (College Station, Texas, USA). Since the haplotype consisting of allele T at rs201571 and allele G at rs6036025 is in perfect linkage disequilibrium with rs1160312 allele A (r^2=1 in the HapMap CEU data), additional genotypes of rs1160312 in the Icelandic population were inferred from the genotypes for rs201571 and rs6036025 applying a likelihood approach as previously described.[9] Association tests in the Icelandic and Dutch cohorts were performed applying a likelihood procedure,[10] implemented in the NEMO software package. We tested the association of an allele to hair loss using a standard likelihood ratio statistic that, if the subjects were unrelated, would have asymptotically a χ^2 distribution with one degree of freedom under the null hypothesis. Allele-specific ORs and associated *P* values were calculated assuming a multiplicative model for the two chromosomes of an individual.[11] For all 3 SNPs (rs201571, rs6036025 and rs1160312) there was no significant deviation from HWE in the population controls ($P > 0.1$). Some individuals in the Icelandic study groups were related to each other, causing the aforementioned χ^2 test statistic to have a mean >1. We estimated the inflation factor by using a previously described procedure[12] (Stefansson, 2005) in which we simulated genotypes through the genealogy of the 1,612 Icelanders analyzed in the present study (number of simulations = 100,000), giving inflation factors of 1.023 and 1.031 for the analysis of males and females, respectively. The results for the Icelandic samples presented are based on adjusting the χ^2 statistics by these factors. Meta-analysis of cohort level data was performed using inverse-variance methods. To assess the risk of androgenic alopecia, sensitivity, specificity, positive and negative predictive values attributed to both the *AR* and 20p11.22 loci, individuals with the risk allele at the rs6625163 (hemizygotes) and one or two risk alleles at rs1160312 were compared to individuals with neither risk allele in the CoLaus cohort. To assess the variance in androgenic alopecia explained by the presence of both risk alleles, as described, a logistic regression was performed. The variance explained was taken to be the resultant pseudo-R^2 from this logistic regression. To assess for genome-wide significance, we used a p-value threshold of 5.0×10^{-8},

which is a Bonferroni correction of 0.05, by the number of independent SNPs identified in HapMap Phase II.[13]

References

1. Firmann, M. et al. The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc Disord* **8**, 6 (2008).
2. Sandhu, M.S. et al. LDL-cholesterol concentrations: a genome-wide association study. *Lancet* **371**, 483-91 (2008).
3. Richards, J.B., Rivadeneira, F. & M Inouye, T.P., N Soranzo, SG Wilson,, M Falchi, R Gwilliam, KR Ahmadi, P Arp, P Whittaker, M Jhamai, V Kumanduri, M Moorhouse, JB van Meurs, A Hofman, HAP Pols,, G Zhai, AM Valdes, BS Kato, T Andrew, BH Mullin, F Zhang, P Deloukas, AG Uitterlinden, TD Spector. Genome-Wide Association Study Reveals Genetic Variants Associated with Bone Mineral Density, Osteoporosis and Osteoporotic Fractures. *The Lancet* **371**, 1505-1512 (2008).
4. Andrew, T. et al. Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res* **4**, 464-477 (2001).
5. Gudmundsson, J. et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* **39**, 977-83 (2007).
6. Purcell, S. et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* **81**, 559-75 (2007).
7. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).
8. Kutyavin, I.V. et al. A novel endonuclease IV post-PCR genotyping system. *Nucleic Acids Res* **34**, e128 (2006).
9. Amundadottir, L.T. et al. A common variant associated with prostate cancer in European and African populations. *Nat Genet* **38**, 652-8 (2006).
10. Gretarsdottir, S. et al. The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat Genet* **35**, 131-8 (2003).
11. Falk, C.T. & Rubinstein, P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann Hum Genet* **51**, 227-33 (1987).
12. Stefansson, H. et al. A common inversion under selection in Europeans. *Nat Genet* **37**, 129-37 (2005).
13. Frazer, K.A. et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-61 (2007).

Figure S1: Quantile-Quantile Plot for the Test Statistics for Stage 1

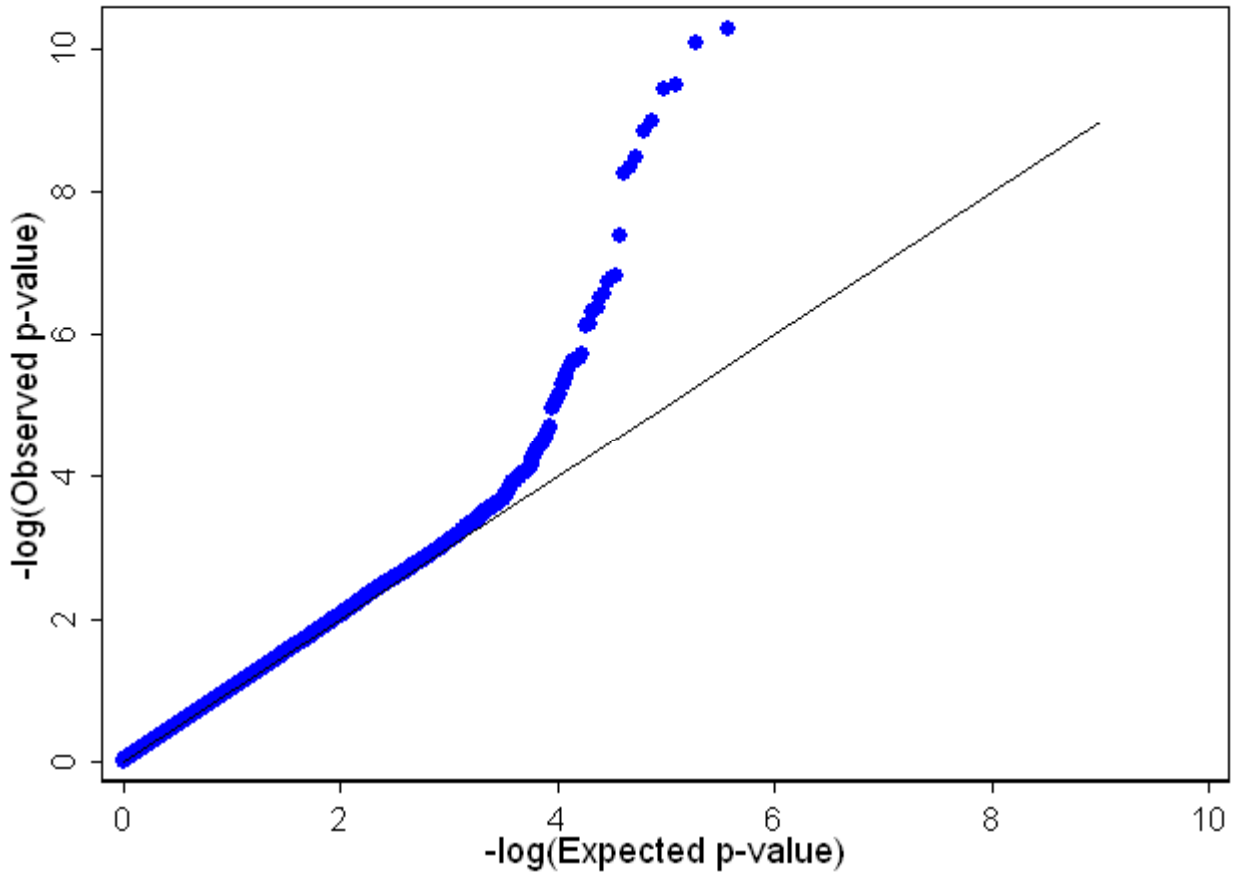

Figure S2: Genome-Wide Manhattan Plot for Androgenic Alopecia