

ARTICLE

Received 27 Oct 2016 | Accepted 23 Jan 2017 | Published 8 Mar 2017

DOI: 10.1038/ncomms14694

OPEN

# Meta-analysis identifies novel risk loci and yields systematic insights into the biology of male-pattern baldness

Stefanie Heilmann-Heimbach<sup>1,2,\*</sup>, Christine Herold<sup>3,\*</sup>, Lara M. Hochfeld<sup>1,2</sup>, Axel M. Hillmer<sup>4,5</sup>, Dale R. Nyholt<sup>6</sup>, Julian Hecker<sup>7</sup>, Asif Javed<sup>4</sup>, Elaine G.Y. Chew<sup>4</sup>, Sonali Pechlivanis<sup>8</sup>, Dmitriy Drichel<sup>9</sup>, Xiu Ting Heng<sup>4</sup>, Ricardo C.-H. del Rosario<sup>10</sup>, Heide L. Fier<sup>7,11</sup>, Ralf Paus<sup>12</sup>, Rico Rueedi<sup>13,14</sup>, Tessel E. Galesloot<sup>15</sup>, Susanne Moebus<sup>8</sup>, Thomas Anhalt<sup>1,2</sup>, Shyam Prabhakar<sup>10</sup>, Rui Li<sup>16</sup>, Stavroula Kanoni<sup>17</sup>, George Papanikolaou<sup>18</sup>, Zoltán Kutalik<sup>14,19</sup>, Panos Deloukas<sup>17,20</sup>, Michael P. Philpott<sup>21</sup>, Gérard Waeber<sup>22</sup>, Tim D. Spector<sup>23</sup>, Peter Vollenweider<sup>22</sup>, Lambertus A.L.M. Kiemeny<sup>15</sup>, George Dedoussis<sup>18</sup>, J. Brent Richards<sup>16,23</sup>, Michael Nothnagel<sup>9</sup>, Nicholas G. Martin<sup>24</sup>, Tim Becker<sup>25</sup>, David A. Hinds<sup>26,\*</sup> & Markus M. Nöthen<sup>1,2,\*</sup>

Male-pattern baldness (MPB) is a common and highly heritable trait characterized by androgen-dependent, progressive hair loss from the scalp. Here, we carry out the largest GWAS meta-analysis of MPB to date, comprising 10,846 early-onset cases and 11,672 controls from eight independent cohorts. We identify 63 MPB-associated loci ( $P < 5 \times 10^{-8}$ , METAL) of which 23 have not been reported previously. The 63 loci explain ~39% of the phenotypic variance in MPB and highlight several plausible candidate genes (*FGF5*, *IRF4*, *DKK2*) and pathways (melatonin signalling, adipogenesis) that are likely to be implicated in the key-pathophysiological features of MPB and may represent promising targets for the development of novel therapeutic options. The data provide molecular evidence that rather than being an isolated trait, MPB shares a substantial biological basis with numerous other human phenotypes and may deserve evaluation as an early prognostic marker, for example, for prostate cancer, sudden cardiac arrest and neurodegenerative disorders.

<sup>1</sup>Institute of Human Genetics, University of Bonn, 53127 Bonn, Germany. <sup>2</sup>Department of Genomics, Life & Brain Center, University of Bonn, 53127 Bonn, Germany. <sup>3</sup>German Center for Neurodegenerative Disorders, 53175 Bonn, Germany. <sup>4</sup>Cancer Therapeutics and Stratified Oncology, Genome Institute of Singapore, Singapore 138672, Singapore. <sup>5</sup>TRON-Translational Oncology at the University Medical Center of Johannes Gutenberg University, TRON gGmbH, 55131 Mainz, Germany. <sup>6</sup>Institute of Health and Biomedical Innovation, Queensland University of Technology, GPO Box 2434, Brisbane, Queensland 4001, Australia. <sup>7</sup>Institute of Genomic Mathematics, University of Bonn, 53127 Bonn, Germany. <sup>8</sup>Institute of Medical Informatics, Biometry and Epidemiology, Medical Faculty, University of Duisburg-Essen, 45122 Essen, Germany. <sup>9</sup>Cologne Center for Genomics, University of Cologne, 50931 Cologne, Germany. <sup>10</sup>Computational and Systems Biology, Genome Institute of Singapore, Singapore 138672, Singapore. <sup>11</sup>Department of Biostatistics, Harvard School T.H. Chan of Public Health, Boston, Massachusetts 02115, USA. <sup>12</sup>Dermatology Research Centre, Institute of Inflammation and Repair, University of Manchester, Manchester M13 9PL, UK. <sup>13</sup>Department of Computational Biology, University of Lausanne, 1011 Lausanne, Switzerland. <sup>14</sup>Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. <sup>15</sup>Radboud Institute for Health Sciences, Department for Health Evidence, Radboud University Medical Center, 6500 HB Nijmegen, The Netherlands. <sup>16</sup>Centre for Clinical Epidemiology, Departments of Medicine, Genetics and Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montreal, Quebec, H3T 1E2 Canada. <sup>17</sup>William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK. <sup>18</sup>School of Health Sciences and Education, Department of Nutrition and Dietetics, Harokopio University, Attica, 17671 Kallithea, Greece. <sup>19</sup>Institute of Social and Preventive Medicine, University Hospital of Lausanne, 1010 Lausanne, Switzerland. <sup>20</sup>Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah 21589, Saudi Arabia. <sup>21</sup>Centre for Cell Biology and Cutaneous Research Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AT, UK. <sup>22</sup>Department of Internal Medicine, Lausanne University Hospital (CHUV), 1011 Lausanne, Switzerland. <sup>23</sup>Department of Twin Research & Genetic Epidemiology, King's College London, London SE1 7EH, UK. <sup>24</sup>QIMR Berghofer Medical Research Institute, Brisbane, Queensland 4006, Australia. <sup>25</sup>Institute for Community Medicine, Ernst Moritz Arndt University Greifswald, 17475 Greifswald, Germany. <sup>26</sup>23andMe Inc., Mountain View, California 94040, USA. \* These authors contributed equally to this work. Correspondence and requests for materials should be addressed to S.H.-H. (email: sheilman@uni-bonn.de) or to M.M.N. (email: markus.noethen@uni-bonn.de).

**M**ale-pattern baldness (MPB) has a lifetime prevalence of ~80% in European men and is characterized by a characteristic pattern of progressive hair loss in distinct populations of androgen-dependent scalp hair follicles<sup>1</sup>. Early-onset MPB in particular can exert profound negative effects on quality of life<sup>2,3</sup>. While the pathobiology of MPB remains incompletely understood<sup>4,5</sup> and no causal therapy is available, current MBP treatments are of limited efficacy in many patients and even can have severe adverse effects<sup>6</sup>. Therefore, studies that clarify the biological underpinnings of key-pathophysiological features of MPB and enable the identification of novel molecular targets for more effective therapeutic intervention are needed. Here, molecular genetic studies that enable the identification of early causal events and allow differentiation between disease cause and consequence<sup>7</sup>, hold great potential and have already proven to be successful in the identification of contributing risk factors<sup>8–12</sup>. However, for the majority of the identified loci, few data are available concerning the manner on how these genetic variants contribute to (i) the strict androgen-dependency of the phenotype; (ii) the restriction of pathophysiological changes to hair follicles in the frontal and vertex scalp regions; (iii) characteristic changes in hair follicle cycling (anagen-shortening/premature catagen entry); and (iv) the gradual (mini-)organ transformation of pigmented terminal hair into unpigmented vellus hair follicles<sup>4</sup>. Given the high heritability of MPB ( $h^2 \sim 80\%$ ; refs 13,14), and the fact that a significant fraction of the overall heritable risk for MPB still awaits identification, large-scale genetic studies are an important tool to elucidate the molecular basis of MPB and to gain systematic insights into the underlying pathobiology.

Here, we report the results of the largest genome-wide association studies (GWAS) meta-analysis of MPB to date, that comprised a total of 22,518 individuals from eight independent GWAS samples of European descent. The analysis identifies 63 genome-wide significant loci that explain ~39% of the phenotypic variance in MPB. More than one-third of these loci ( $N = 23$ ) have not been reported previously. Our data highlight highly plausible candidate genes and pathways that are likely to contribute to key-pathophysiological characteristics of MPB such as the deregulation of anagen-to-catagen transition (*FGF5*, *EBF1*, *DKK2*, adipogenesis); increased androgen sensitivity (*SRD5A2*, melatonin signalling); and the transformation of pigmented terminal hair into unpigmented vellus hair (*IRF4*). Some of these genes and pathways may represent promising targets for the development of novel therapeutic options. In addition, our data provide molecular evidence that MPB shares a substantial biological basis with numerous other human phenotypes, which may have major implications in terms of the evaluation of MPB as an early prognostic marker for different phenotypes such as prostate cancer, sudden cardiac arrest or neurodegenerative disorders.

## Results

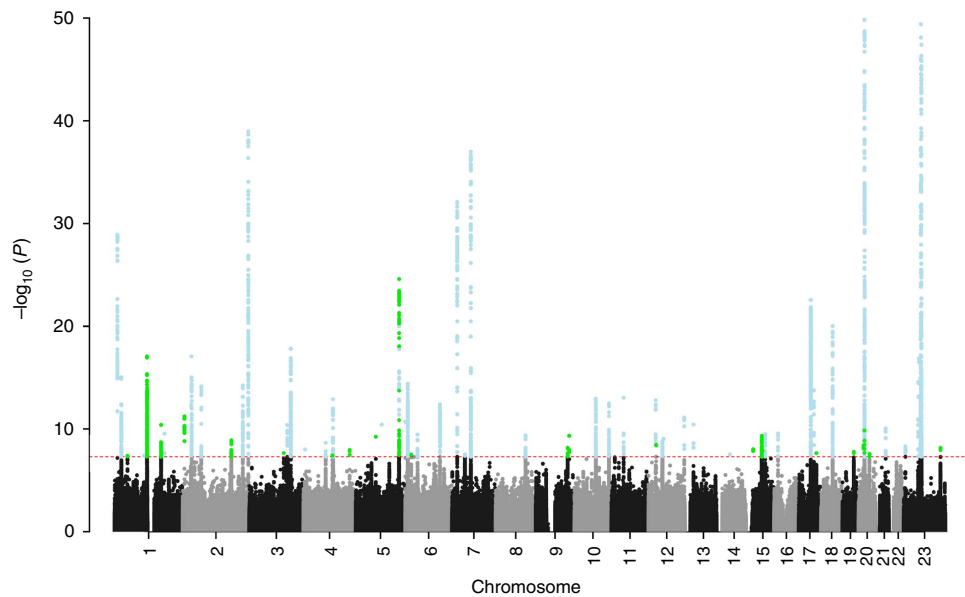
**Meta-analysis.** Our meta-analysis of eight independent GWAS samples of European descent (10,846 early-onset MPB cases, 11,672 controls; Supplementary Note 1; Supplementary Table 1) identified a total of 63 genome-wide significant MPB-risk loci ( $P < 5 \times 10^{-8}$ , METAL), which account for ~39% of the phenotypic variance in MPB (Fig. 1). These include 40 previously reported and 23 novel MPB-risk loci (Table 1; Supplementary Fig. 1). Regarding the genomic context of the 63 genome-wide significant risk loci, the majority of the association peaks ( $N = 45$ , 71%) mapped to intergenic regions. However, 60 (95%) of the association peaks were located within  $\leq 500$  kb of a protein coding gene, and a total of 18 (29%) association peaks mapped to

intronic or coding regions. Notably, six of the MPB-risk loci are located on the X-chromosome, and may thus contribute to the anecdotal similarity in hair status observed between men and their maternal grandfathers. Among them are the well-established *AR/EDA2R*-locus on Xq12, a locus near *FAM9A* and *FAM9B* on Xp22.31 and two loci near *KLF8* and *TRSD2* on Xp11.21. The remaining 57 loci are located on the autosomes. Heritability estimates on the liability scale were 0.14 ( $\pm 0.03$ ) for the X-chromosome, and 0.34 ( $\pm 0.12$ ) for the autosomes.

**Risk score analysis.** To evaluate the potential value of these association findings for the prediction of MPB risk, a weighted genotype-risk score for MPB was constructed from the lead SNPs of the 63 risk loci. The resulting score was divided into quartiles, and the risk for MPB was tested in each quartile, using the lowest quartile as a reference. As shown in Table 2, an increased risk for MPB was observed across all quartiles, with a substantially increased risk for MPB in quartile four (odds ratio (OR) = 4.16, 95% confidence interval (CI) = (2.03–8.55)). This effect was even stronger after adjustment for age (OR = 5.14, 95% CI = (2.44–10.86)), which underlines the strong age-dependency of the trait.

**Integration with mRNA and miRNA expression data.** To enable biological interpretation of the association findings and to pinpoint plausible candidate genes, the genetic data were integrated with own unpublished data on hair follicle mRNA-, and micro(mi) RNA-expression, an unpublished expression quantitative trait locus (eQTL) dataset from hair follicle (for details see Supplementary Materials and Methods) and two published eQTL data sets from skin and blood<sup>15,16</sup>. The comparison with eQTL data sets revealed a colocalization of MPB-risk variants with known regulatory variants ( $r^2 > 0.5$  between eQTL single nucleotide polymorphism (SNP) and MPB lead SNP) in 25 loci. For 10 of these 25 loci, previous studies have found an association between the MPB lead SNP itself and gene expression for example, of *CRHR1*, which encodes for a receptor for corticotropin releasing hormone (CRH), a known hair growth inhibitor and catagen inductor<sup>17</sup>, *FAM53B* a positive regulator of WNT/beta-catenin-signalling<sup>18</sup> and other genes, for example, *ANXA2*, *SUCNR1*, *WARS2* that have not yet been associated with hair biology (Supplementary Table 2). Detailed functional follow-up studies are now warranted to confirm these regulatory interactions, and to investigate the contribution of these candidate genes to the development of key MPB pathophysiological signs. A total of 19 association peaks (~30%) were located within  $\leq 500$  kb of a miRNA gene. Eighteen of these 34 miRNA-genes were expressed in human scalp hair follicles, and were predicted to target numerous mRNA genes at MPB-risk loci (Supplementary Table 3). Since miRNAs have been implicated in various aspects of hair biology, such as the control of hair follicle cycling, keratinocyte differentiation/proliferation and melanogenesis<sup>19</sup>, these miRNA-genes and their target genes may constitute plausible candidate genes at these MPB-associated loci.

**DEPICT analysis and enhancer enrichment.** The DEPICT analysis (Supplementary Tables 4–6) and literature search identified highly plausible candidate genes, such as *FGF5* at 4q21.21 (rs982804;  $P = 2.2 \times 10^{-9}$ , METAL) and *DKK2* at 4q25 (rs145945174;  $P = 1.3 \times 10^{-13}$ , METAL). *FGF5* plays an important role in the regulation of anagen-to-catagen transition and the control of human hair length<sup>20–22</sup>. *DKK2* encodes for a member of the family of dickkopf WNT-signalling inhibitors, which are reported to be secreted by dermal papilla cells (DPCs)



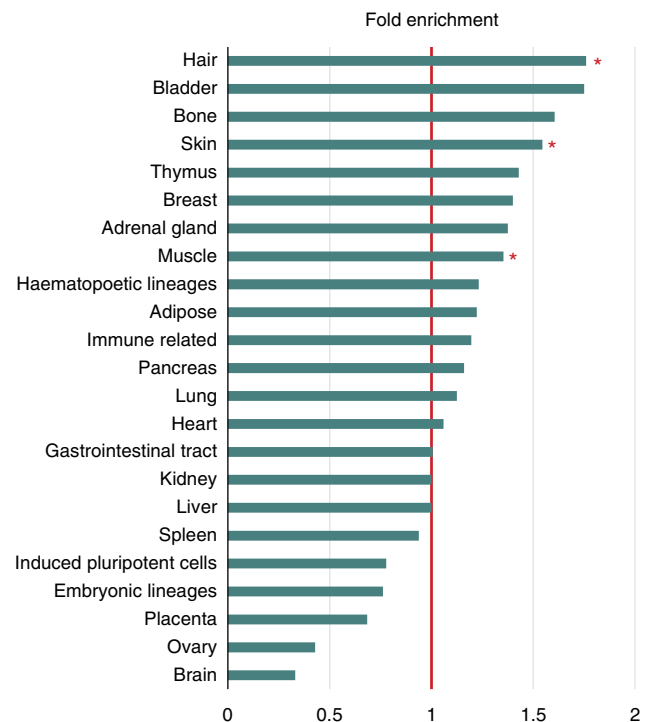
**Figure 1 | Manhattan plot of association results from the genome-wide meta-analysis.** Manhattan plot showing the results of the genome-wide association meta-analysis in 10,846 early-onset MPB cases and 11,672 controls. The x axis shows the chromosomal position, and the y axis shows the  $-\log_{10}(P)$  (METAL) of the association analysis. Previously reported genome-wide significant loci are depicted in blue; novel MPB-risk loci are depicted in green. Data points with  $P < 1 \times 10^{-50}$  (METAL) were truncated. The lowest  $P$ -value was  $1 \times 10^{-320}$  (METAL) at the *AR/EDA2R*-locus.

in response to androgens and to promote androgen-induced (premature) anagen-to-catagen transition<sup>23,24</sup>. Notably, our data indicated a nominally significant enrichment ( $P < 0.05$ , 1 million permutations) of credible MPB SNPs in enhancer regions from DPCs (treated with dihydrotestosterone, DHT), which further supports the hypothesis that DPCs are involved in MPB aetiology<sup>25</sup> (Fig. 2; Supplementary Data 1). An interesting focus for future research will be to map androgen receptor (AR) binding sites in DPCs and other hair-related cell types to test for enrichment of MPB-risk variants.

Moreover, the strongest MPB-risk allele at 6p25.3 (rs12203592-T;  $P = 2 \times 10^{-11}$ , METAL) is located within a melanocyte-specific enhancer element, and is reported to have a negative regulatory effect on *IRF4* expression. As *IRF4* contributes to the pigmentation of human skin, hair and eye<sup>2,26,27</sup>, this locus is likely to contribute to the gradual transformation of pigmented terminal hair into unpigmented vellus hair in MPB<sup>28</sup>. In geographical areas distant from the equator, less-pigmented skin, hair and eyes have been under positive selection, presumably due to the fact that this optimizes use of available ultraviolet radiation (UVR) for vitamin D3 generation<sup>29</sup>. Thus, the association between this functional *IRF4* variant and MPB may contribute to the relatively high prevalence of MPB in Europeans.

Another highly plausible candidate gene is located at 2p23.1. Here, the most strongly associated SNP (rs9282858;  $P = 8.9 \times 10^{-18}$ , METAL) represents a missense variant (c.145G > A; p.Ala49Thr) in *SRD5A2*. This gene encodes for the 5-alpha-reductase type II enzyme, which plays a critical role in androgen metabolism and MBP pathobiology<sup>30</sup>. Interestingly, elevated *SRD5A2* levels have been detected in MPB affected scalp, and the *SRD5A2* inhibitor finasteride is an effective therapy for MPB<sup>31,32</sup>. Additional candidate genes and their biological functions are summarized in Table 1 and Supplementary Data 2.

**Pathway-based analysis.** Pathway-based analysis of genes at MPB-risk loci revealed significant enrichment ( $P$ -value of the right-tailed Fisher's exact test  $< 0.05$ ) in 37 pathways, including the previously implicated androgen metabolism and WNT-



**Figure 2 | Enrichment of credible SNPs in tissue enhancers.** Barplot showing enrichment of MPB credible SNPs across 23 tissue specific enhancer sets. A total of 144 cell lineages and tissues were grouped in 23 categories for this evaluation. The expected number of credible SNPs and the significance of enrichment in each category were estimated in one million enhancer matched permutations, using the rest of the tissue types to define the candidate null set (Supplementary Data 2). Red asterisks indicate tissue types that contain individual cell lineages or tissues that show an enrichment of credible SNPs with an empirical  $P$ -value of  $< 0.05$ : non-balding DPCs treated with DHT (hair); one of three foreskin fibroblasts (skin); and one of three psoas muscle lineages.

signalling pathways (Supplementary Table 7). In addition to these known pathways, our data provide the first genetic evidence for the involvement of additional hormonal pathways such as epidermal growth factor (EGF)-signalling and oestrogen biosynthesis in MPB<sup>5,8,33–36</sup>. Enrichment was also found in pathways that, to our knowledge have not been associated previously with the key-pathophysiological signs of MPB such as melatonin signalling/degradation, adipogenesis and immune related pathways. Notably, both melatonin degradation and adipogenesis are reported to interact with, and to be controlled by, sex hormones<sup>37,38</sup> and decreasing melatonin levels are a potential marker of puberty progression<sup>39</sup>. These findings highlight the indispensability of sex hormones to the MPB phenotype<sup>40</sup>, and point to a novel potential link between intrafollicular melatonin synthesis and its recognized effects on oestrogen receptor expression and thus the sensitivity of human scalp hair follicles to stimulation with oestrogens, which effectively counteract the anagen-shortening effects of DHT in MPB development<sup>36,41</sup>.

Regarding adipogenesis, maturation of adipocyte precursors in the skin is reported to occur in parallel with the activation of hair follicle stem cells and to drive anagen induction and hair growth during synchronized hair regeneration in mice<sup>42</sup>. However, few data are available on whether and in what manner adipogenesis in the skin may impact hair follicle cycling in humans. Notably, adipogenesis and follicular stem cell activation are impaired in mice lacking the gene *EBF1*, which is located at another MPB-risk locus on 5q33.3, which suggest that adipogenesis plays a role in healthy hair cycle regulation and MPB development in humans. Regarding immunological processes, no conclusive data on their role in MPB aetiology are yet available<sup>43</sup>. However, skin resident macrophages contribute to the cyclic activation of adult hair follicles via induction of WNT- and FGF5-signalling, which were also implicated in the present meta-analysis<sup>44,45</sup>, and perifollicular inflammatory infiltrates have long been suspected to participate in the terminal-to-vellus transformation in MBP<sup>46,47</sup>. Thus, our data represent the first genetic evidence that cells residing in the immediate hair follicle micro-environment impact on MPB development and may warrant investigation as therapeutic targets.

**Overlap with other human traits.** On an epidemiological level, early-onset MPB has been associated with several severe late-onset somatic disorders, such as cardiovascular disease (CVD)<sup>48</sup>; prostate hyperplasia and cancer<sup>49–51</sup>; Parkinson's disease<sup>10</sup>; and amyotrophic lateral sclerosis<sup>52</sup>. We therefore investigated a possible genetic overlap between MPB and other phenotypes, and compared our data to reported GWAS signals from the NHGRI GWAS catalogue. The results are summarized in Supplementary Data 3. A total of 124 GWAS catalogue entries mapped to MPB-risk loci ( $r^2 \geq 0.3$  and/or  $D' > 0.8$ ). These included the above mentioned associations with hormone-dependent traits and a reduced pigmentation of facial skin, scalp hair and eyes. As regards to the well-established associations between MPB and CVD and prostate cancer, a total of seven overlapping associations were identified. These were found between MPB and: (i) blood pressure ( $N = 3$ ); (ii) QT-interval length ( $N = 1$ ); (iii) atrial fibrillation ( $N = 1$ ); (iv) sudden cardiac arrest ( $N = 1$ ); (v) and prostate cancer ( $N = 1$ ). Here, our analysis confirmed the positive epidemiological association between prostate cancer and MPB at Xq12 (*AR/EDA2R*-locus), pointing towards a shared pathophysiological mechanism that may involve EDA2R-signalling and AR-transactivation<sup>53</sup>. Surprisingly, for the majority of CVD GWAS SNPs, the direction of effect for MPB and CVD differed, thus opposing the reported positive association between MPB and CVD at an epidemiological level.

This was not the case for the positive associations between MPB and diastolic blood pressure at 4q21.21, and sudden cardiac arrest at 12q13.12. Here, further analyses are warranted to elucidate the exact underlying genes and biological pathways, and how they relate to the epidemiological findings. Notably, while the 4q21.21 locus pointed towards a positive association between MPB and diastolic blood pressure levels, opposite effect direction were observed for a second overlapping association between these traits on 5q33.3. This indicates that the effect direction of the genetic correlation between two complex traits may differ between individual loci or pathways. This finding highlights the need for systematic studies to assess not only the quantitative genetic overlap but also individual overlapping genetic factors and the underlying genes and pathways. In addition, associations with four loci were found for MPB and lower body height, which may be driven by an accelerated progression of puberty and premature induction of epiphyseal closure<sup>54</sup>. MPB-risk alleles at 17q21.31 and 6q22.32 were associated with increased bone mineral density, which may be a consequence of optimized UVR-induced vitamin D synthesis in subjects with MPB. This is consistent with the observation that MPB-associated alleles confer a reduced risk for immune related phenotypes, such as type 1 diabetes; multiple sclerosis (MS); and rheumatoid arthritis (Supplementary Data 3). An increased incidence of these diseases has been reported in subjects with poor vitamin D intake and low serum vitamin D levels<sup>55</sup>. Furthermore, a recent Mendelian-randomization study by Mokry *et al.*<sup>56</sup>, found an association between genetically lowered 25-hydroxyvitamin D levels and increased susceptibility to MS. Moreover, our data indicate shared genetic determinants for MPB and a reduced risk for ovarian cancer, colorectal cancer, and chronic lymphatic leukaemia, as well as overlapping associations with progressive supranuclear palsy and decreased intracranial volume. The indispensability of sex hormones for the MPB phenotype is supported by the identification of overlapping genetic association between MPB and other hormone-dependent traits, such as an earlier age-at-onset of menarche in females and earlier sexual maturation and higher serum androgen levels in males (6q22.32, 16p13.12, Xp22.31)<sup>12,57–60</sup>.

## Discussion

The present genome-wide meta-analysis identified 63 risk loci for MPB, and highlights highly plausible candidate genes that are likely to be implicated in the key-pathophysiological features of MPB, such as deregulation of anagen-to-catagen transition (*FGF5*, *EBF1*, *DKK2*, adipogenesis); increased androgen sensitivity (*SRD5A2*, melatonin signalling); and transformation of pigmented terminal hair into unpigmented vellus hair (*IRF4*). As demonstrated for *SRD5A2*, the GWAS approach is a valuable tool for drug target identification and the newly identified candidate genes and pathways, perhaps most notably *FGF5* and melatonin signalling are promising targets for the development of novel therapeutic options for MPB. In addition to confirming the involvement of well-established pathways that control hormonal status (androgen metabolism, oestrogen signalling) and hair follicle cycling (WNT-signalling, EGF-signalling), our data support the importance of less-well studied biological contexts, such as the involvement of perifollicular macrophages and adipocytes.

Moreover, our data provide molecular evidence that rather than being an isolated trait, MPB shares a substantial biological basis with numerous other human phenotypes. This may have major implications in terms of the evaluation of MPB as an early prognostic marker for different phenotypes such as prostate cancer, sudden cardiac arrest or neurodegenerative disorders, and for the repurposing of existing drugs for use in MPB therapy. The latter is illustrated by the fact that the efficacy of the two U.S. Food and Drug Administration (FDA) approved drugs, minoxidil

**Table 1 | Summary results for the lead SNPs for the 63 genome-wide significant MPB-risk loci.**

Chromosomal region	SNP	Chr:Start-Stop	EA/OA	EAF	P-value	$\beta$	s.e.m.	Genomic context	Plausible candidate genes	Previously reported with genome-wide significance
1p36.22	rs2095921	1:11,032,215-11,050,962	G/C	0.23	$1.24 \times 10^{-29}$	-0.27	0.02	[C1orf127], Intron	<b>ANGPTL7<sup>EF</sup>, SRM<sup>EF</sup>, RUNX3<sup>DEF</sup>, SYF2<sup>EN</sup>, TMEM50A<sup>EQ</sup>, CMPK1<sup>EQ</sup></b>	4, 6
1p36.11	rs7534070	1:25,318,225-25,511,358	T/G	0.35	$9.21 \times 10^{-16}$	-0.20	0.03	MIR4425-[-]-SYF2		6
1p33*	rs17371253	1:47,945,306-47,957,081	T/C	0.42	$4.19 \times 10^{-8}$	0.12	0.02	FOXD2-[-]-TRABD2B		6
1p12	rs12083887	1:118,849,762-118,983,663	G/A	0.39	$8.65 \times 10^{-18}$	-0.18	0.02	SPAG17-[-]-		6
1p12	rs17185996	1:119,203,617-119,789,289	T/C	0.13	$9.96 \times 10^{-15}$	0.23	0.03	[WARS2], Intron	<b>TBX15<sup>DF</sup>, WARS2<sup>EN</sup>, GORAB<sup>FN</sup>, GORAB<sup>FQ</sup></b>	6
1q24.2	rs2421326	1:170,229,218-170,546,697	T/A	0.38	$3.84 \times 10^{-11}$	0.14	0.02	[GORAB], Intron		6
1q24.2*	rs2206310	1:170,712,490-170,791,811	G/A	0.37	$3.71 \times 10^{-8}$	0.12	0.02	PRRX1-[-]-MROH9		6
2p25.2	rs12997617	2:6,458,933-6,573,741	T/C	0.14	$5.82 \times 10^{-12}$	-0.20	0.03	MIR7158 -[-]- MIR7515HG		6
2p23.1	rs9282858	2:31,609,942-33,040,343	T/C	0.03	$8.85 \times 10^{-18}$	-0.51	0.06	[SRD5A2], Exon	<b>SRD5A2<sup>DF</sup>, MEMO<sup>EF</sup></b>	6
2p14	rs6546334	2:68,075,994-68,165,380	T/C	0.35	$7.28 \times 10^{-15}$	-0.16	0.02	ETAA1 -[-]-CID		6
2q31.1	rs1819008	2:177,588,381-177,796,823	T/C	0.46	$1.26 \times 10^{-9}$	-0.12	0.02	MIR1246-[-]- HNRNPA3	<b>NFE2L2<sup>EF</sup></b>	6
2q35	rs74333950	2:219,691,566-219,854,317	T/G	0.14	$5.74 \times 10^{-15}$	-0.24	0.03	[WNT10A], Intron	<b>CYP27A1<sup>EQ</sup>, WNT10A<sup>EF</sup>, TWIST2<sup>FN</sup>, HDAC4<sup>EF</sup></b>	5, 6
2q37.3	rs11684254	2:239,638,300-239,764,032	G/C	0.34	$1.08 \times 10^{-39}$	0.29	0.02	ASB1-[-]-TWIST2		4, 6
3q21.3*	rs9850626	3:126,066,714-126,095,012	T/G	0.24	$2.23 \times 10^{-8}$	-0.13	0.02	[KLF15], Intron		6
3q23	rs10935316	3:139,020,298-139,032,333	G/C	0.20	$4.09 \times 10^{-11}$	0.17	0.03	PISR71-[-]-MRPS22	<b>MRPS22<sup>EN</sup></b>	6
3q25.1	rs4679956	3:151,599,393-151,781,175	T/C	0.41	$1.50 \times 10^{-18}$	0.18	0.02	SUCNR1-[-]-MBNL1		5, 6
4q21.21	rs982804	4:81,170,173-81,207,039	T/C	0.42	$2.20 \times 10^{-9}$	-0.13	0.02	[FGF5], Intron	<b>PRDM8<sup>EQ</sup>, FGF5<sup>FN</sup></b>	6
4q24*	rs142756290	4:105,851,807-105,903,045	G/C	0.07	$3.81 \times 10^{-8}$	0.22	0.04	CXXC4-[-]-TET2	<b>TET2<sup>FN</sup></b>	6
4q25	rs145945174	4:107,507,213-107,888,788	G/A	0.07	$1.26 \times 10^{-13}$	0.30	0.04	[DKK2], Intron	<b>DKK2<sup>FN</sup></b>	6
4q32.3	rs17053607	4:168,717,680-168,789,572	G/A	0.05	$1.06 \times 10^{-8}$	0.26	0.05	[-]-ANXA10	<b>ANXA10<sup>DN</sup></b>	6
5q13.3	rs141577316	5:73,425,638-73,479,375	T/C	0.02	$5.71 \times 10^{-10}$	-0.58	0.09	ARHGGEF28-[-]-ENC1		6
5q33.3*	rs6556350	5:157,844,535-157,880,002	T/C	0.47	$3.17 \times 10^{-08}$	0.11	0.02	[-]-EBF1	<b>EBF1<sup>DF</sup></b>	6
5q33.3	rs17643057	5:158,140,780-158,499,367	G/A	0.40	$2.58 \times 10^{-25}$	-0.21	0.02	[EBF1], Intron	<b>EBF1<sup>DF</sup>, RNF145<sup>EQ</sup></b>	5, 6
6p25.3	rs12203592	6:396,321-421,281	T/C	0.17	$2.00 \times 10^{-11}$	0.20	0.03	[IRF4], Intron	<b>IRF4<sup>DF</sup>, EXOC2<sup>EF</sup></b>	6
6p24.3	rs9380830	6:9,292,574-9,699,467	G/A	0.05	$3.88 \times 10^{-15}$	-0.36	0.05			6
6p22.3*	rs34061913	6:21,835,656-21,926,313	T/-	0.10	$3.22 \times 10^{-8}$	-0.20	0.04	SOX4-[-]-NBA1	<b>SOX4<sup>EF</sup></b>	6
6p21.1	rs4714811	6:44,639,723-44,676,420	T/C	0.19	$3.43 \times 10^{-10}$	-0.16	0.03	CDCSL-[-]-SUPT3H	<b>CDCSL<sup>EN</sup>, SUPT3H<sup>NQ</sup></b>	6
6q22.32	rs144578168	6:126,609,124-127,137,371	AAAAT/-	0.49	$3.91 \times 10^{-13}$	-0.15	0.02	MIRS88-[-]-RSPO3		6
7p21.1	rs6461387	7:18,765,520-18,933,411	G/C	0.40	$7.92 \times 10^{-33}$	-0.25	0.02	[HDAC9] Intron	<b>HDAC9<sup>EF</sup>, TWIST1<sup>F</sup></b>	3, 4, 6
7p12.3*	rs12702262	7:46,846,164-46,948,331	G/C	0.24	$2.94 \times 10^{-8}$	-0.14	0.02	[-]-TNS3	<b>TNS3<sup>EN</sup></b>	6
7q11.22	rs6952233	7:68,566,391-68,918,814	G/A	0.46	$1.04 \times 10^{-37}$	-0.26	0.02	[-]-AUTS2	<b>AUTS2<sup>EN</sup></b>	4, 6
8q23.1	rs74321310	8:108,847,459-109,469,932	G/A	0.01	$4.25 \times 10^{-10}$	0.65	0.11	E1F3E-[-]-EMC2	<b>E1F3E<sup>EN</sup></b>	6
9q32	rs201655553	9:116,511,364-116,691,072	AAAG/-	0.33	$6.66 \times 10^{-9}$	0.14	0.02	RGS3-[-]-ZNF618	<b>ZNF618<sup>EN</sup></b>	6
9q33.1	rs2416699	9:122,192,838-122,197,736	T/C	0.13	$4.53 \times 10^{-10}$	0.18	0.03	BRINP1-[-]-		6
10q22.3	rs1907352	10:78,168,310-78,638,957	G/C	0.41	$1.08 \times 10^{-13}$	0.15	0.02	[C10orf11], Intron	<b>C10orf11<sup>FN</sup></b>	6
10q26.13	rs3781458	10:126,278,648-126,537,000	T/C	0.38	$3.31 \times 10^{-13}$	-0.15	0.02	[FAMS3B], Intron	<b>OAT<sup>EF</sup>, LHPP<sup>EQ</sup>, FAMS3B<sup>EN</sup>, METTL10<sup>EQ</sup></b>	6
11p11.2	rs11037975	11:44,384,928-44,449,006	G/C	0.30	$8.99 \times 10^{-14}$	-0.17	0.02	ALX4-[-]-CD82	<b>ALX4<sup>FN</sup></b>	6
12p12.1	rs9300169	12:26,438,616-26,446,289	G/A	0.27	$1.59 \times 10^{-13}$	-0.17	0.02	SSPN-[-]-ITPR2	<b>SSPN<sup>FN</sup></b>	5, 6
12p11.22	rs10843026	12:27,986,524-28,102,722	G/A	0.05	$3.37 \times 10^{-9}$	0.30	0.05	KLHL42-[-]-PTHLH	<b>PTHLH<sup>EF</sup></b>	6
12q13.12	rs4307773	12:50,878,227-51,319,795	T/C	0.41	$1.47 \times 10^{-9}$	0.12	0.02	DIP2B-[-]-ATF1	<b>LIM1<sup>EF</sup>, DIP2B<sup>EN</sup>, ATF1<sup>DE</sup>, METTL7A<sup>EQ</sup>, SLC11A2<sup>EQ</sup>, TFCP2<sup>EQ</sup></b>	6
12q24.33	rs78472887	12:130,562,418-130,593,594	G/C	0.14	$7.65 \times 10^{-12}$	0.20	0.03	TMEM132D-[-]- FZD10	<b>FZD10<sup>F</sup></b>	6
13q12.3	rs7997640	13:30,741,949-30,751,681	C/A	0.43	$3.61 \times 10^{-11}$	-0.14	0.02	UBL3-[-]-KATNAL1	<b>KATNAL1<sup>FN</sup></b>	6
15q12	rs4906800	15:26,120,550-26,121,173	T/A	0.07	$9.29 \times 10^{-9}$	0.24	0.04	MIR4715-[-]-	<b>MIR4715<sup>N</sup></b>	6
15q21.3	rs7173477	15:57,079,417-57,583,301	C/A	0.18	$4.60 \times 10^{-10}$	-0.16	0.03	[TCF12], Intron	<b>TCF12<sup>EN</sup></b>	6
15q22.2	rs17270216	15:60,836,794-60,842,258	G/A	0.22	$1.38 \times 10^{-8}$	0.14	0.02	[RORA], Intron	<b>ANXA2<sup>EQ</sup>, RORA<sup>DF</sup>, RORA<sup>FN</sup></b>	6
15q23	rs7182742	15:69,992,647-70,048,157	T/A	0.47	$3.20 \times 10^{-10}$	-0.12	0.02	PCAT29-[-]-TLE3		6
16p13.12	rs246185	16:14,377,400-14,406,119	T/C	0.33	$2.65 \times 10^{-10}$	0.14	0.02	MKL2-[-]-MIR193B	<b>PLA2G10<sup>F</sup></b>	6
17q21.31	rs112550936	17:43,292,035-44,877,887	G/A	0.18	$2.72 \times 10^{-23}$	-0.29	0.03	PLEKHM1-[-]-CRHR1	<b>ARHGAP27<sup>EQ</sup>, CRHR1- ITP<sup>EQ</sup>, CRHR1<sup>FQ</sup></b>	4, 6
17q22	rs62060349	17:55,226,801-55,237,542	T/C	0.45	$1.64 \times 10^{-14}$	-0.16	0.02	AKAP1-[-]-MSI2	<b>AKAP1<sup>FN</sup></b>	6
17q24.2	rs79806428	17:64,251,733-64,278,885	T/C	0.07	$2.26 \times 10^{-8}$	-0.21	0.04	APOH-[-]-PRKCA	<b>PRKCA<sup>EF</sup></b>	6
18q12.3	rs12606816	18:42,669,587-42,829,042	G/A	0.27	$9.36 \times 10^{-21}$	-0.21	0.02	[SLC14A2], Intron		4, 6
19q13.2	rs17318596	19:41,885,151-41,946,095	G/A	0.37	$1.67 \times 10^{-8}$	0.12	0.02	B3GNT8-[-]-ATPSSL	<b>TGFB1<sup>F</sup>, B9D2<sup>EF</sup>, TMEM91<sup>EQ</sup>, EXOSC5<sup>EQ</sup>, BCKDHA<sup>EQ</sup>, B3GNT8<sup>DE</sup>, ATPSSL<sup>EN</sup>, DSTN<sup>EQ</sup>, RRBP1<sup>EQ</sup>, OVOL2<sup>EF</sup>, RN2<sup>EQ</sup></b>	6
20p12.2	rs733626	20:17,681,551-17,696,340	T/G	0.14	$4.14 \times 10^{-9}$	0.17	0.03	[BANF2], Intron		6
20p11.22	rs148541866	20:21,365,463-21,517,472	T/G	0.01	$1.38 \times 10^{-10}$	-0.58	0.09	NKX2-4-[-]-NKX2-2		1, 2, 3, 4, 6
20p11.22	rs6113492	20:21,655,888-22,375,754	G/C	0.43	$1.23 \times 10^{-162}$	0.51	0.02	PAX1-[-]-		6
20q12*	rs6072223	20:39,616,593-39,647,516	T/A	0.16	$2.63 \times 10^{-8}$	0.16	0.03	MAFB-[-]-TOP1	<b>TOP1<sup>DE</sup>, ZHX3<sup>EQ</sup>, LPIN3<sup>EQ</sup>, RUNX1<sup>DE</sup>, RUNX1<sup>EF</sup></b>	6
21q22.12	rs7280071	21:36,208,167-36,238,517	T/A	0.23	$9.19 \times 10^{-11}$	-0.17	0.03	[RUNX1], Intron		6
Xp22.31	rs5933688	X:8,880,680-8,917,206	G/A	0.29	$4.77 \times 10^{-9}$	0.10	0.02	FAM9A-[-]-FAM9B		6
Xp11.21	rs185597083	X:54,875,521-55,133,845	C/A	0.04	$1.15 \times 10^{-11}$	0.30	0.04	PAGE2-[-]-FAM104B	<b>TSR2<sup>EF</sup></b>	6
Xp11.21	rs143212632	X:55,775,616-56,492,197	G/A	0.04	$2.58 \times 10^{-15}$	0.38	0.05	RRAGB-[-]-KLF8	<b>KLF8<sup>FN</sup></b>	6
Xp11.1	rs150535953	X:57,199,041-58,352,816	G/A	0.04	$1.39 \times 10^{-17}$	-0.41	0.05	ZXDA-[-]-		6
Xq12	rs2497911	X:62,340,191-67,813,682	C/A	0.19	$1 \times 10^{-320}$	0.80	0.02	[-]-AR	<b>AR<sup>EF</sup></b>	1, 2, 3, 4, 6
Xq26.3	rs4829906	X:136,618,815-136,957,253	G/A	0.40	$6.98 \times 10^{-9}$	0.09	0.02	ZIC3-[-]-		6

D, evidence from DEPICT ( $P < 0.05$ ); E, expressed in human hair follicle; EAF, effect allele frequency; EA, effect allele; F, functional evidence from literature; MPB, male-pattern baldness; N, neighbouring gene; Q-eQTL; OA, other allele; P-value, association P-value of the meta-analysis;  $\beta$ -effect size; SNP, single nucleotide polymorphism; plausible genes within 500 kb of lead SNP are printed in bold. \*Region where the lead SNP showed an association with  $5 \times 10^{-8} < P < 1.3 \times 10^{-7}$  after double GC correction with the residual inflation factors from LD score regression analysis. <sup>1</sup>Hillmer et al.<sup>9</sup>; <sup>2</sup>Richards et al.<sup>11</sup>; <sup>3</sup>Brockschmidt et al., 2011<sup>7</sup>; <sup>4</sup>Li et al.<sup>10</sup>; <sup>5</sup>Heilmann et al.<sup>8,64</sup>; <sup>6</sup>Pickrell et al.<sup>12</sup>

**Table 2 | Association between the weighted genotype-risk score and MPB.**

	Model I		Model II	
	OR	95% CI	OR	95% CI
Quartile 1	1.00 (reference)		1.00 (reference)	
Quartile 2	1.547	0.90–2.65	1.87	1.06–3.32
Quartile 3	1.796	1.03–3.13	2.12	1.17–3.83
Quartile 4	4.16	2.03–8.55	5.14	2.45–10.86

95% CI, 95% confidence interval; Model I, weighted genotype-risk score; Model II, age-adjusted weighted genotype-risk score; MPB, male-pattern baldness; OR, odds ratio.

and finasteride in MPB, was a serendipitous finding in patients administered these medications for hypertension and prostate hyperplasia. Finally, these novel insights into the genetic basis of MPB and its association with other traits may help to elucidate the evolutionary forces responsible for the relatively high prevalence of MPB in the European population.

## Methods

**Study participants.** Participants were drawn from eight independent genome-wide association studies early-onset MPB samples: 23andMe (9,009 cases; 8,491 controls); Bonn (581 cases; 416 controls); CoLaus (622 cases; 655 controls); Nijmegen Biomedical Study (145 cases; 247 controls); QIMRB1 (216 cases; 1,162 controls); QIMRB2 (59 cases; 498 controls); THISEAS (52 cases; 150 controls); and TwinsUK (163 cases; 210 controls). A detailed description of the studies and their phenotype definitions is provided in Supplementary Note 1. All eight studies were approved by the respective institutional ethics review committees (specified in Supplementary Note 1), and written informed consent was obtained from all participants prior to inclusion.

**Genome-wide association analyses.** A summary of the study specific genotyping platforms, imputation methods and GWAS is provided in Supplementary Table 1. For each of the eight studies fulfillment of the following two criteria was required: (i) a minor allele frequency (MAF) of > 1%; and (ii) a call rate of > 98%, a variance ratio of  $\geq 0.3$  (MACH) or a proper info statistic of  $\geq 0.4$  (IMPUTE2). Imputed data were analysed using logistic regression and the dosage data options of either PLINK or SNPTEST.

**Meta-analysis.** The present meta-analysis was performed in accordance with the GWAS meta-analysis standards outlined in de Bakker *et al.*<sup>61</sup> A fixed effects model was used to combine the logistic regression effect estimates of individual studies into a joint estimate, as implemented in METAL<sup>62</sup>. Results were cross-validated using the respective implementations in METAINTER and YAMAS. The Cochran's  $Q$  statistics and the  $I^2$  measurement were used to test for cross-study heterogeneity.  $P$ -values for the test of heterogeneity were calculated according to Higgins *et al.*<sup>63</sup> The meta-analysis included a total of  $N = 8,004,650$  SNPs that were available in the 23andMe cohort and (i) at least four additional studies for non-X-chromosomal SNPs, and (ii) three additional studies for X-chromosomal SNPs. The QQ-plot of the meta-analysis and a plot of the across-study homogeneity are shown in Supplementary Figs 2 and 3.

**Identification of independent risk loci.** SNP associations with a  $P$ -value of  $< 10^{-6}$  (METAL) were extracted from the meta-analysis and filtered for a heterogeneity  $P$ -value ( $P_{\text{Het}}$ ) of  $< 0.01$ . SNPs separated by a distances of  $\leq 100$  kb on the autosomes and  $\leq 500$  kb on the X-chromosome were assigned to the same genomic region. Loci were considered to be significantly associated with MPB if at least one SNP within the defined region showed an association to MPB at  $P < 5 \times 10^{-8}$  (METAL) and (i) had a MAF of  $\geq 0.05$ ; or (ii) was supported by at least one SNP in linkage disequilibrium (LD;  $r^2 > 0.5$ ), with a MAF of  $\geq 0.05$  and an association with MPB at  $P < 10^{-6}$  (METAL). MPB-risk loci were considered to be independent if the lead SNPs of the regions showed an LD of  $r^2 < 0.2$ .

**Test for polygenicity.** For the meta-analysis overall, an inflation factor of  $\lambda = 1.303$  was observed. After removal of LD-SNPs surrounding the two major loci (chr20, chrX), the inflation decreased to  $\lambda = 1.20$ . After exclusion of all 63 associated regions, the inflation decreased to  $\lambda = 1.16$ . This indicates that a relevant proportion of the observed inflation is driven by genome-wide significant associations. Furthermore, it is reasonable to assume from this finding that the remaining inflation is driven primarily by true genetic association. This is consistent with previous reports of a polygenic contribution to MPB, and with observations from meta-analyses of other complex genetic traits in which large

numbers of common risk factors were identified<sup>164,65</sup>. To determine whether the inflation was attributable to polygenicity or to confounding factors, the LD score regression method<sup>66</sup> was applied using LD Score v1.0.0. The analysis confirmed that most of the observed inflation is because of polygenicity (63.5%). The residual genomic inflation is  $\lambda_{\text{GC}} = 1.09$ , indicating that the observed association findings are not due to population stratification.

To estimate the proportion of potentially-false positive findings, association  $P$ -values (logistic regression) and effect directions for the 63 lead SNPs were compared between the 23andMe cohort (23andMe) and a meta-analysis of the remaining seven cohorts (MAwo23andMe) (Supplementary Table 8). For 62 of the 63 loci (98%) the analysis revealed a consistency of effect directions between 23andMe and MAwo23andMe. Of these loci, 76% (47/62) showed at least nominal significant association to MPB in the MAwo23andMe study (METAL), with 72% (34/47) also achieving significant association at  $P < 0.05/63 = 7.9 \times 10^{-4}$  (METAL  $P$ -value Bonferroni corrected for number of risk loci). Lastly, in contrast to the self-report based phenotyping used in the 23andMe cohort, phenotyping in the Bonn sample (BN) was based on clinical assessment. Thus for true positive association findings, stronger effect sizes would be expected on average in the BN sample. This was indeed the case: for 52 out of 61 lead SNPs available in the BN sample (85%) which would be highly unlikely if these loci included a relevant proportion of false positives. On the basis of these results, no correction was made for the residual inflation. The level of significance of the 63 MPB lead SNPs after double GC correction with the residual inflation factors from the LD score regression analyses is however indicated in Table 1.

**Definition of credible sets of SNPs.** For each risk locus, a credible set of SNPs was defined, as reported in Ripke *et al.*<sup>65</sup> (Supplementary Table 9). In brief, here the connection between the deviance and the single-variant test statistics is used to calculate the probability that a specific variant is associated with MPB for the given set of data. These probabilities are used to determine sets of variants within the risk loci—that is, credible sets of SNPs—which contain the causal variant 99% of the time.

**Estimation of heritability.** The Bonn cohort was used to estimate the heritability of MPB, since individual genotypes were available. All imputed variants with an imputation info score of  $> 0.6$  were extracted, and the corresponding genotypes were called using a probability threshold of 0.9. The resulting dataset comprised 997 male individuals and 8,271,786 SNPs. Variants with a MAF of  $> 0.01$  and a genotype SNP-call rate of  $> 95\%$  were then filtered. Only SNPs with a  $P$ -value of  $> 5\%$  for the corresponding test of differences in missingness between cases and controls were retained. After outlier detection with EIGENSTRAT based on autosomal data and the exclusion of individuals based on a genetic similarity of  $> 0.05$ , calculated by GCTA, we obtained the Genetic Relationship Matrix (GRM) for the autosomes and the X-chromosome with respect to 991 male individuals. With these two GRMs and 10 eigenvectors from the EIGENSTRAT step above, the heritability components were estimated using GCTA. The estimated heritability on the observed scale was 0.666 (0.234) for the autosomes and 0.277 (0.054) for the X-chromosome (standard errors in brackets). To obtain the estimated heritability on the liability scale, as described in Lee *et al.*<sup>67</sup>, a straightforward extension of this transformation was used to account for the extreme case and extreme control sampling. Table 2 in Hamilton<sup>68</sup> was used to divide the population into four bins. On the basis of age structure and Hamilton-Norwood (HN) grade, the transformation was calculated under the assumption that controls were selected from the first bin and cases from the last, thus reflecting the extreme sampling scheme. The resulting factor was approximately 0.509. This generated an estimated heritability for autosomal chromosomes of 0.339 (0.119), and 0.141 (0.027) for the X-chromosome.

**Estimation of explained phenotypic variance.** The amount of phenotypic variance explained by the 63 genome-wide significant risk loci was estimated as the correlation coefficient  $r^2$  in a multivariate linear model, in which the significant index SNPs were predictors and the outcome was modelled as 1 = case, 0 = control.

**Genotype-risk score analysis.** Using the lead SNPs from the 63 genome-wide significant loci, a genotype-risk score was constructed based on the weighted number of susceptibility alleles in an independent replication sample from the Heinz-Nixdorff Recall (HNR) cohort ( $N = 1,201$ ). Individuals with HN grade  $> \text{II}$  ( $N = 1,108$ ) were defined as cases, and individuals with HN-grade I or II at age  $< 65$  years were defined as controls ( $N = 93$ ). The weights were established using the beta coefficient from the meta-analysis. The resulting genotype score was divided into four quartiles. The risk for MPB was then tested in each quartile using a logistic regression model and the lowest quartile as a reference. Two models were tested: (I) genotype-risk score; and (II) age-adjusted risk score. The results of this analysis are shown in (Table 2).

**DEPICT analysis.** DEPICT was used to identify plausible candidate genes at each of the 63 risk loci. As recommended in Pers *et al.*<sup>69</sup>, all independent SNPs with a  $P$ -value of  $< 5 \times 10^{-8}$  were included in the DEPICT analysis. Independent SNP sets were generated by retaining the most significant SNP from each set of SNPs with a pairwise LD of  $r^2 > 0.1$  and a physical distance of  $< 500$  kb. Pairwise LD coefficients were computed based on the imputing panel used in the eight GWAS

(1000 Genomes Project Phase I CEU, GBR and TSI genotype data). The results of these analyses are shown in Supplementary Tables 4–6.

**Identification of hair follicle eQTLs.** DNA from peripheral blood and RNA from occipital scalp hair follicle samples were obtained from 125 volunteer healthy male donors of German descent (mean age 27.9 years). Genome-wide genotyping and imputation of blood DNA samples were performed using Illumina's Human OmniExpress-12v1.0 bead array and IMPUTE2 (1000 Genomes, Phase I, June 2014). Whole-transcriptome profiling was performed on Illumina's HT-12v4 bead arrays after amplification and biotinylation of the hair follicle-RNA using the TotalPrep-96 RNA Amplification Kit (Illumina, San Diego, CA, USA). Expression data were quantile normalized, and only probes with a detection  $P$ -value of  $<0.01$  (Illumina GenomeStudio Software) in at least 5% of the samples were taken into account. The selected expression probes were subsequently filtered for: a unique alignment to the transcriptome; a perfect or good probe quality, as reported in the R package illuminaHumanv4.db; and mapping to an ENTREZ gene ID. After quality control and filtering, data for 14,687 expression probes and 6,593,881 SNPs were included in the eQTL analysis. Associations between gene expression levels and SNP genotypes in cis (distance between SNP and expression probe  $\leq 1$  Mb) was tested in MatrixEQTL using an additive linear regression model. Association tests were corrected for the top five principle components. All eQTL findings at a false discovery rate (FDR) of  $<0.001$  were considered significant.

**mRNA and miRNA expression analysis.** Whole-transcriptome profiling of hair follicle-RNA samples ( $N=125$ ) was performed as described above. MiRNA profiling of hair follicle samples ( $N=25$ ) was performed on the AffymetrixGeneChip miRNA 4.0 (Affymetrix, Santa Clara, CA, USA) using a total of 250 ng of hair follicle miRNA. Poly(A) tailing and biotinylation were performed with the AffymetrixGeneChip Hybridization, Wash and Stain Kit, in accordance with the manufacturer's instructions. Data were analysed using the Affymetrix Expression Console software (v.1.4) (Affymetrix). MiRNAs were defined as 'present' ( $N=1,169$ ) or 'absent' ( $N=1,409$ ) according to the implemented RMA (robust multichip average) and DABG (detected above background) methods.

**Identification of miRNA target genes at MPB-risk loci.** MiRNA target genes at MPB-risk loci ( $\pm 500$  kb from MPB lead SNP) were identified using the miR-Walk2.0 algorithm (last accessed 24 March 2016). Only validated target genes and genes that were predicted by the miRWalk algorithm and at least three additional implemented databases were taken into account.

**Overlap between eQTLs and MPB-risk variants.** To identify regulatory effects at known and novel MPB-risk loci, the present meta-analysis data were compared with the eQTL data from: (i) the Blood eQTL Browser (<http://genenetwork.nl/bloodseqbrowser>; last accessed on 21 December 2015); (ii) the GTEx Browser (<http://www.gtexportal.org/home>); and (iii) an unpublished eQTL dataset from human hair follicle. MPB-risk variants were considered to coincide with an eQTL if the MPB lead SNP itself, or any SNP in  $r^2 > 0.5$ , showed an eQTL effect. The complete list of overlapping eQTL findings is provided in Supplementary Table 2.

**Enhancer enrichment analysis.** Enhancers (including promoters) were defined by chromatin immunoprecipitation-sequencing (ChIP-seq) peaks of histone 3 lysine 27 acetylation (H3K27ac). A total of 140 H3K27ac data sets were downloaded from the roadmap epigenomics project (<http://www.roadmapepigenomics.org>), and reprocessed using DFilter. Only peaks with  $P < 10^{-10}$  (DFilter) were retained. Four in house data sets from a human balding and a non-balding dermal papilla cell line with and without 10 nM of DHT treatment were included<sup>70</sup>. In these lines, a H3K27ac antibody (Abcam, Cat. ab4729; 5  $\mu$ g antibody per 100  $\mu$ g chromatin) was used for ChIP. The MPB credible SNPs were overlapped with the enhancer (H3K27ac) sets of each of the 144 cell lineages and tissues. For tissue (or cell lineage) A, the remaining 143 enhancer sets were pooled together to define a tissue agnostic superset. The superset was classified in feature categories based on: (i) distance to transcription start site; (ii) GC content; (iii) length of the enhancer site. To obtain for the enhancer set of tissue A an appropriate null distribution, we randomly drew for every enhancer of tissue A an enhancer from a matched feature category of the superset. The number of credible SNP overlaps across this matched enhancer set was computed, and a million permutations of feature matched enhancer sets were conducted and overlapped with the credible SNP set to define a distribution for tissue A. This procedure was conducted for each of the 144 tissue specific enhancer sets. For Fig. 2, enhancers of cell lineages/tissues that belonged to the same tissue type were merged, resulting in 23 grouped enhancer sets. The above framework was repeated for these grouped enhancer sets. The results of these analyses are provided in Supplementary Data 1.

**Ingenuity pathway enrichment analysis.** To test for an enrichment of MPB candidate genes in canonical pathways, the Ingenuity Pathway Analysis was used (Qiagen, Hilden, Germany). IPA considers 658 pathways, and calculates enrichment based on the right-tailed Fisher's exact test. All genes within 500 kb of the MPB lead SNPs were included in the analysis. Only pathways with  $\geq 3$  annotated

genes were taken into account. The complete list of nominally significant pathways is provided in Supplementary Table 7.

**NHGRI GWAS catalogue.** Previously reported GWAS associations that mapped to the associated regions and showed an  $r^2 \geq 0.3$  and/or  $D' > 0.8$  (1000 Genomes Project, Phase 3) with the respective MPB index SNP were extracted from the NHGRI GWAS catalogue (<http://www.ebi.ac.uk/gwas/>; last accessed on 14 January 2016). This resulted in the identification of 124 overlapping associations, as shown in Supplementary Data 3.

**Data availability.** All data that support the findings of this study are available from the corresponding author upon reasonable request.

## References

- Qi, J. & Garza, L. A. An overview of alopecias. *Cold Spring Harb. Perspect. Med.* **4**, a013615 (2014).
- Han, J. *et al.* A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.* **4**, e1000074 (2008).
- Yu, N. L., Tan, H., Song, Z. Q. & Yang, X. C. Illness perception in patients with androgenetic alopecia and alopecia areata in China. *J. Psychosom. Res.* **86**, 1–6 (2016).
- Heilmann-Heimbach, S., Hochfeld, L. M., Paus, R. & Nothen, M. M. Hunting the genes in male-pattern alopecia: how important are they, how close are we and what will they tell us? *Exp. Dermatol.* **25**, 251–257 (2016).
- Inui, S. & Itami, S. Androgen actions on the human hair follicle: perspectives. *Exp. Dermatol.* **22**, 168–171 (2013).
- Ganzer, C. A., Jacobs, A. R. & Iqbal, F. Persistent sexual, emotional, and cognitive impairment post-finasteride: a survey of men reporting symptoms. *Am. J. Mens Health* **9**, 222–228 (2015).
- Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nat. Rev. Drug. Discov.* **12**, 581–594 (2013).
- Heilmann, S. *et al.* Androgenetic alopecia: identification of four genetic risk loci and evidence for the contribution of WNT signaling to its etiology. *J. Invest. Dermatol.* **133**, 1489–1496 (2013).
- Hillmer, A. M. *et al.* Susceptibility variants for male-pattern baldness on chromosome 20p11. *Nat. Genet.* **40**, 1279–1281 (2008).
- Li, R. *et al.* Six novel susceptibility loci for early-onset androgenetic alopecia and their unexpected association with common diseases. *PLoS Genet.* **8**, e1002746 (2012).
- Richards, J. B. *et al.* Male-pattern baldness susceptibility locus at 20p11. *Nat. Genet.* **40**, 1282–1284 (2008).
- Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 1296 (2016).
- Nyholt, D. R., Gillespie, N. A., Heath, A. C. & Martin, N. G. Genetic basis of male pattern baldness. *J. Invest. Dermatol.* **121**, 1561–1564 (2003).
- Rexbye, H. *et al.* Hair loss among elderly men: etiology and impact on perceived age. *J. Gerontol. A Biol. Sci. Med. Sci.* **60**, 1077–1082 (2005).
- Consortium GTEx. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
- Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
- Paus, R., Langan, E. A., Vidali, S., Ramot, Y. & Andersen, B. Neuroendocrinology of the hair follicle: principles and clinical perspectives. *Trends Mol. Med.* **20**, 559–570 (2014).
- Kizil, C. *et al.* Simplex/Fam53b is required for Wnt signal transduction by regulating beta-catenin nuclear localization. *Development* **141**, 3529–3539 (2014).
- Andl, T. & Botchkareva, N. V. MicroRNAs (miRNAs) in the control of HF development and cycling: the next frontiers in hair research. *Exp. Dermatol.* **24**, 821–826 (2015).
- Hebert, J. M., Rosenquist, T., Gotz, J. & Martin, G. R. FGF5 as a regulator of the hair growth cycle: evidence from targeted and spontaneous mutations. *Cell* **78**, 1017–1025 (1994).
- Higgins, C. A. *et al.* FGF5 is a crucial regulator of hair length in humans. *Proc. Natl Acad. Sci. USA* **111**, 10648–10653 (2014).
- Ota, Y. *et al.* Fibroblast growth factor 5 inhibits hair growth by blocking dermal papilla cell activation. *Biochem. Biophys. Res. Commun.* **290**, 169–176 (2002).
- Kwack, M. H., Kim, M. K., Kim, J. C. & Sung, Y. K. Dickkopf 1 promotes regression of hair follicles. *J. Invest. Dermatol.* **132**, 1554–1560 (2012).
- Kwack, M. H. *et al.* Dihydrotestosterone-inducible dickkopf 1 from balding dermal papilla cells causes apoptosis in follicular keratinocytes. *J. Invest. Dermatol.* **128**, 262–269 (2008).
- Inui, S. & Itami, S. Molecular basis of androgenetic alopecia: From androgen to paracrine mediators through dermal papilla. *J. Dermatol. Sci.* **61**, 1–6 (2011).
- Eriksson, N. *et al.* Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet.* **6**, e1000993 (2010).
- Sulem, P. *et al.* Two newly identified genetic determinants of pigmentation in Europeans. *Nat. Genet.* **40**, 835–837 (2008).

28. Praetorius, C. *et al.* A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell* **155**, 1022–1033 (2013).
29. Sturm, R. A. Molecular genetics of human pigmentation diversity. *Hum. Mol. Genet.* **18**, R9–R17 (2009).
30. Stough, D. *et al.* Psychological effect, pathophysiology, and management of androgenetic alopecia in men. *Mayo Clin. Proc.* **80**, 1316–1322 (2005).
31. Jain, R. & De-Eknankul, W. Potential targets in the discovery of new hair growth promoters for androgenetic alopecia. *Expert Opin. Ther. Targets* **18**, 787–806 (2014).
32. Schweikert, H. U. & Wilson, J. D. Regulation of human hair growth by steroid hormones. I. Testosterone metabolism in isolated hairs. *J. Clin. Endocrinol. Metab.* **38**, 811–819 (1974).
33. Hillmer, A. M. *et al.* Genetic variation in the human androgen receptor gene is the major determinant of common early-onset androgenetic alopecia. *Am. J. Hum. Genet.* **77**, 140–148 (2005).
34. Wehner, G. & Schweikert, H. U. Estrone sulfate source of estrone and estradiol formation in isolated human hair roots: identification of a pathway linked to hair growth phase and subject to site-, gender-, and age-related modulations. *J. Clin. Endocrinol. Metab.* **99**, 1393–1399 (2014).
35. Mak, K. K. & Chan, S. Y. Epidermal growth factor as a biologic switch in hair growth cycle. *J. Biol. Chem.* **278**, 26120–26126 (2003).
36. Ohnemus, U., Uenal, M., Inzunza, J., Gustafsson, J. A. & Paus, R. The hair follicle as an estrogen target and source. *Endocr. Rev.* **27**, 677–706 (2006).
37. Li, C. & Zhou, X. Melatonin and male reproduction. *Clin. Chim. Acta* **446**, 175–180 (2015).
38. Zerradi, M., Dereumetz, J., Boulet, M. M. & Tcherno, A. Androgens, body fat distribution and adipogenesis. *Curr. Opin. Endocrinol. Metab.* **3**, 396–403 (2014).
39. Crowley, S. J., Acebo, C. & Carskadon, M. A. Human puberty: salivary melatonin profiles in constant conditions. *Dev. Psychobiol.* **54**, 468–473 (2012).
40. Hamilton, J. B. Effect of castration in adolescent and young adult males upon further changes in the proportions of bare and hairy scalp. *J. Clin. Endocrinol. Metab.* **20**, 1309–1318 (1960).
41. Kobayashi, H. *et al.* A role of melatonin in neuroectodermal-mesodermal interactions: the hair follicle synthesizes melatonin and expresses functional melatonin receptors. *FASEB J.* **19**, 1710–1712 (2005).
42. Festa, E. *et al.* Adipocyte lineage cells contribute to the skin stem cell niche to drive hair cycling. *Cell* **146**, 761–771 (2011).
43. Trueb, R. M. Molecular mechanisms of androgenetic alopecia. *Exp. Gerontol.* **37**, 981–990 (2002).
44. Castellana, D., Paus, R. & Perez-Moreno, M. Macrophages contribute to the cyclic activation of adult hair follicle stem cells. *PLoS Biol.* **12**, e1002002 (2014).
45. Suzuki, S. *et al.* Localization of rat FGF-5 protein in skin macrophage-like cells and FGF-5 protein in hair follicle: possible involvement of two Fgf-5 gene products in hair growth cycle regulation. *J. Invest. Dermatol.* **111**, 963–972 (1998).
46. Jaworsky, C., Kligman, A. M. & Murphy, G. F. Characterization of inflammatory infiltrates in male pattern alopecia: implications for pathogenesis. *Br. J. Dermatol.* **127**, 239–246 (1992).
47. Piérard, G., Piérard-Franchimont, C., Nikkels-Tassoudji, N., Nikkels, A. & Léger, D. S. Improvement in the inflammatory aspect of androgenetic alopecia. A pilot study with an antimicrobial lotion. *J. Dermatol. Treat.* **7**, 153–157 (1996).
48. Trieu, N. & Eslick, G. D. Alopecia and its association with coronary heart disease and cardiovascular risk factors: a meta-analysis. *Int. J. Cardiol.* **176**, 687–695 (2014).
49. Arias-Santiago, S. *et al.* Androgenetic alopecia as an early marker of benign prostatic hyperplasia. *J. Am. Acad. Dermatol.* **66**, 401–408 (2012).
50. Hawk, E., Breslow, R. A. & Graubard, B. I. 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 (2000).
51. Amoretti, A., Laydner, H. & Bergfeld, W. Androgenetic alopecia and risk of prostate cancer: A systematic review and meta-analysis. *J. Am. Acad. Dermatol.* **68**, 937–943 (2013).
52. Fondell, E., Fitzgerald, K. C., Falcone, G. J., O'Reilly, E. J. & Ascherio, A. Early-onset alopecia and amyotrophic lateral sclerosis: a cohort study. *Am. J. Epidemiol.* **178**, 1146–1149 (2013).
53. Ryan, C. J. & Chan, J. M. Hair, hormones, and high-risk prostate cancer. *J. Clin. Oncol.* **33**, 386–387 (2015).
54. Vanderschueren, D., Gaytan, J., Boonen, S. & Venken, K. Androgens and bone. *Curr. Opin. Endocrinol. Diabetes Obes.* **15**, 250–254 (2008).
55. Khan, Q. J. & Fabian, C. J. How I treat vitamin D deficiency. *J. Oncol. Pract.* **6**, 97–101 (2010).
56. Mokry, L. E. *et al.* Vitamin D and risk of multiple sclerosis: a mendelian randomization study. *PLoS Med.* **12**, e1001866 (2015).
57. Cousminer, D. L. *et al.* Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty. *Hum. Mol. Genet.* **23**, 4452–4464 (2014).
58. Elks, C. E. *et al.* Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat. Genet.* **42**, 1077–1085 (2010).
59. Jin, G. *et al.* Genome-wide association study identifies a new locus JMJD1C at 10q21 that may influence serum androgen levels in men. *Hum. Mol. Genet.* **21**, 5222–5228 (2012).
60. Perry, J. R. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).
61. de Bakker, P. I. W. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17**, R122–R128 (2008).
62. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
63. Higgins, J. P., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. *BMJ* **327**, 557–560 (2003).
64. Heilmann, S. *et al.* Evidence for a polygenic contribution to androgenetic alopecia. *Br. J. Dermatol.* **169**, 927–930 (2013).
65. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
66. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
67. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).
68. Hamilton, J. B. Patterned loss of hair in man; types and incidence. *Ann. N.Y. Acad. Sci.* **53**, 708–728 (1951).
69. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
70. Chew, E. G. *et al.* Differential expression between human dermal papilla cells from balding and non-balding scalps reveals new candidate genes for androgenetic alopecia. *J. Invest. Dermatol.* **136**, 1559–1567 (2016).
71. Brockschmidt, F. F. *et al.* Susceptibility variants on chromosome 7p21.1 suggest HDAC9 as a new candidate gene for male-pattern baldness. *Br. J. Dermatol.* **165**, 1293–1302 (2011).

## Acknowledgements

The study was supported by the Life & Brain GmbH; the BONFOR programme of the University of Bonn; and the Agency for Science, Technology and Research (A\*STAR). M.M.N. is a member of the DFG Excellence Cluster ImmunoSensation. E.G.Y.C. is supported by the A\*STAR Graduate Scholarship programme. X.T.H. is supported by the A\*STAR Joint Council Office. We thank the participants of the following cohorts: 23andMe, CoLaus, Bonn, Twins UK, Nijmegen Biomedical Study, QIMRB1, QIMRB2 and THISEAS.

## Author contributions

S.H.-H., M.M.N. conceived and designed the experiments. C.H., T.B. performed the statistical analyses; S.H.-H., L.M.H., A.M.H., R.L., E.G.Y.C., D.R.N., G.W., P.V. performed the experiments; S.H.-H., L.M.H., J.H., H.L.F., A.M.H., S.P., M.N., D.D., J.B.R., R.L., X.T.H., E.G.Y.C., R.C.-H.d.R., A.J., Sh.P., D.R.N., S.K., G.P., R.R., D.A.H. analysed the data; S.M., T.A., J.B.R., T.D.S., R.C.-H.d.R., D.R.N., N.G.M., L.A.L.M.K., T.E.G., M.P.P., P.D., G.D., Z.K., G.W., P.V. contributed reagents, materials, analysis tools; S.H.-H., C.H., T.B., A.M.H., M.M.N. wrote the manuscript and R.P. revised the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** D.A.H. is an employee of the company 23andMe. The remaining authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Heilmann-Heimbach, S. *et al.* Meta-analysis identifies novel risk loci and yields systematic insights into the biology of male-pattern baldness. *Nat. Commun.* **8**, 14694 doi: 10.1038/ncomms14694 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017