Assessment of rosacea symptom severity by genome-wide association study and expression analysis highlights immuno-inflammatory and skin pigmentation genes

Jennifer L. Aponte1,*, Mathias N. Chiano2, Laura M. Yerges-Armstrong3, David A. Hinds4, Chao Tian4, Akanksha Gupta5, Cong Guo3, Dana J. Fraser1, Johannes M. Freudenberg3, Deepak K. Rajpal3, Margaret G. Ehm3 and Dawn M. Waterworth3

1Genomic Medicine, PAREXEL International, Research Triangle Park, NC 27713, USA, 2Target Sciences, GlaxoSmithKline, Stevenage, UK, 3Target Sciences, GlaxoSmithKline, Collegeville, PA 19426, USA, 423andMe Inc., Mountain View, CA 94041, USA and 5Translational Science, Dermatology, GlaxoSmithKline, Research Triangle Park, NC 27709, USA

*To whom correspondence should be addressed at: PAREXEL International, 2520 Meridian Parkway, Durham, NC 27713, USA. Tel: +1 9196664812; Fax: +1 9195445410; Email: Jennifer.Aponte@PAREXEL.com

Abstract

Rosacea is a common, chronic skin disease of variable severity with limited treatment options. The cause of rosacea is unknown, but it is believed to be due to a combination of hereditary and environmental factors. Little is known about the genetics of the disease. We performed a genome-wide association study (GWAS) of rosacea symptom severity with data from 73,265 research participants of European ancestry from the 23andMe customer base. Seven loci had variants associated with rosacea at the genome-wide significance level (P < 5 x 10^-8). Further analyses highlighted likely gene regions or effector genes including IRF4 (P = 1.5 x 10^-17), a human leukocyte antigen (HLA) region flanked by PSMB9 and HLA-DMB (P = 2.2 x 10^-15), HERC2-OCA2 (P = 4.2 x 10^-13), SLC45A2 (P = 1.7 x 10^-10), IL13 (P = 2.8 x 10^-7), a region flanked by NRXN3 and DIO2 (P = 4.1 x 10^-9), and a region flanked by OVOL1 and SNX32 (P = 1.2 x 10^-8). All associations with rosacea were novel except for the HLA locus. Two of these loci (HERC-OCA2 and SLC45A2) and another precedent variant (rs1805007 in melanocortin 1 receptor) with an association P value just below the significance threshold (P = 1.3 x 10^-7) have been previously associated with skin phenotypes and pigmentation, two of these loci are linked to immuno-inflammation phenotypes (IL13 and PSMB9-HLA-DM) and one has been associated with both categories (IRF4). Genes within three loci (PSMB9-HLA-DMA, HERC-OCA2 and NRX3-DIO2) were differentially expressed in a previously published clinical rosacea transcriptomics study that compared lesional to non-lesional samples. The identified loci provide specificity of inflammatory mechanisms in rosacea, and identify potential pathways for therapeutic intervention.
Introduction

Over 16 million Americans are affected with rosacea, a chronic inflammatory skin disease with poorly understood etiology, genetics and pathophysiology (1,2). Studies have suggested that rosacea may occur three times more often in women than in men. Although women are more likely to experience symptoms on the cheeks and chin, men are more likely to have severe symptoms and are more commonly affected on the nose (3). Although the clinical presentation is highly variable, rosacea has been classified into four common subtypes: (i) erythematotelengiastic rosacea (ETR) characterized by flushing and persistent redness on face; (ii) papulopustular rosacea (PPR) characterized by persistent facial redness with papules or pustules that come and go; (iii) phymatous rosacea (PhR) characterized by thickening skin, irregular nodularities and enlargement primarily of the nose and (iv) ocular rosacea where the eyes display a host of symptoms that may include irritation, dryness, stinging, light sensitivity and decreased visual acuity (1). Rosacea patients often experience progression from one subtype to another (4). The onset of rosacea typically occurs between the ages of 30 and 60 years as intermittent facial redness, progressing into visible blood vessels, bumps, papules, pustules, eye irritation and swelling of the nose if left untreated (5).

The exact pathogenesis of rosacea is unknown, but dysregulation of the innate and adaptive immune system, overgrowth of commensal skin organisms and abnormal neurovascular signaling may all have a role (1). Studies suggest that rosaces occurs in individuals with an inappropriate innate immune response to environmental triggers, such as stress and sun exposure, leading to inflammation and vascular abnormalities (6). A recent transcriptome-wide study of biopsies from rosacea patients and healthy volunteers suggested distinct differences between three rosacea subtypes (ocular rosacea not reported) and observed that all subtypes are characterized by Th1/Th17 polarized inflammation, and macrophage and mast cells infiltration (7). Despite progress in understanding pathophysiology of rosacea, work remains to understand the complex dysregulation of cells and mechanisms of disease response as well as genetic determinants of disease. Twin studies have demonstrated that genetics may contribute to rosacea incidence and severity (8), but only one genome-wide scan has been performed to date, also with 23andMe, Inc. research participants (9). The study identified an association between rosacea risk and a single nucleotide polymorphism (SNP) in the human leukocyte antigen (HLA) region located between HLA-DRA and BTNL2, which supported the concept of an inflammatory genetic component to rosacea. However, the study’s sample size was limited to 22,952 individuals, employed a binary disease definition (2618 cases/20,334 controls) and did not investigate rosacea symptom severity. The absence of animal models further emphasizes the need about their rosacea symptoms and had greater than 97% European ancestry, as determined through an analysis of local ancestry. Rosacea symptom severity was evaluated using a composite 32-point score derived from survey questions about rosacea-related facial, nose and ocular symptoms analyzed as a continuous trait. Association analysis used a linear model adjusting for age, sex and the first five ancestry principal components. Table 1 summarizes the distribution of the rosacea symptom severity score for the participant group by age and gender. Although the number of men and women with symptoms of any severity are similar, the number of women reporting rosacea-like symptoms increases to three times that of men as the score increases. Higher scores were observed for women as compared with men and for the older age groups (>45 years) as compared with the 30–45 years age category.

The results of the GWAS are shown in Figure 1. Variation within seven genomic regions showed significant association with rosacea symptom severity at the genome-wide level ($P < 5 \times 10^{-8}$). Patterns of association in these genomic regions are illustrated using regional plots (Supplementary Material, Fig. S1). Summary statistics for all variants with suggestive associations ($P < 5 \times 10^{-7}$) to rosacea symptom severity are listed in Supplementary Material, Table S1. Although the genomic control value was large ($\hat{G}_C$ of 1.0864), linkage disequilibrium (LD) score regression analysis had an intercept of 1.0099 suggesting that polygenicity accounts for the majority of the increase in the mean $\hat{\beta}$ statistic. We analyzed the associated loci with the Probabilistic Identification of Causal SNPs (PICS) method which uses LD and the pattern of association in order to determine the most likely causal variant or variants (referred to as the PICS set, Supplementary Material, Table S2) (11). We calculated the PICS probability assuming a European reference ancestry for each associated locus and identified the most likely causal SNP or SNPs (i.e. the PICS set) using a $>90\%$ probability threshold. Chromatin regulatory marks in blood and immune cells were evaluated to identify possible regulatory variants in each region (Supplementary Material, Figs S2–S7). To provide support for proposed effector genes within associated loci, variants within these regions were further evaluated to determine cis eQTLs using publicly available data in GTEx (12) (Supplementary Material, Table S3). Summary information for the most significant SNP in each region, including the most likely ‘effector’ gene for each locus, as well as information on previous reports implicating these variants can be found in Table 2.

Of the seven loci, two had previously been genetically associated with skin phenotypes and pigmentation (SLC45A2 and HERC2–OCA2), two with immuno-inflammation phenotypes (IL13 and HLA-DMA/B) and one gene (IRF4) was previously associated with both phenotypic categories. Two loci were intergenic (NRX3-DIO2 and OVO1–SNX32) and did not contain any obvious candidate genes for rosacea. In addition to these loci, another variant with a $P$ value just below the significance threshold, rs1805007 in melanocortin 1 receptor (MC1R; $P < 1.3 \times 10^{-7}$), has been previously genetically associated with skin phenotypes and pigmentation. A recently published microarray-based mRNA expression dataset (7) was reanalyzed to evaluate the genes and molecular pathways underlying rosacea and its subtypes that are dysregulated on the transcriptional level by comparing PPR, ETR and PhR samples to samples derived from healthy volunteers. Overall, transcriptional signatures for the three subtypes were highly similar (Fig. 2A). Differential expression analysis of the seven loci with $P$ values less than $5.0 \times 10^{-6}$ plus the MC1R locus showed that genes (including HLA-DMA, PSM89, HERC2, OCA2, NRX3 and DIO2) in three of the seven gene regions (PSMB9–HLA-
DMA, HERC2-OCA2 and NRX3-DIO2) were differentially regulated in rosacea gene expression data (Fig. 2B).

Our results show that the index SNP rs12203592 in the intronic region of the interferon regulatory factor 4 (IRF4) gene was strongly associated with rosacea symptom severity; the T allele is significantly associated with increased rosacea symptom severity (\( P = 1.5 \times 10^{-17} \)). The associated variant is located within a melanocyte-specific enhancer (13) and is significantly associated with pigmentation traits in previously conducted GWAS (14). The rs12203592 variant is the only variant identified associated with pigmentation traits in previously conducted GWAS (14). The rs12203592 variant is the only variant identified associated with pigmentation traits in previously conducted GWAS (14).

The third locus associated with rosacea symptom severity was at 15q13 (rs1129038, \( P = 4.2 \times 10^{-12} \)). This SNP is located in the 3’UTR of the HERC2 (probable E3 ubiquitin-protein ligase) gene exon 93 and upstream of OCA2 (probable E3 ubiquitin-protein ligase) and is part of a haplotype block that predicts eye color. Our results show that the minor allele (T) is significantly associated with increased rosacea symptom severity. Variant rs12913832, also significantly associated in our study, \( P = 5.3 \times 10^{-12} \), is located in an intron in HERC2 affecting the expression of OCA2 in the human iris, and is common to nearly all people with blue eyes (15). Two markers, rs1129038 and rs12913832, are in the PICS set and both overlap enhancers in skin cells. Although rs12913832 has a significant eQTL for HERC2 in whole blood (GTEX, \( P = 1.7 \times 10^{-5} \)), this may be a spurious or coincidental finding as the top eQTL for HERC2 in whole blood is the different, nearby SNP rs8023410 (GTEX, eQTL \( P = 1.7 \times 10^{-3} \)), which was only nominally associated with rosacea symptom severity (\( P = 8.41 \times 10^{-4} \)). On the other hand, rs12913832 resides in a putative enhancer in foreskin melanocytes (Supplementary Material, Fig. S3) and has a higher PICS probability (0.66). Both
Table 2. Characteristics of the most significant GWAS SNPs associated with rosacea symptom severity score

| rsID     | Chromosome position | Allele | MAF   | Gene context | Possible effector gene(s) | P value     | Effect β (95% CI) | Gene description | Prior published associations with index SNP or SNP(s) within 500 kb and moderate LD (r² > 0.5) |
|----------|---------------------|--------|-------|--------------|--------------------------|-------------|-------------------|------------------|---------------------------------------------------------------|
| rs12203592 | 6p25.3: 396321       | C/T    | 0.174 | [IRF4]       | IRF4                     | 1.5 x 10⁻¹⁷ | 0.221 (0.169, 0.273) | Interferon regulatory factor 4 | Hair color (14,35,36), eye color (14,35), freckling (35), non-melanoma skin cancer (36), sunburns/tanning (14,36) |
| rs57390839 | 6p21.32: 32900383    | TT/–   | 0.07  | PSMB8,[J]- HLA-DM, HLA-DMB | HLA-DMA, HLA-DMB, OCA2, HERC2 | 2.2 x 10⁻¹⁵ | -0.303 (-0.380, -0.227) | HLA class II histocompatibility antigen Probable E3 ubiquitin-protein ligase | Autoimmune and inflammatory disease (21) Generalized vitiligo (40), tanning (36), eye color (35,36,57–61), hair color (14,35,36,57) |
| rs1129038  | 15q13.1: 28356859    | C/T    | 0.266 | [HERC2]     | OCA2, HERC2              | 4.2 x 10⁻¹² | 0.157 (0.111, 0.202) | Probable E3 ubiquitin-protein ligase | Generalized vitiligo (40), skin color (36), tanning (36), freckling (35,57), melanoma (66) |
| rs16891982 | 5p13.2: 33951693     | C/G    | 0.042 | [SLC45A2] [MATP] | SLC45A2                  | 1.7 x 10⁻¹⁰ | 0.317 (0.218, 0.202) | Solute carrier family 45 member 2 | Asthma (23,24), atopic dermatitis (25,26), psoriasis (28), self-reported allergy (27), Hodgkin's lymphoma (62), IgE levels (63) |
| rs847     | 5q31.1: 131996669    | C/T    | 0.2   | [IL13]      | IL13                     | 2.8 x 10⁻⁹ | -0.143 (-0.191, -0.095) | Interleukin 13 | Hair color (14,35), eye color (35), skin pigmentation (64), tanning (65), malignant melanoma (66) |
| rs149851565 | 14q31.1: 80509104    | C/A    | 0.116 | NRNX3[–]–DIQ2 | NRNX3[–]–DIQ2             | 4.1 x 10⁻⁹ | 0.175 (0.115, 0.235) | | No previous associations IBD (18) |
| rs77779142 | 11q13.1: 65599656    | C/T    | 0.159 | OVOL1[–]–SNX32 | OVOL1[–]–SNX32           | 1.2 x 10⁻⁸ | 0.150 (0.097, 0.203) | | |
| rs1805007  | 16q24.3: 89986117    | C/T    | 0.074 | TCF25[–]–TUBB3 | MCIR                     | 1.3 x 10⁻⁷ | 0.189 (0.117, 0.261) | Melanocortin 1 receptor | Hair color (14,36,57), skin sensitivity to sun (57), freckling (35,57), basal cell carcinoma (67), non-melanoma skin cancer (36), melanoma (66,68,69), homocysteine levels (70) |

The table of index or most significant SNPs in each associated locus. Regions were defined by identifying SNPs with P < 1 x 10⁻⁷, then grouping these into intervals separated by gaps of at least 250 kb, and choosing the SNP with smallest p within each interval. Mapping and gene context was based on NCBI Build 37, and the gene context for the most likely genes were derived using the HG19 release of the UCSC Known Genes tables. The gene context field has the following interpretations: Gene1, Gene2: The SNP is contained within the transcripts of the specified gene(s); Gene1[–]–Gene2: The SNP is flanked by Gene 1 and Gene2. Dashes indicate distance: ‘<10 kb’, ‘<100 kb’, ‘<1000 kb’. The two SNP alleles are in order of major/minor allele. Putative effector genes at each locus were identified by first defining the credible set of SNPs via PICS, and then annotating variants and genes within each credible set using functional fine mapping, differential expression analysis, and prior evidence of association.
OCA2 and HERC2 are significantly down-regulated in the rosacea gene expression data making it difficult to distinguish between OCA2 and HERC2 as the likely effector gene.

The fourth SNP associated with rosacea symptom severity (rs16891982, \(P = 1.7 \times 10^{-13}\)) at 5p13 is in the solute carrier family 45 member 2 gene (SLC45A2), which encodes a transporter protein that mediates melanin synthesis with the G allele associated with increased symptom severity. Evaluation of the PICS data and the LD pattern across the region identified six SNPs likely to have a causal effect on rosacea. Of these, rs16891982 has the highest PICS probability (0.88) and is a missense variant (Leu374Phe) predicted to be functional (16). A gene regulatory mechanism is unlikely, because of a lack of overlap with active chromatin marks (Supplementary Material, Fig. S4). SLC45A2 is already known to play a role in pigmentation phenotypes suggesting that it is the likely effector gene.

The most significant variant observed within the fifth locus is an intronic SNP (rs847) within the interleukin 13 (IL13) gene, with the minor allele (T) strongly associated with less severe forms of rosacea (\(P = 2.8 \times 10^{-9}\)). This variant is in strong LD \((r^2 > 0.96)\) with a missense SNP, rs20541 (Gln144Arg), that is also associated with rosacea symptom severity \((P = 8.1 \times 10^{-4}\)). IL13 is a cytokine derived from helper T-cell lymphocytes that inhibits mononuclear cell inflammation (17). Five SNPs were identified in the PICS set (rs847, rs848, rs1295685, rs20541 and rs1295686). The SNPs rs1295686 and rs20541 reside within a putative enhancer in B-lymphocytes (Supplementary Material, Fig. S5) and all, except the rs848 variant, were associated with IL13 expression in testis in GTEx \((P < 7 \times 10^{-5})\). Given the patterns of association and LD across the locus, this association signal is likely modulating IL13 expression and inflammation.

At the next locus, the strongest association was with an intergenic variant SNP located in the NRXN3-DIO2 genomic region at 14q31 (rs149851565, \(P = 4.1 \times 10^{-8}\)), with the minor allele (A) associated with increased rosacea symptom severity. There are no previous associations reported for this SNP and no
significant eQTLs are reported in this region. Determining the effector gene for this region is further complicated by the fact that the associated variants are in high LD with 53 other potentially causal variants and by the lack of robust overlap with regulatory markers in relevant tissues (Supplementary Material, Fig. S6).

For the final locus, the most significant association is with an intergenic variant located between OVOL1-SNX32 at 11q13 (rs77779142, \(P = 1.2 \times 10^{-10}\)), with the T allele associated with more severe forms of rosacea. No previous associations are reported for this SNP; however, this variant is in high LD (\(r^2 = 0.94\)) with rs2231884 that has previously been associated with inflammatory bowel disease (IBD) (18). Within the locus, 32 SNPs that may be causally related to rosacea overlap enhancer and promoter sites in skin and blood. The SNP rs17854357 is located in a blood and skin cell promoter of SNX32. The SNP rs11227332 resides within a regulatory element present in both tissue types near CFL1. SNPs rs12225345 and rs658524 are located within a blood cell type-specific promoter element. Lastly, there is a cluster of variants in the promoter of FIBP (Supplementary Material, Fig. S7). The plethora of regulatory mechanisms at the locus makes it difficult to pinpoint a causal mechanism. In GTEx, the SNPs are multiple significant (\(P < 1 \times 10^{-11}\)) eQTLs in multiple tissues including skin in GTEx for genes in this region (SNX32, CTSW and FIBP). However, only the eQTL results for SNX32 were concordant with the most significant eQTL results for that gene making the CTSW and FIBP results more likely coincidental or spurious. In addition, SNX32, FIBP and CTSW were not differentially regulated in the rosacea gene expression data. This evidence, collectively, does not point to a clear effector gene in this locus.

In addition to the seven loci described, a non-synonymous deleterious missense variant (Arg151Cys) mapping to the MC1R gene was associated with rosacea symptom severity and the (T) allele associated with more severe forms of the disease, but did not meet the threshold for genome-wide significance (\(P = 1.3 \times 10^{-7}\)). The pattern of association and LD across this region suggests that this variant, as well as 23 other variants within this locus could be causally related to rosacea. MC1R plays an important role in normal pigmentation (see Table 2). The receptor is primarily located on the surface of melanocytes, specialized cells that produce melanin. There are numerous eQTL associations in GTEx in multiple tissues including brain but no significant signals were identified for skin.

### Table 3. Genetic correlation between rosacea symptom severity score and other immune mediated diseases

| Trait                        | \(r_G\) | SE\(_r_G\) | \(P\)  |
|------------------------------|---------|------------|-------|
| Eczema (71)                  | -0.268  | 0.122      | 0.028 |
| Crohn’s disease (72)         | 0.187   | 0.07       | 0.008 |
| IBD (72)                     | 0.228   | 0.069      | 0.001 |
| UC (72)                      | 0.217   | 0.074      | 0.003 |
| Asthma (73)                  | -0.094  | 0.116      | 0.417 |
| Rheumatoid arthritis (74)    | -0.051  | 0.068      | 0.649 |
| Multiple sclerosis (75)      | 0.109   | 0.158      | 0.492 |
| Systemic lupus erythematosus (76) | 0.132 | 0.083      | 0.113 |
| Primary biliary cirrhosis (77) | 0.199 | 0.094      | 0.033 |
| Celiac disease               | 0.2409  | 0.107      | 0.024 |
| Primary sclerosing cholangitis | -0.0815 | 0.122      | 0.5039 |

\(r_G\) refers to the genetic correlation between rosacea symptom severity and the listed traits, SE\(_r_G\) is the standard error of the genetic correlation and \(P\) is the \(P\) value of the genetic correlation.

Given the overlap observed with immune and pigmentation loci, we evaluated the genetic correlation between rosacea symptom severity and relevant conditions where GWAS summary statistics were available using LD score regression. Unfortunately, pigmentation GWAS were not available, but we compared the association results with 10 other GWAS for immune conditions. We observed significant (\(P < 0.05/10 = 0.005\)) positive genetic correlations with IBD (\(r_G = 0.228, SE_{r_G} = 0.069; P = 0.001\)) and ulcerative colitis (UC) (\(r_G = 0.217; SE_{r_G} = 0.074; P = 0.003\)), Table 3 summarizes results from the analysis. Several large epidemiological studies have shown correlations between these conditions, with a UK study indicating an increased risk for rosacea in patients with a history of UC (19) and the Nurse’s Health Study in the US demonstrating an increased risk of Crohn’s disease in women with a history of rosacea (20). Therefore, evidence exists connecting these two conditions, but more investigation is required to understand causality.

### Discussion

In this study of rosacea symptom severity, we leveraged a large database of research participants’ self-reported data with rosacea-related symptom information and identified novel susceptibility loci. This information coupled with gene expression analyses highlighted gene regions previously implicated in skin pigmentation (HERC2-OCA2, SLC45A2 and MC1R), inflammation (IL13, HLA and potentially OVOL1-SNX32) or both (IRF4), supporting the hypothesis that both skin type and inflammatory mechanisms play a critical role in rosacea pathophysiology. LD score regression analysis also demonstrated significant correlation with the genetic associations in the immune conditions, IBD and UC. Together with the inflammatory mechanisms identified in the published transcriptomics study (7), this study further highlights the importance of understanding immune-mediated mechanisms and to explore them further for therapeutic resolution. This study comprises more than 73 000 participants and provides insight beyond the HLA region, previously associated with rosacea risk (9). Supplementary Material, Table S4 summarizes the case and control status for participants included in the present study and the primary cohort published in Chang et al. (9) illustrating that participants with higher scores are more likely to be diagnosed with rosacea versus those with lower scores and that the symptom score captures information beyond a traditional doctor diagnosis.

Analysis of these data provided insight into the value of self-reported data. The phenotype analyzed in this study included a burden score that has not been previously studied using a genome scan approach. Even though this is not a clinically validated score, the increased yield of significant results beyond the simple case–control definition suggests that this score is meaningful in disease pathogenesis terms and supports the idea of a continuum of disease symptoms across the subtypes. Although the loci identified in this study have not been replicated, the size of the study, the strength of statistical associations and the likely functional impacts of the effector genes in these regions provide compelling evidence for prioritizing these loci for further in-depth investigations.

The associations of rosacea with the IL13, IRF4 and HLA gene regions are consistent with the known inflammatory component of the disease. The HLA-DMA/B genes that map most closely to our associated variant are part of the MHC class II, expressed in antigen presenting cells. Many associations with autoimmune conditions have been shown to be specific to class II alleles (21) including susceptibility to rheumatoid arthritis,
multiple sclerosis and celiac disease. For rosacea, it has also been suggested that HLA associations may help explain the connection of various microbes with rosacea (22). The HLA gene locus identified here for rosacea symptom severity overlaps with a region previously associated with rosacea risk (9). Although our study provides greater power to precisely dissect contributions of MHC variants to rosacea symptoms, accurate mapping of disease loci and detailed interpretation of MHC disease associations is still problematic because of the density of MHC genes, the strong LD between genes and the effects of multiple HLA loci (21).

IL-13 is a cytokine secreted by many cell types, but especially T helper type 2 (Th2) cells (17), and is a mediator of allergic inflammation and disease. Our associated SNPs in the region, rs1295686 and rs20541, have been previously implicated in autoimmune diseases (rs1295686 with asthma (23,24) and atopic dermatitis (25,26), and rs20541 with self-reported allergy (27) and psoriasis (28)). However, not all these associations are in the same direction, for example the minor allele of rs20541 confers risk of atopic dermatitis, but is protective for rosacea and psoriasis. Recent studies also report positive findings on the efficacy of new therapies targeting cytokines interleukin IL-13 and IL-4 in asthma and atopic dermatitis (25,30).

IRF4, a T cell receptor-induced transcription factor, functions in partnership with multiple molecular partners to modulate adaptive immune response (31). IRF4 plays a role in the initiation and progression of autoimmune diseases (32) such as celiac disease (33) and is strongly associated with human pigmentation traits (34). The IRF4 variant described here affects skin color by modulating enhancer-mediated transcriptional regulation (13) and has been previously associated with hair color (14,35,36) eye color (14,35), skin color (14), freckling (35), non-melanoma skin cancer (36) and sunburn/tanning (14,36). Additionally, IRF4 plays a critical role in the fate of T-cells as it promotes differentiation of naive CD4(+) T cells into various cell types such as T helper 2 (Th2), Th9, Th17 or T follicular helper (Tfh) cells (37). These results support the possibility of IRF4 playing a dual role in influencing skin susceptibility to rosacea as well as a potentially mediating inflammation in the development of rosacea. Collectively, the HLA, IL13 and IRF4 findings highlight the critical role of immune modulation and rosacea disease pathology.

This study in individuals of 97% European ancestry has identified significant associations with rosacea genes previously implicated in skin pigmentation (HERC2-OCA2, SLC45A2, IRF4 and MC1R). Skin pigmentation is primarily due to the function of melanocytes and cutaneous homeostatic mechanisms involving multiple cell types interacting with melanocytes. Melanosomes, components of melanocytes, synthesize two types of melanins: eumelanin (black-brown) and phaeomelanin (red/yellow). The type and quality of melanin produced, and the cutaneous cell interactions determine the skin color of an individual. MC1R regulates the quality and quantity of melanin production and minor vitiligo autoantigen, and upon activation switches to eumelanin production (38). MC1R agonism along with ultraviolet B exposure assisting repigmentation has been reported (39). Associations between genetic variation in MC1R and malignant melanoma, vitiligo (a disease characterized by depigmented skin and hair resulting from autoimmune destruction of epidermal melanocytes) (40) and pigmentation traits have been reported (Table 2). Reduced production of melanin due to genetic mutations in the protein encoded by SLC45A2 results in oculocutaneous albinism type IV (OCA4), an autosomal recessive inherited disorder with reduced melanin pigmentation in skin, hair and eyes (41). SLC45A2 is also genetically associated with a variety of pigmentation phenotypes (Table 2). Another noteworthy association is with variation within the HERC2-OCA2 locus, in particular a variant in an evolutionarily conserved region of HERC2 that downregulates expression of OCA2. OCA2 is causal for oculocutaneous albinism type 2, encodes a melanosomal membrane transporter and has a major role in determining skin, hair and eye color (Table 2). HERC2 has also been associated with human pigmentation traits, and may have multiple roles in genomic stability (42).

Further investigation is necessary to determine the mechanistic link between the gene variants identified in pigmentation genes and the rosacea phenotype.

This large study of rosacea symptom severity represents the first genome wide study of this phenotype. Together, these loci provide genetic insight into underlying disease processes in rosacea as a combination of skin type and susceptibility to inflammation. Further work will be needed to identify and replicate causal variants within associated loci and to assess functional variants identified as possible therapeutic targets.

Materials and Methods

Study design and 23andMe cohort

GWAS participants were customers of the genetics company 23andMe, Inc. (Mountain View, CA) who consented to research and answered surveys online according to the 23andMe human subjects protocol, which was reviewed and approved by Ethical & Independent Review Services, a private institutional review board (http://www.eandireview.com).

Responses to eight questions related to rosacea symptoms were added to generate a total symptom severity score, range 0–32 (Table 1). Participants responded to the following questions with a 0 (None), 1 (Slight), 2 (Mild), 3, (Moderate) or 4 (Severe) rating: ‘Do you have persistent redness or rosy appearance of the face?’; ‘Do you have tiny, visible blood vessels on the surface of your facial skin?’; ‘Do you have papules (small red bumps, sometimes filled with white pus) on the face?’; ‘Do you have a recurrent gritty feeling in your eyes or dry eyes?’.

The rosacea risk phenotype used in the GWAS previously reported (9) was defined on the basis of participant’s response to whether a healthcare professional had ever diagnosed them with rosacea. Participants who answered ‘yes’ were defined as ‘cases’ (n = 2618), and those who answered ‘no’ (n = 20334) were defined as ‘controls.’ The symptom severity score in our study was derived at a later point in time and included individuals who responded to a survey related to rosacea symptoms independent of a diagnosis of rosacea and therefore included more subjects (73265). Almost all participants that were studied as part of the GWAS of rosacea risk were included in our analysis but the variable under study in this case is a quantitative trait on the basis of rosacea-related symptoms ranging from ‘0’ (no rosacea symptoms) to ‘32’ (severe rosacea symptoms). Many patients who reported a diagnosis of rosacea in Chang et al. (9) also reported very mild symptomatology (Supplementary Material, Table S1). Moreover, many more patients with moderate to severe symptoms (score 5–32) were added in this analysis.
(n = 9389). Therefore, this sample set, the phenotype and these results are distinct to those described in the earlier paper. A cross-tabulation comparison of the cohorts across the two studies can be found in Supplementary Material, Table S1.

Genotyping and quality control

Samples were genotyped on one of four Illumina genotyping platforms. The V1 and V2 platforms were variants of the Illumina HumanHap550+ BeadChip, including about 25,000 custom SNPs selected by 23andMe, totaling about 560,000 SNPs. The V3 platform was based on the Illumina OmniExpress+ BeadChips, with custom content with a total of about 950,000 SNPs. The V4 platform was a fully custom array with additional coverage of low-frequency coding variation, and about 570,000 SNPs. Participant genotype data were imputed against the September 2013 release of 1000 Genomes Phase1 reference haplotypes. The genotyping platforms, quality control procedures, ancestry determination and imputation protocols have been previously described (43).

GWAS analysis

Participants were restricted to a set of individuals who have >97% European ancestry, determined through an analysis of local ancestry (44). A maximal set of unrelated individuals, as defined by sharing no more than 700 cm IBD, was chosen for each analysis using a segmental identity-by-descent (IBD) estimation algorithm (45).

We tested for genetic associations using linear regression assuming an additive model for allelic effects, adjusting for age, gender, ancestry differences by including the first five GWAS principal components and genotype batch effects by including indicators for genotype platforms in the model: rosacea_severity ~ age + sex + pc.0 + pc.1 + pc.2 + pc.3 + pc.4 + v2_platform + v3_1_platform + v4_platform + genotype. To test associations between HLA allele dosages and rosacea symptom severity, linear regression was performed using the same set of covariates used in the SNP-based GWAS. Separate association tests were performed for each imputed allele. Q-Q plots of the association P values for the study are shown in Supplementary Material, Figure S8.

Significant association was declared at the GWAS level (P < 5.0 × 10⁻⁸). LD score regression was applied to the summary statistics to distinguish between polygenicity and test statistic inflation as implemented in LD Hub (www.ldsc.broadinstitute.org). We report the LD score regression intercept (46). All signals were evaluated for imputation quality (avg r² < 0.5 or min. r² < 0.3) and association evidence provided by adjacent variants or variants in LD with the primary associated marker(s) to ensure robustness of the association.

The results were compared with other GWAS for autoimmune traits using LD score regression (46). Genetic correlation between traits was considered to be significant when the P value was less than 0.05 divided by the number of traits evaluated (P < 0.05/10 = 0.005).

In silico functional assessment

We estimated the probability that an individual SNP is the causal variant using a haplotype structure-based method called PICS for each of the top signals (11). Specifically, we calculated the PICS probability assuming a European reference ancestry for each associated locus and identified the most likely causal SNP or SNPs (i.e. the PICS set) using a >90% probability threshold.

We then completed functional annotation of SNPs in the PICS set in rosacea associated loci.

We annotated coding variation as well as putative regulatory information using HaploReg (47) and used cis-eQTL data available in V6 of GTEx (GTEx Consortium). Active chromatin marks within each locus, excluding the HLA locus due to lack of data, were visualized in the UCSC Genome Browser (48). Tracks containing active histone marks (H3K27ac, H3K9ac, H3K4me3), open chromatin marks (DNase) and transcription factor binding sites (TF ChIP-seq) from multiple cell lines were loaded from the ENCODE Project (49) and the Roadmap Epigenomics Project (50). Represented skin cell lines include dermal fibroblast (NHDF), fetal skin (FSK), fetal abdominal skin (FFSA) and forekin melanocytes (FFM). Represented immune cell lines include B-lymphocytes (GM12878), CD8 primary cells, CD8 memory cells, CD4 primary cells and CD4 memory cells. The cis-eQTL assessment was further refined by evaluating if the rosacea associated variant was concordant with the top cis-eQTL for the identified tissue gene pairs as this would increase confidence that the colocalization of the cis-eQTL and disease associations were not coincidental.

Gene expression analysis

We reanalyzed a recently published rosacea mRNA expression dataset (7). This dataset included samples derived from 29 participants which were grouped as follows: ETR—7 patients, PFR—6 patients, PhR—6 patients, healthy volunteers—10 participants. Patients were diagnosed on the basis of the classification system of the American Rosacea Society (7). Total RNA was extracted from biopsies taken from the participants’ nasolabial fold and hybridized in duplicate to Affymetrix U133 plus 2.0 arrays (7). Raw data files (CEL files) were downloaded from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo; accession GSE65914) (7) and then preprocessed using the RMA (Robust Multi-chip Average) pipeline (51) in combination with the most current reannotated probe set definitions (52). To determine differential mRNA expression, a linear model was fit using an empirical Bayes methodology for more robust variance estimates (53,54). The false discovery rate (FDR) was computed as an adjusted P value (55) to account for multiple testing and a cut-off of 10% FDR, as well as an absolute fold change of 1.5 or greater was used to define differential expression. Clustering analysis was performed as previously described (56).

Supplementary Material

Supplementary Material is available at HMG online.

Data availability statement

The full summary statistics will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Interested investigators should email dataset-request@23andme.com for more information and to apply to access the data.

Acknowledgements

We thank the research participants of 23andMe who answered surveys and consented to participate in research. We also thank the employees of 23andMe who contributed to the development of the infrastructure that made this research possible.
would like to thank Li Li and Chuan Gao from PAREXEL and Robert Scott, Charles Cox, Matthew Nelson, and David Rubenstein at GlaxoSmithKline who contributed expertise in support of this research.

LD Hub: We gratefully acknowledge all the studies and databases that made GWAS summary data available: ADIPOGen (Adiponectin genetics consortium), C4D (Coronary Artery Disease Genetics Consortium), CARDiOGRAM (Coronary Artery Disease Genome wide Replication and Meta-analysis), CKDGen (Chronic Kidney Disease Genetics consortium), dbGAP (database of Genotypes and Phenotypes), DIAGRAM (Diabetes Replication And Meta-analysis), ENIGMA (Enhancing Neuro Imaging Genetics through Meta Analysis), EAGLE (Early Genetics & Life course Epidemiology Eczema Consortium, excluding 23andMe), EGG (Early Growth Genetics Consortium), GABRIEL (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community), GCAN (Genetic Consortium for Anorexia Nervosa), GEFOS (Genetic Factors for OSteoporosis Consortium), GIANT (Genetic Investigation of ANthropometric Traits), GIS (Genetics of Iron Status consortium), GLGC (Global Lipids Genetics Consortium), GPC (Genetics of Personality Consortium), GUGC (Global Urate Status consortium), HaemGen (haemotological and platelet traits genetics consortium), HRGene (Heart Rate consortium), IIBDGC (International Inflammatory Bowel Disease Genetics Consortium), ILCCO (International Lung Cancer Consortium), IMSGC (International Multiple Sclerosis Genetic Consortium), MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium), MESA (Multi-Ethnic Study of Atherosclerosis), PGC (Psychiatric Genomics Consortium), Project MinE consortium, ReproGen (Reproductive Genetics Consortium), SSGAC (Social Genomics Spanning Age and Conditions, Transdisciplinary Research in Cancer of the Lung consortium), UK Biobank. We gratefully acknowledge the contributions of Alkes Price (the systemic lupus erythematosus GWAS and primary biliary cirrhosis GWAS) and Johannes Kettunen (lipids metabolites GWAS).

**Conflict of Interest statement.** J.L.A., M.N.C., L.M.Y.-A., A.G., C.G., D.J.F., J.M.F., D.K.R., M.G.E. and D.M.W. are current or former employees of GlaxoSmithKline, during the term of the research project and may own GlaxoSmithKline stock. J.L.A. and D.J.F. are employees of PAREXEL International. D.A.H. and C.T. are employees of 23andMe, Inc. and own stock or stock options in 23andMe, Inc.

**Funding**

The work was funded by GlaxoSmithKline.

**References**

1. Two, A.M., Wu, W., Gallo, R.L. and Hata, T.R. (2015) Rosacea: introduction, categorization, histology, pathogenesis, and risk factors. *J. Am. Acad. Dermatol.*, 72, 749–758.

2. Fleischer, A.B., Jr. (2011) Inflammation in rosacea and acne: implications for patient care. *J. Drugs Dermatol.*, 10, 614–620.

3. (2000) More women face embarrassment from conspicuous effects of rosacea. Rosacea review. *Newsletter of the National Rosacea Society*. https://www.rosacea.org/tr/2000/summer/article_1.php; date last accessed May 1, 2018.

4. (2005) New survey shows rosacea tends to evolve beyond one subtype. Rosacea review. *Newsletter of the National Rosacea Society*. https://www.rosacea.org/press/archive/20050421.php; date last accessed May 1, 2018.

5. (2016) All about rosacea. *Newsletter of the National Rosacea Society*. https://www.rosacea.org/patients/allaboutrosacea.php; date last accessed May 1, 2018.

6. Steinhoff, M., Schauber, J. and Leyden, J.J. (2013) New insights into rosacea pathophysiology: a review of findings. *J. Am. Acad. Dermatol.*, 69, S15–S26.

7. Buhl, T., Sulk, M., Nowak, P., Buddenkotte, J., McDonald, I., Aubert, J., Carlavan, I., DÉret, S., Reiniche, P. and Rivier, M. (2015) Molecular and morphological characterization of inflammatory infiltrate in rosacea reveals activation of Th1/Th17 pathways. *J. Invest. Dermatol.*, 135, 2198–2208.

8. Aldrich, N., Gerstenblith, M., Fu, P., Tuttle, M.S., Varma, P., Gotow, E., Cooper, K.D., Mann, M. and Popkin, D.L. (2015) Genetic vs environmental factors that correlate with rosacea: a cohort-based survey of twins. *JAMA Dermatol.*, 151, 1213–1219.

9. Chang, A.L.S., Raber, I., Xu, J., Li, R., Spitalre, R., Chen, J., Kiefer, A.K., Tian, C., Eriksson, N.K., Hinds, D.A. and Tung, J.Y. (2015) Assessment of the genetic basis of rosacea by genome-wide association study. *J. Invest. Dermatol.*, 135, 1548–1555.

10. Holmes, A.D. and Steinhoff, M. (2017) Integrative concepts of rosacea pathophysiology, clinical presentation and new therapeutics. *Exp. Dermatol.*, 26, 659–667.

11. Farh, K.K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Belk, S., Shores, N., Whitton, H., Ryan, R.J.H., Shishkin, A.A. et al. (2015) Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518, 337–343.

12. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N. et al. (2013) GTEx consortium. The genotype–tissue expression (GTEx) project. *Nat. Genet.*, 45, 580–585.

13. Visscher, M., Palstra, R.J. and Kayser, M. (2015) Allele-specific transcriptional regulation of IRF4 in melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the IRF4 promoter. *Hum. Mol. Genet.*, 24, 2649–2661.

14. Han, J., Kraft, P., Nan, H., Guo, Q., Chen, C., Qureshi, A., Hankinson, S.E., Hu, F.B., Duffy, D.L., Zhao, Z.Z. et al. (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.*, 4, e1000074.

15. Sturm, R.A., Duffy, D.L., Zhao, Z.Z., Leite, F.P., Stark, M.S., Hayward, N.K., Martin, N.G. and Montgomery, G.W. (2008) A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. *Am. J. Hum. Genet.*, 82, 424–431.

16. Cook, A.L., Chen, W., Thurber, A.E., Smit, D.J., Smith, A.G., Bladen, T.G., Brown, D.L., Duffy, D.L., Pastorio, L., Bianchi-Scarra, G. et al. (2009) Analysis of cultured human melanocytes based on polymorphisms within the SLC45A2/MATP, SLC24A5/NCX5, and OCA2/P loci. *J. Invest. Dermatol.*, 129, 392–405.

17. Wynn, T.A. (2003) IL-13 effector functions. *Annu. Rev. Immunol.*, 21, 425–456.

18. Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.J., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A. et al. (2012) Host-microbe interactions shape the genetic architecture of inflammatory bowel disease. *Nature*, 491, 119–124.

19. Spoendlin, J., Karatas, G., Furlano, R.I., Jick, S.S. and Meier, C.R. (2016) Rosacea, use of tetracycline, and risk of incident Crohn’s disease: a population-based case-control study. *J. Invest. Dermatol.*, 151, 392–405.
inflammatory bowel disease in women. Clin. Gastroenterol. Hepatol., 14, 220–225.
21. Trowsdale, J. and Knight, J.C. (2013) Major histocompatibility complex genomics and human disease. Annu. Rev. Genomics Hum. Genet., 14, 301–323.
22. Weinstock, L.B. and Steinhoff, M. (2013) Rosacea and small intestinal bacterial overgrowth: prevalence and response to rifaximin. J. Am. Acad. Dermatol., 68, 875–876.
23. Bonnellyke, K., Sleiman, P., Nielsen, K., Kreiner-Moller, E., Mercarder, J.M., Belgrave, D., den Dekker, H.T., Husby, A., Sevelsted, A., Faura-Tellez, G. et al. (2014) A genome-wide association study identifies CDH3 as a susceptibility locus for early childhood asthma with severe exacerbations. Nat. Genet., 46, 51–55.
24. Moffatt, M.F., Gut, I.G., Demenais, F., Strachan, D.P., Bouzigon, E., Heath, S., von Mutius, E., Farrall, M., Lathrop, M. and Cookson, W.O.C.M. (2010) A large-scale, consortium-based genome wide association study of asthma. N. Engl. J. Med., 363, 1211–1221.
25. Weidinger, S., Willis-Owen, S.A., Kamatani, Y., Baurecht, H., Morar, N., Liang, L., Edser, P., Street, T., Rodriguez, E., O’Regan, G.M. et al. (2013) A genome-wide association study of atopic dermatitis identifies loci with overlapping effects on asthma and psoriasis. Hum. Mol. Genet., 22, 4841–4856.
26. Hirota, T., Takahashi, A., Kubo, M., Tsunoda, T., Tomita, K., Sakashita, M., Yamada, T., Fujieda, S., Tanaka, S., Doi, S. et al. (2012) Genome-wide association study identifies eight new susceptibility loci for atopic dermatitus in the Japanese population. Nat. Genet., 44, 1222–1226.
27. Hinds, D.A., McMahon, G., Kiefer, A.K., Do, C.B., Eriksson, N., Evans, D.M., St Pourcain, B., Ring, S.M., Mountain, J.L., Francke, U. et al. (2013) A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. Nat. Genet., 45, 907–911.
28. Nair, R.P., Duffin, K.C., Helms, C., Ding, J., Stuart, P.E., Goldgar, D., Gudjonsson, J.E., Li, Y., Tejasvi, T., Feng, B.J. et al. (2009) Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. Nat. Genet., 41, 199–204.
29. Buzney, C.D., Gottlieb, A.B. and Rosmarin, D. (2016) Asthma and atopic dermatitis: a review of targeted inhibition of interleukin-4 and interleukin-13 as therapy for atopic disease. J. Drugs Dermatol., 15, 165–171.
30. Luo, J., Liu, D. and Liu, C.T. (2016) The efficacy and safety of anti-interleukin 13, a monoclonal antibody, in adult patients with asthma: a systematic review and meta-analysis. Medicine (Baltimore), 95, e2556.
31. Man, K. and Kallies, A. (2015) Synchronizing transcriptional control of T cell metabolism and function. Nat. Rev. Immunol., 15, 574–584.
32. Xu, W.D., Pan, H.F., Ye, D.Q. and Xu, Y. (2012) Targeting IRF4 in autoimmune diseases. Autoimmun. Rev., 11, 918–924.
33. Trynka, G., Hunt, K.A., Bockett, N.A., Romanos, J., Mistry, V., Szerel, A., Bakker, S.F., Bardella, M.T., Bhaw-Rosun, L., Castillejo, G. et al. (2011) Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat. Genet., 43, 1193–1201.
34. Praetorius, C., Grill, C., Stacey, S.N., Metcalf, A.M., Corkin, D.U., Robinson, K.C., Van Otterloo, E., Kim, R.S., Bergsteinsdottir, K., Ogmundsdottir, M.H. et al. (2013) A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. Cell, 155, 1022–1033.
35. Eriksson, N., Macpherson, J.M., Tung, J.Y., Hon, I.S., Naughton, B., Saxonov, S., Avey, L., Wojcicki, A., Pe’er, I. and Mountain, J. (2010) Web-based, participant-driven studies yield novel genetic associations for common traits. PloS Genet, 6, e1000993.
36. Zhang, M., Song, F., Liang, L., Nan, H., Zhang, J., Liu, H., Wang, L.E., Wei, Q., Lee, J.E., Amos, C.I. et al. (2013) Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans. Hum. Mol. Genet., 22, 2948–2959.
37. Huber, M. and Hofh, M. (2014) IRF4 at the crossroads of effector T-cell fate decision. Eur. J. Immunol., 44, 1886–1895.
38. Bastonini, E., Kovacs, D. and Ricardo, M. (2016) Skin pigmentation and pigmentary disorders: focus on epidermal/dermal cross-talk. Ann. Dermatol., 28, 279–289.
39. Haddadean, C., Lai, C., Cho, S.Y. and Healy, E. (2015) Variants of the melanocortin-1 receptor: do they matter clinically? Exp. Dermatol., 24, 5–9.
40. Jin, Y., Birlea, S.A., Fain, P.R., Ferrara, T.M., Ben, S., Riccardi, S.L., Cole, J.B., Gowen, K., Holland, P.J., Bennett, D.C. et al. (2012) Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. Nat. Genet., 44, 676–680.
41. Kamaraj, B. and Purohit, R. (2016) Mutational analysis on membrane associated transporter protein (MATP) and their structural consequences in oculocutaneous albinism type 4 (OCA4)—a molecular dynamics approach. J. Cell Biochem., 117, 2608–2619.
42. Sánchez-Tena, S., Cubillos-Rojas, M., Schneider, T. and Rosa, J.L. (2016) Functional and pathological relevance of HERC family proteins: a decade later. Cell Mol. Life Sci., 73, 1955–1968.
43. Hyde, C.L., Nagle, M.W., Tian, C., Chen, X., Paciga, S.A., Wendland, J.R., Tung, J.Y., Hinds, D.A., Perlis, R.H. and Witslow, A.R. (2016) Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat. Genet., 48, 1031–1036.
44. Durand, E.Y., Do, C.B., Mountain, J.L. and Macpherson, J.M. (2014) Ancestry composition: a novel efficient pipeline for ancestry deconvolution. https://doi.org/10.1101/010512.
45. Henn, B.M., Hon, L., Macpherson, J.M., Eriksson, N., Saxonov, S., Pe’er, I. and Mountain, J.L. (2012) Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. PLoS One, 7, e34267.
46. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., Price, A.L. and Neale, B.M. (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet., 47, 291–295.
47. Ward, L.M. and Kellis, M. (2016) HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res., 44, D877–D881.
48. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. Genome Res., 12, 996–1006.
49. Rosenberg, K.R., Sloan, C.A., Malladi, V.S., Dreszer, T.R., Learned, K., Kirkup, V.M., Wong, M.C., Maddren, M., Fang, R., Heitner, S.G. et al. (2013) ENCODE data in the UCSC genome browser: year 5 update. Nucleic Acids Res., 41, D56–D63.
50. Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R. et al. (2010) The NIH roadmap epigenomics mapping consortium. Nat. Biotechnol., 28, 1045–1045.
51. Gautier, L., Cope, L., Bolstad, B.M. and Irizarry, R.A. (2004) affy – analysis of Affymetrix GeneChip data at the probe level. Bioinformatics, 20, 307–315.
52. Dai, M., Wang, P., Boyd, A.D., Kostov, G., Athey, B., Jones, E.G., Bunney, W.E., Myers, R.M., Speed, T.P., Akil, H. et al. (2005) Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res., 33, e175.
53. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res., 43, e47.
54. Smyth, G.K. (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol., 3, 1. 10.2202/1544-6115.1027.
55. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol., 7, 289–300.
56. Freudenberg, J.M., Joshi, V.K., Hu, Z. and Medvedovic, M. (2009) CLEAN: Clustering Enrichment Analysis. BMC Bioinformatics, 10, 234. doi: 10.1186/1471-2105-10-234.
57. Sulem, P., Gudjardarsson, S., Johannsdottir, J., Helgadottir, A., Rafnar, T., Magnusson, K.F., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G. et al. (2007) Genetic determinants of human height. Science, 316, 836–843.
58. Eberhart, J., Kjaer, K.W. and Hansen, L. (2008) Blue eye color in a Danish nationwide sample: a linkage analysis identifies HERC2 as a human iris color gene. Heredity, 100, 920–925.
59. van Meurs, J.B., Pare, G., Schwartz, S.M., Hazra, A., Tanaka, T., Vermeulen, S.H., Cotlarciuc, I., Yuan, X., Målsarstig, A., Bandinelli, S. et al. (2013) Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease. Am. J. Clin. Nutr., 98, 668–676.
60. Paternoster, L., Stanøe, T., Waage, J., Baurecht, H., Hotze, M., Strachan, D.P., Curtin, J.A., Benneykë, K., Tian, C., Takahashi, A. et al. (2015) Multi-ancestry genome-wide association study of cutaneous melanoma. Nature, 521, 1449–1456.
61. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T. et al. (2015) Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat. Genet., 47, 979–986.
62. Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., von Berg, A., Bufe, A., Rietschel, E. et al. (2007) Genetic variants regulating ORMEL3 expression contribute to the risk of childhood asthma. Nature, 448, 470–473.
63. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S. et al. (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature, 506, 376–381.
64. International Multiple Sclerosis Genetics Consortium; Wellcome Trust Case Control Consortium 2, Sawyer, S., Hellenthal, G., Pirinen, M., Spencer, C.C., Patzopoulos, N.A., Moutsianas, L., Dílthey, A., Su, Z., Freeman, C., Hunt, S.E. et al. (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature, 476, 214–219.
65. Bentham, J., Morris, D.L., Graham, D.S.C., Pinder, C.L., Tombrelos, P., Behrens, T.W., Martin, J., Fairfax, B.P., Knight, J.C., Chen, L. et al. (2015) Genetic association analyses identify aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat. Genet., 47, 1457–1464.
66. Cordell, H.J., Han, Y., Mells, G.F., Li, Y., Hirschfeld, G.M., Greene, C.S., Xie, G., Juran, B.D., Zhu, D., Qian, D.C. et al. (2015) International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. Nat. Commun., 6, 8019.