 Recent positive selection of a human androgen receptor/ectodysplasin A2 receptor haplotype and its relationship to male pattern baldness

Axel M. Hillmer · Jan Freudenberg · Sean Myles · Stefan Herms · Kun Tang · David A. Hughes · Felix F. Brockschmidt · Yijun Ruan · Mark Stoneking · Markus M. Nöthen

Abstract Genetic variants in the human androgen receptor gene (AR) are associated with male pattern baldness (androgenetic alopecia, AGA) in Europeans. Previous observations of long-range linkage disequilibrium at the AR locus are consistent with the hypothesis of recent positive selection. Here, we further investigate this signature and its relationship to the AGA risk haplotype. The haplotype homozygosity suggests that the AGA risk haplotype was driven to high frequency by positive selection in Europeans although a low meiotic recombination rate contributed to the high haplotype homozygosity. Further, we find high levels of population differentiation as measured by $F_{ST}$ and a series of fixed derived alleles along an extended region centromeric to AR in the Asian HapMap sample. The predominant AGA risk haplotype also carries the putatively functional variant 57K in the flanking ectodysplasin A2 receptor gene (EDA2R). It is therefore probable that the AGA risk haplotype rose to high frequency in combination with this EDA2R variant, possibly by hitchhiking on a positively selected 57K haplotype.

Introduction

Androgens mediate a wide range of developmental and physiological responses through the androgen receptor (AR), and play a crucial role in several stages of male development (Lee and Chang 2003). Mutations in the androgen receptor gene (AR) may cause androgen insensitivity syndrome (Ris-Stalpers et al. 1990) and spinobulbar muscular atrophy (La Spada et al. 1991) and are found in prostate and breast cancer tissues (reviewed by Gottlieb et al. 2004; Heinlein and Chang 2004; Weiss et al. 2005). Various studies have tested for association between...
variability in AR and a number of phenotypes including male infertility, breast cancer, and prostate cancer, but the results have been inconsistent (reviewed by Rajender et al. 2007). We recently found genetic variability in a 1 Mb region within and centromeric to AR to be associated with androgenetic alopecia (AGA) at a highly significant level (Hillmer et al. 2005, 2008a). The large size of the associated linkage disequilibrium (LD) block and the high frequency of the derived haplotype were suggestive of recent positive selection in Europeans (Hillmer et al. 2005). Since AR is linked to the male reproductive system, it is conceivable that variability in AR might have an impact on reproductive success and therefore be subject to natural selection. Consistent with this hypothesis are more recent studies which have reported significant signals of positive selection at the AR locus (Altshuler et al. 2005; Voight et al. 2006), although the major signals were located closer to the centromere.

In the present study, we have further analyzed the signature of positive selection in and around the AR locus and its relationship to the AGA susceptibility variants there. Our data demonstrate that the AR locus is a unique X-chromosomal region with striking differences between human populations. Interestingly, it is the AGA risk haplotype that shows evidence for positive selection in the European population. The AGA risk haplotype also carries a derived non-synonymous allele (57K) in the flanking ectodysplasin A2 receptor gene (EDA2R). The 57K allele may have been a target of positive selection in East Asians: it shows large allele frequency differences between populations (Sabeti et al. 2007), lies in a likely functional domain of the EDA2R protein, and the ancestral allele, 57R, is conserved from human to rat. Thus, the AGA risk haplotype may have hitchhiked to high frequency in Europeans as a result of positive selection on the linked 57K allele in EDA2R.

Materials and methods

Haplotype bifurcation and extended haplotype homozgyosity (EHH)

Phased haplotype data of the non-pseudoautosomal region of the X chromosome of CEU (90 X chromosomes), CHB + JPT (134 X chromosomes), and YRI (90 X chromosomes) were downloaded from the International HapMap project home page (http://www.hapmap.org/, release 16) (Altshuler et al. 2005; Gibbs et al. 2003). The European Americans (CEU) are referred to as “European”, and the Yoruba (YRI) from Nigeria are referred to as “African”. The Han Chinese and Japanese (CHB + JPT) data were pooled and are referred to as “East Asian”. SNPs polymorphic in the respective populations were selected. Haplotype bifurcation, EHH, and relative EHH (REHH) were analyzed using Sweep software v1.1 (Sabeti et al. 2002), which was downloaded from (http://www.broad.mit.edu/mpg/sweep). The longest non-overlapping haplotype blocks with between 3 and 20 SNPs were determined using the block definition of Gabriel et al. (2002). The haplotype block which overlapped with the 5’ end of AR was defined as the core for bifurcation, EHH, and REHH analysis. The EHH approach proposed by Sabeti et al. (2002) compares rates of LD decay between haplotypes. The REHH corrects for local variation in the recombination rate by comparing the EHHs of different core haplotypes present at a locus. Only core haplotypes with frequencies ≥0.05 were selected for presentation. To determine the significance of EHH, the EHHs within a distance of 0.5 cM of the selected core were compared to the EHHs within a distance of 0.5 cM of all other cores of the non-pseudoautosomal X chromosome.

Haplotype block structure

Genotype data, one Mb upstream and downstream from the AR start site (chrX:65,546,894…67,546,894, build 35) of CEU, CHB + JPT, and YRI were downloaded from the International HapMap project home page (http://www.hapmap.org/, release 21a) (Altshuler et al. 2005) and haplotype structure was analyzed with Haplovew software v4.0 (Barrett et al. 2005) using the definition of Gabriel et al. (2002).

Derived allele frequencies

The phased haplotype data from 1 Mb upstream and downstream of AR of those SNPs, which were polymorphic in at least one population were downloaded (HapMap release 21a), and chimpanzee alleles were defined as ancestral alleles (Mikkelsen et al. 2005). In addition, macaque alleles were used to identify the ancestral alleles and SNPs with discrepant predictions from chimpanzee and macaque were excluded.

$F_{ST}$ analysis

Genotype data (build 35) were downloaded from the HapMap project home page (release 20). Only unrelated individuals were included: 60 European American, 45 Han Chinese, 45 Japanese, and 60 Yoruba. SNPs of the pseudoautosomal region of the X chromosome were not considered. A weighted average $F_{ST}$ was calculated according to equation (10) of Weir and Cockerm (1984) for 759 non-overlapping 200 kb windows across the X chromosome. The windows were designed such that AR was centered inside a single window. For each $F_{ST}$ comparison, P values were defined as the proportion of 200 kb windows with $F_{ST}$ values equal to or greater than the $F_{ST}$ value in the window...
containing AR. Similarly, the 95 and 99% thresholds represent the \( F_{ST} \) values above which 5 and 1% of the windows’ \( F_{ST} \) values lie, respectively. We define the value of the test statistic as “significant” if it lies within the most extreme 5% of the empirical distribution of that test statistic. We acknowledge that our \( P \) value only represents the position of a value relative to the empirical distribution and that this is not a statistical “significance test”.

Results

Haplotype homozygosity

Under the influence of positive selection, an allele may rise to a high frequency before successive recombination events disrupt the haplotype on which the allele lies (reviewed by Bamshad and Wooding 2003). This phenomenon is the rationale behind the extended haplotype homozygosity (EHH) test which exploits the assumption that, under neutrality, LD of high-frequency haplotypes tends to be less extensive than that of low-frequency haplotypes. To test for deviation from neutrality at the human AR locus, we investigated EHH using HapMap data extending 1 cM upstream and downstream of AR. The 5’ end of AR was chosen as the core haplotype region since we had previously found strong association between variants within this region and AGA in the German population (Hillmer et al. 2005). Analysis of the European genotype data revealed three frequent core haplotypes, with the most frequent haplotype (Eur-H1; European haplotype 1) carrying four AGA risk alleles (Fig. 1a). Bifurcation of Eur-H1 toward the centromere is low with respect to the high frequency of this haplotype, and compared to the less frequent haplotypes Eur-H2 and Eur-H3 (Fig. 1b). This is reflected by the EHH of Eur-H1 toward the centromere, which is higher than the EHH of the less frequent haplotypes Eur-H2 and Eur-H3 (Fig. 1d). The Eur-H1 EHH of 0.79 centromeric to the core is significantly higher (\( P \) value = 0.028) than the EHH of other X-chromosomal core haplotypes of comparable frequency (Fig. 2). However, the relative EHH test statistic (REHH), which corrects for local variation in the recombination rate and is 2.75 for Eur-H1 centromeric to the core, is not unusually high compared to the empirical distribution (\( P \) value = 0.151). Toward the telomere, both EHH, and REHH were not significantly different than expectation.

Haplotype specific LD differed considerably between populations. In Africans, the diversity of haplotypes was higher and the bifurcations more complex than in the European population (Supplementary Fig. 1). Afr-H3 (which is

![Fig. 1](https://example.com/Fig1.png)

AR promoter haplotype frequency, bifurcation, EHH, and REHH in the European population. Phased HapMap data was analyzed using Sweep software. a Haplotypes with frequencies >5% are shown. The three core haplotypes are numbered 1–3 and are colored to match the colors in parts b–e. The ancestral alleles were determined using the chimpanzee sequence. Positions that carry the ancestral allele are represented by dot and in cases where the ancestral allele could not be determined, the nucleotides carried by each haplotype are indicated in light gray. b Freq, frequency. b Haplotype bifurcation plots of the three common AR promoter haplotypes. The thickness of the line represents the frequency of each haplotype. Bifurcation occurs when haplotype homozygosity breaks down. c Relative gene positions (boxes) and analyzed SNPs (gray vertical lines) are shown; SNPs for core haplotypes are in blue, non-synonymous SNP rs1385699 is in red. d Plots of decay of EHH and REHH (e) for increasing distances from AR for each common haplotype are shown.
closest to Eur-H1) has a frequency of only 12% and an EHH of 0.65 \( (P \text{ value} = 2.3 \times 10^{-4}) \) centromeric to the core (Supplementary Fig. 1). However, the centromeric REHH of Afr-H3 is 5.5 \( (P \text{ value} = 0.13) \). The variability centromeric to AR is markedly lower in the East Asian population compared to the European and African populations. Therefore, haplotype comparisons between the populations are not straightforward, and none of the East Asian core SNPs has been shown to be associated with AGA. Due to low variability within and centromeric to AR, bifurcations of the core toward the centromere are low (Supplementary Fig. 2). In the East Asian population, EHH for all three high-frequency haplotypes in both directions of the core have \( P \) values < 0.05, but the respective REHH values were not unusually high.

Thus, the EHH analysis shows that it is the AGA risk haplotype that displays extended haplotype homozygosity from AR toward EDA2R, but it does not provide a significant REHH signal.

Differentiation of common SNPs between the three major continental populations

To understand more clearly the overall LD structure at the AR locus, we compared HapMap genotype data of the three major continental populations. The LD and haplotype block structure differ markedly across the populations. Centromeric of AR is a region of high LD in the European population (Supplementary Fig. 3b). This region includes EDA2R which is located more than 900 kb centromeric to AR. The LD in this region is lower in the African sample, in agreement with previous studies which have consistently found lower LD in African than in non-African populations (Hinds et al. 2005; Altshuler et al. 2005) (Supplementary Fig. 3c). In contrast, only a few SNPs between EDA2R and AR are polymorphic in the East Asian population, and the distant polymorphic SNPs are in low LD (Supplementary Fig. 3d). This is in accordance with the low haplotype bifurcation in the East Asian population. Again, the most striking differences between the investigated populations were found centromeric to AR.

The observation of low variability in the East Asian population prompted us to investigate the distribution of ancestral and derived alleles in East Asia. Comparison of the derived allele frequencies of Europeans, Africans, and East Asians revealed remarkable differences. Most SNPs in and between EDA2R and AR which are polymorphic in at least one of the three populations were observed to be fixed for the derived allele in the East Asian sample (Supplementary Fig. 4), consistent with a recent selective sweep in East Asians at this locus.

Geographically restricted positive selection can create large allele frequency differences between populations. The \( F_{ST} \) statistic captures the difference in allele frequency between populations and has values that range from 0 (no differentiation) to 1 (fixed difference between populations). It has previously been shown that positively selected alleles tend to accumulate in the top tail of the \( F_{ST} \) distribution (Beaumont and Balding 2004; Pollinger et al. 2005) and that putatively functional variants are more likely to display increased \( F_{ST} \) (Barreiro et al. 2008). We calculated average \( F_{ST} \) in non-overlapping 200 kb windows across the X chromosome. \( F_{ST} \) of three consecutive windows (600 kb) between EDA2R and AR were significantly high \( (P < 0.01) \) in the comparisons between the East Asian and African samples (Fig. 3). The flanking windows containing EDA2R (2nd centromeric window) and AR (1st telomeric window) were also significantly higher than expected \( (P = 0.024 \text{ and } P = 0.03) \), respectively. The European/East Asian comparison showed no significant differences at this locus. In contrast, the European/African analysis revealed significant average \( F_{ST} \) values for the windows between EDA2R and AR, but no significance for the windows containing the genes themselves \( (P = 0.078 \text{ and } P = 0.089) \). Thus the \( F_{ST} \) analysis shows that extended haplotype homozygosity of the AGA risk haplotype is accompanied by strong differentiation between human populations and once more
highlights that the region between $EDA2R$ and $AR$ is a region of particular interest.

Family based recombination rate at the AR locus

Although the earlier-reported haplotype homozygosity at the $AR$ locus shows a pattern consistent with the influence of positive selection, it is important to note that the influence of recombination rate variation on extended haplotype homozygosity is still being explored (O’Reilly et al. 2008). Such research has pointed out that the existing methods for identifying loci with signatures of recent positive selection are prone to biases towards regions with low recombination rates (Myles et al. 2008; Nielsen et al. 2007). What is more,
an increased population recombination rate is associated with those functional categories of genes that show faster and accelerated sequence evolution between species, an observation which led to the hypothesis that evolution could favor recombination at loci subject to stronger directional selection (Frazer et al. 2007; Freudenberg et al. 2007). For these reasons, it is important to consider the local recombination rate when hypothesizing about a recent history of non-neutral evolution at a genetic locus. A low recombination rate can be expected for the AR locus, since recombination is known to be lower towards the centromere, and the X chromosome only recombines in females. According to the Decode genetic map (Kong et al. 2002), the female recombination rate for the interval between the EDA2R and the AR (65.6–66.6 Mb on the NCBI 35 assembly) is 0.756 cM/Mb, which is below average. In order to interpret the possible influence of the local recombination rate on LD in a population, the female recombination rate must be translated into the sex-averaged rate (0.378 cM/Mb). This value is clearly at the lower end of the distribution of recombination rates for all 1 Mb windows across the human genome (Supplementary Fig. 5). A low recombination rate is thus one factor that is likely to have contributed to the extent of LD and haplotype homozygosity at the AR locus. The lack of formal significance for our REHH analysis, which corrects for local variation in the recombination rate, suggests that the high EHH might be influenced by the low recombination rate at this locus. Thus, low recombination rate certainly has contributed to the strong LD from EDA2R to AR, which applies in either the presence or the absence of recent positive selection.

**Discussion**

The present study was initiated in order to understand the putative signatures of recent positive selection at the human AR/EDA2R locus and their relationship to haplotypes that have been shown to be the major genetic determinant of AGA (Ellis et al. 2001; 2007; Hayes et al. 2005; Hillmer et al. 2005, 2008a; Richards et al. 2008). In the European population, the AGA-associated alleles are on the most frequent haplotype Eur-H1, which showed an unusually high EHH toward EDA2R. Although REHH was not extreme, the levels of EHH and REHH are clearly higher compared to the levels of the less frequent haplotypes in the European population. The high haplotype specific LD, which is indicative of a young haplotype with a high frequency, suggests a rapid increase in the frequency of Eur-H1.

The chromosomal region extending from AR to EDA2R displays several unusual features: (1) The interval of more than 1 Mb displays high LD in the European population. It has been previously shown that the average pairwise LD of the AR locus was the highest of all X-chromosomal windows with the same recombination rate (Hillmer et al. 2005). (2) The LD of the AR/EDA2R locus differs between the three continental populations, with comparatively low LD observed in Africans. (3) In East Asians, genetic variability in this region is remarkably low due to fixation of the derived allele of most SNPs. Since a locus that has experienced positive selection will subsequently show low levels of genetic variation (Kim and Stephan 2002) and an excess of high-frequency derived variants (Fay and Wu 2000), the pattern of variation at the AR/EDA2R locus in East Asia may be best explained by a selective sweep. When those SNPs which are non-polymorphic in East Asia are included in an LD analysis, the region displays high LD and the pattern is more similar to that of Europeans.

In order to explain the pattern of genetic variability observed at the AR locus, genetic drift needs to be considered as the null hypothesis. In fact, population range expansions can lead to the geographic spread of alleles and thereby to strong differences in allele frequencies between continents (Hofer et al. 2009; Klopfstein et al. 2006). It has been reported that East Asians generally have fewer rare derived alleles than Europeans, which has been interpreted as an indication of increased genetic drift in East Asians (Keinan et al. 2007). Furthermore, significantly larger $F_{ST}$ values have been observed between East Asian and Africans than between Europeans and Africans, with a more pronounced bottleneck in East Asians being the probable cause (Keinan et al. 2007). Compared to the pattern of the autosomes, the X chromosome shows a stronger genetic drift in North Europeans and East Asians and both populations harbor more high-frequency derived alleles on the X chromosome than expected (Keinan et al. 2009).

On other hand, the influence of positive selection on AR/EDA2R haplotype diversity gains further support from the evidence that the non-synonymous substitution R57K (rs1385699) in EDA2R has been subject to positive selection, with frequencies of the derived allele 57K (rs1385699 T) of 0% in Africa, 70% in Europe, and 100% in East Asia (Sabeti et al. 2007). The frequency of the haplotype carrying rs1385699 T (57K) and the AGA risk alleles on Eur-H1 is 0.71 in the European population (Table 1). Given the extent of LD between the two genes, it is possible that the AGA risk alleles rose to high frequency due to a hitchhiking event that was driven by selection for the 57K allele. Alternatively, the 57K allele may itself be considered as a candidate for conferring AGA susceptibility, since the ectodysplasin pathway is involved in hair development (Botchkarev and Fessing 2005; Fujimoto et al. 2008). EDAR (ectodysplasin-A receptor) is an ortholog of EDA2R and the derived amino acid 370A in EDAR results in enhanced NF-kB activity (Mou et al. 2008; Bryk et al. 2008), larger hair follicles and thicker hair fibers in mice (Mou et al. 2008).
and is associated with increased hair thickness in humans (Fujimoto et al. 2008; Mou et al. 2008). Since hair follicle miniaturization is a feature of AGA (Randall 2007), EDAR2R contribution to AGA development is plausible. Amino acid 57 is located in one of three cysteine-rich regions of the extracellular domain of EDAR2R which specifically binds the ligand EDA-A2 (ectodysplasin A isoform 2) (Yan et al. 2000). A possible functional role for R57K gains further support from the fact that the ancestral allele 57R is conserved between human, chimpanzee, mouse, and rat. As the 57K allele is strongly associated with AGA (Prodi et al. 2008), this SNP is a good candidate for the functional AGA variant at this locus. Interestingly, EDAR shows strong signatures of positive selection (Barreiro et al. 2008; Bryk et al. 2008; Myles et al. 2008). The hair phenotype as a driving force due to climate adaptation or sexual selection has been suggested and phenotypic hitchhiking on another trait (e.g. tooth morphology) has been considered. However, in contrast to EDAR deficiency, genetic EDAR2R ablation in mice is not accompanied by any visible skin or hair follicle abnormality (Newton et al. 2004). Further, our data showed stronger association for SNPs other than rs1385699 at the AR/EDAR2R locus (Table 1 and Hillmer et al. 2005, 2008a) and there appears to be no independent effect of AR and EDAR2R variants on AGA susceptibility (own unpublished data). The androgen signaling pathway is of great importance for the development of AGA since castration and complete androgen insensitivity syndrome inhibit the development of AGA (Hamilton 1960, Randall 2007). Both genes, AR and EDAR2R, are therefore good candidates for AGA. It seems likely that a regulatory variant between the two genes is of functional relevance for AGA, but it remains unclear whether AR, EDAR2R or both genes are involved.

Table 1 Association of rs1385699 and AR core haplotype SNPs with AGA and haplotype frequencies

| Haplotype freq. | rs1385699 | rs2497935 | rs962458 | rs12007229 | rs12396249 |
|-----------------|-----------|-----------|----------|------------|------------|
| 198 German affected versus 157 German unaffected (Hillmer et al. 2005) | | | | | |
| Association P with AGA | 9.41 × 10^(-5) | 3.17 × 10^(-11) | 2.56 × 10^(-6) | | |
| OR for AGA | 3 | 8.69 | 11.1 | | |
| OR (95% CI) | 1.71–5.32 | 4.23–17.87 | 3.27–37.74 | | |
| AGA risk allele | T | A | A | | |
| 296 German affected versus 347 German controls (Hillmer et al. 2008b) | | | | | |
| Association P with AGA | 1.6 × 10^(-4) | 4.61 × 10^(-6) | 0.00234 | 9.05 × 10^(-9) | |
| OR for AGA | 2.22 | 9.63 | 2.96 | 5.22 | |
| OR (95% CI) | 1.46–3.37 | 2.99–31.17 | 1.43–6.13 | 2.82–9.67 | |
| AGA risk allele | T | A | C | G | 0.703^b |
| European haplotype^c | | | | | |
| T | A | A | C | G | 0.711 |
| C | A | A | C | G | 0.078 |
| C | G | A | A | A | 0.078 |
| T | G | A | A | A | 0.078 |
| C | G | G | C | A | 0.033 |
| C | G | A | C | G | 0.022 |

^a Haplotype frequency in 188 male German blood donors
^b Haplotype frequency in 182 male German population controls
^c HapMap CEU data (Altshuler et al. 2005; Gibbs et al. 2003)
identified, the lower prevalence of AGA in Asia and presumably in Africa might indicate a higher frequency of the functional AGA allele in the European population, either due to its origin on a Europe-specific background or differences in population demography. Alleles of 21 SNPs between AR and EDA2R are more frequent in Europeans than in East Asians and Africans (Supplementary Table 1). Three of these have been tested for association with AGA in a German sample (Hillmer et al. 2008a) and the AGA risk allele of one SNP (rs471205) is more frequent in Europeans than in East Asians and Africans. Although the association of rs471205 was only moderate and a functional relevance for AGA development, therefore, seems unlikely, the other SNPs with this frequency pattern are good candidates for further investigation. Since the HapMap data does not cover all SNPs in this region, the presence of more SNPs showing the respective population differentiation is possible. Alternatively, it may be that the penetrance of the mutation is modified by genetic background or environmental factors. Genetic studies of AGA in non-European populations are warranted in order to address this question.

Importantly, the markers with the strongest population differences (as measured by average $F_{ST}$ values in the European/African and East Asian/African comparison) are found in the region extending from AR toward the centromere, including EDA2R. It is tempting to speculate that the proposed positive selection on 57K (Sabeti et al. 2007) is a driving force behind the population haplotype distribution centromeric to AR and that hitchhiking of the AGA susceptibility mutation on the 57K background accounts for the high prevalence of the phenotype. In either case, it seems safe to say that the mutation which confers risk for AGA in the European population rose to high frequency in combination with the derived allele 57K in EDA2R. 57K being the causative variant for AGA might be regarded as less likely as this allele is fixed in East Asia whereby the prevalence of AGA is lower in this population compared to Europeans. As the signature of positive selection is less profound in Europeans compared to East Asians, but AGA prevalence is higher in Europeans, it seems likely that the AGA causative variant and the cause for positive selection are distinct from each other. Resequencing of the AGA risk haplotype of the AR/EDA2R locus in different populations and subsequent association analysis are needed to clarify whether a European specific mutation accounts for the higher prevalence of AGA in this population.

Acknowledgments This work was supported by the European Research Grant for Research on Alopecia funded by Pfizer Consumer Healthcare, the Deutsche Forschungsgemeinschaft (FOR 423), and the Alfried Krupp von Bohlen und Halbach-Stiftung. We thank Christine Schmael for revision of the manuscript. The work by S.M., D.A.H., K.T., and M.S. was funded by the Max Planck Society.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

Altschuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P (2005) A haplotype map of the human genome. Nature 437:1299–1320

Bamshad M, Wooding SP (2003) Signatures of natural selection in the human genome. Nat Rev Genet 4:99–111

Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L (2008) Natural selection has driven population differentiation in modern humans. Nat Genet 40:340–345

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol 13:969–980

Botchkarev VA, Fessing MY (2005) Edar signaling in the control of hair follicle development. J Invest Dermatol Symp Proc 10:247–251

Bryk J, Hardouin E, Pugach I, Hughes D, Strutmann R, Stoneking M, Myles S (2008) Positive selection in East Asians for an EDAR allele that enhances NF-kappaB activation. PLoS ONE 3:e2209

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, Cobb JE, Zaloumis SG, Duncan AE, Harrap SB (2007) Baldness and the androgen receptor: the AR polyglycine repeat polymorphism does not confer susceptibility to androgenetic alopecia. Hum Genet 121:451–457

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413

Frazier KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Fan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Zhu Z, Shen D, Shi Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Sun W, Wang H, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallec C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T et al (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851–861

Freudenberg J, Fu YH, Pucek LJ (2007) Enrichment of HapMap recombination hotspot predictions among human nervous system genes: evidence for positive selection? Eur J Hum Genet 15:1071–1078

Fujimoto A, Kimura R, Ohashi J, Omi K, Yuluwilandaru R, Babatbara L, Mustofa MS, Samakkarn U, Settheetham-Ishida W, Ishida T, Morishita Y, Furusawa T, Nakazawa M, Ohtsuka R, Tokunaga K (2008) A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. Hum Mol Genet 17:835–843
EDA2R is associated with androgenetic alopecia. J Invest Dermatol 128:2268–2270
Rajender S, Singh L, Thangaraj K (2007) Phenotypic heterogeneity of mutations in androgen receptor gene. Asian J Androl 9:147–179
Randall VA (2007) Hormonal regulation of hair follicles exhibits a biological paradox. Semin Cell Dev Biol 18:274–285
Richards JB, Yuan X, Geller F, Waterworth D, Bataille V, Glass D, Song K, Waerbe G, Vollenweider P, Aben KK, Kiemeney LA, Walters B, Soranzo N, Thorsteindottir U, Kong A, Rafnar T, Deloukas P, Sulem P, Stefansson H, Stefansson K, Spector TD, Mooserv V (2008) Male-pattern baldness susceptibility locus at 20p11. Nat Genet 40:1282–1284
Ris-Stalpers C, Kuiper GG, Faber PW, Schweikert HU, van Rooij HC, Zegers ND, Hodgins MB, Degenhart HJ, Trapman J, Brinkmann AO (1990) Aberrant splicing of androgen receptor mRNA results in synthesis of a nonfunctional receptor protein in a patient with androgen insensitivity. Proc Natl Acad Sci USA 87:7866–7870
Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837
Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, Lander ES, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Sun W, Wang H, Wang Y, Xiong X, Xu L, Waye MM, Tsai SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Lebouef M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC et al (2007) Genome-wide detection and characterization of positive selection in human populations. Nature 449:913–918
Setty LR (1970) Hair patterns of scalp of white and Negro males. Am J Phys Anthropol 33:49–55
Tang PH, Chia HP, Cheong LL, Koh D (2000) A community study of male androgenetic alopecia in Bishan, Singapore. Singapore Med J 41:202–205
Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. PLoS Biol 4:e72
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370
Weiss JR, Moysich KB, Swede H (2005) Epidemiology of male breast cancer. Cancer Epidemiol Biomarkers Prev 14:20–26
Yan M, Wang LC, Hymowitz SG, Schilbach S, Lee J, Goddard A, de Vos AM, Gao WQ, Dixit VM (2000) Two-amino acid molecular switch in an epithelial morphogen that regulates binding to two distinct receptors. Science 290:523–527