Genome-wide association study in frontal fibrosing alopecia identifies four susceptibility loci including HLA-B*07:02

Christos Tziotzios et al.

Frontal fibrosing alopecia (FFA) is a recently described inflammatory and scarring type of hair loss affecting almost exclusively women. Despite a dramatic recent increase in incidence the aetiopathogenesis of FFA remains unknown. We undertake genome-wide association studies in females from a UK cohort, comprising 844 cases and 3,760 controls, a Spanish cohort of 172 cases and 385 controls, and perform statistical meta-analysis. We observe genome-wide significant association with FFA at four genomic loci: 2p22.2, 6p21.1, 8q24.22 and 15q2.1. Within the 6p21.1 locus, fine-mapping indicates that the association is driven by the HLA-B*07:02 allele. At 2p22.1, we implicate a putative causal missense variant in CYP1B1, encoding the homonymous xenobiotic- and hormone-processing enzyme. Transcriptomic analysis of affected scalp tissue highlights overrepresentation of transcripts encoding components of innate and adaptive immune response pathways. These findings provide insight into disease pathogenesis and characterise FFA as a genetically predisposed immuno-inflammatory disorder driven by HLA-B*07:02.

Correspondence and requests for materials should be addressed to C.T. (email: christos.tziotzios@kcl.ac.uk) or to J.A.M. (email: john.mcgrath@kcl.ac.uk). A full list of authors and their affiliations appears at the end of the paper.
Frontal fibrosing alopecia (FFA) is a recently reported lichenoid and scarring inflammatory skin disorder associated with widespread cutaneous inflammation and irreversible hair loss, which occurs predominantly in women of post-menopausal age (Fig. 1). Since FFA was first identified by Kossard in 1994, there has been rapid increase in reported incidence culminating in intense clinical and public interest in the condition. FFA is often referred to as a dermatological epidemic condition. FFA is considered to be a clinical sub-variant of lichen planus, a more common inflammatory skin condition of unresolved aetiology, while also representing a variant of the prototypic primary lymphocytic cicatrical (or scarring) alopecia lichen planopilaris (LPP). A key molecular event in the pathology of scarring hair loss has been postulated to be the immune privilege collapse at the level of the immunologically shielded hair follicle bulge, which is home to epithelial hair follicle stem cells (eHFSC): T-cell mediated inflammatory presence culminates in stem cell apoptosis and irreversible alopecia. Dissection of the genetic basis of FFA and its interplay with environmental risk factors, therefore, could provide insight into the molecular profile of lichenoid inflammation, scarring and mechanisms of immune privilege collapse. Furthermore, the identification of environmental triggers that interact with FFA genetic susceptibility loci could ultimately contribute to disease prevention by avoidance of exposures in genetically predisposed individuals (Supplementary Note 1).

To date there have been no systematic investigations into the molecular genetic basis of FFA or any other lichenoid inflammatory disorder. We hypothesise that common genetic variation contributes to FFA susceptibility and undertake a genome-wide association study and meta-analysis of two independent European cohorts of females with FFA and controls and investigate transcriptomic and metabolomic involvement in the disease.

**Results**

**Genome-wide association study.** We undertook a genome-wide association analysis across 8,405,903 common variants in a UK cohort of 844 FFA female cases and 3760 female controls. Inspection of the quantile-quantile plot indicated adequate control of confounding bias ($\lambda_{\text{GC_MAP}} = 1.03$; Supplementary Table 2; Supplementary Figure 3). We observed genetic variants with genome-wide significant association ($P < 5.0 \times 10^{-8}$) with FFA at three genomic loci; 6p21.1, 2p22.2 and 15q26.1 (Table 1). We estimate the genome-wide SNP heritability for FFA as 46.66% ($SE = 3.00\%$).

In an attempt to replicate the observed associations at each of the three loci and identify additional FFA susceptibility loci, we performed a genome-wide association study across 7,964,651 common variants in our independent Spanish cohort comprising 172 affected females and 385 controls. We observed allelic associations with FFA at each of the three loci that had been implicated in FFA susceptibility in the UK cohort. The direction and magnitude of the effect of these associations was consistent between the UK and Spanish cohorts (Table 1). To identify further loci harbouring variation contributing to FFA risk we performed a statistical meta-analysis of the association summary statistics from the UK and Spanish cohorts. This revealed a single additional risk locus at 8q24.22, again with a consistent direction and magnitude of effect in both studies (Fig. 2 and Table 1) and a number of loci at which there is suggestive evidence of association ($P < 5 \times 10^{-5}$; Supplementary Table 3).

We sought to further investigate the allelic basis for the observed FFA association at 6p21.1, which is located within the MHC region. We undertook imputation of classical MHC Class I alleles and evaluated the association of each allele with FFA. The strongest evidence of association was observed for the HLA-B allele HLA-B*07:02 ($P_{\text{Meta}} = 9.44 \times 10^{-117}$, OR $= 5.22$ (4.53–6.01); Supplementary Table 4), indicating that this is the most likely classical HLA allele to be underpinning the observed SNP associations in this region. Although full characterization of

**Fig. 1** Clinical features of frontal fibrosing alopecia. Scalp with frontal hairline recession (a) involving the temporal areas bilaterally (b), as well as eyebrows (c). Histopathology (d) shows two hair follicles with focal interface changes, and a moderately dense perifollicular lymphoid cell infiltrate with perifollicular fibrosis, characteristic of FFA ($\times 200$).
the role of HLA genes in FFA is complicated by the complex linkage disequilibrium structure across the region, a sequential conditional analysis with both SNPs and imputed HLA alleles indicates that there may be at least a further two independent HLA-alleles that contribute independently to disease risk (Supplementary Figure 3 and Supplementary Tables 4A and 4B).

To further investigate causal genes and alleles at the three remaining FFA susceptibility loci, we performed Bayesian fine-mapping of the association signals. This process identified a single putative causal variant with a posterior probability >0.5 of being the causal variant underlying the association signal at two of the three loci (Fig. 3 and Supplementary Table 5). At 2p22.2, rs1800440, is likely to be the causal variant underlying the association signal with a posterior probability >0.5 of being the causal variant with a posterior probability of 0.98 (Fig. 4). The FFA protective allele is a missense allele (c.1358A>G p. = Asn453Ser) in the CYP1B1 gene, which introduces a serine residue in the haem binding domain of the enzyme. In silico pathogenicity prediction tools predict that this allele has a deleterious effect on the function of the protein (SIFT = 0.015; CADD = 32)13,14, which is corroborated by published functional investigations of the p.Asn453Ser substitution15. At the 8q24.22 locus, rs760327 is the most likely causal allele (posterior probability = 0.68). The variant is located within intron 1 of the ST3GAL1 gene encoding the homonymous galactoside sialyltransferase enzyme, which has been studied in the context of T cell homeostasis16,17. At 15q26.1, statistical fine mapping was unable to clearly resolve the causal variant at this locus (Supplementary Table 5), though the most likely causal variant rs34560261 (posterior probability = 0.4) is located within intron 1 of SEMA4B. Co-localization with skin eQTLs (Supplementary Figure 5) provides evidence that the same variant(s) underlying the observed FFA association at this locus are also associated with variation in the expression of SEMA4B in the skin (Pcoloc = 0.99), providing support that SEMA4B may be the causal gene at this locus.

PLasma metaboLomic analysis. At the 2p22.2 locus statistical fine mapping of the causal allele to a functional missense variant in CYP1B1 implicates variation in xenobiotic and endogenous hormone metabolism18–22 as a potential mechanism influencing disease susceptibility. To evaluate if there are systematic differences in metabolomic profiles between FFA cases and controls we compared levels of plasma metabolites in 52 treatment-naïve FFA cases and 35 ethnicity-, gender-, age- and BMI-matched healthy controls (Supplementary Figure 6). We did not observe differences in the distribution of individual metabolite levels between cases and controls of the magnitude detectable by this experiment following multiple-testing correction (Supplementary Table 6), nor did we observe any enrichment of xenobiotic or endogenous hormone metabolites in the extremes of the distribution of observed mean differences.

Transcriptomic analysis. To further investigate genes and biological pathways involved in FFA pathobiology we performed transcriptome sequencing in lesional scalp skin from seven treatment-naïve postmenopausal FFA cases and compared to transcriptome profiles from scalp skin in seven

<table>
<thead>
<tr>
<th>Locus</th>
<th>Gene</th>
<th>Position (hg19)</th>
<th>SNP ID</th>
<th>RA</th>
<th>PA</th>
<th>RAF Cases</th>
<th>RAF Controls</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>2p22.2</td>
<td>CYP1B1</td>
<td>38,298,139</td>
<td>rs1800440</td>
<td>T</td>
<td>C</td>
<td>0.87</td>
<td>0.81</td>
<td>1.62 (1.38–1.90)</td>
<td>5.89x10^-9</td>
<td>1.81 (1.28–2.58)</td>
<td>0.00090</td>
<td>1.65 (1.43–1.91)</td>
<td>2.44x10^-11</td>
</tr>
<tr>
<td>6q21.1</td>
<td>HLA-B</td>
<td>31,320,562</td>
<td>rs2523816</td>
<td>T</td>
<td>C</td>
<td>0.47</td>
<td>0.19</td>
<td>4.69 (4.07–5.40)</td>
<td>8.52x10^-10</td>
<td>4.97 (3.52–7.02)</td>
<td>8.09x10^-9</td>
<td>4.73 (4.56–5.09)</td>
<td>7.60x10^-4</td>
</tr>
<tr>
<td>8q24.22</td>
<td>ST3GAL1</td>
<td>118,114,278</td>
<td>rs760327</td>
<td>G</td>
<td>C</td>
<td>0.46</td>
<td>0.39</td>
<td>1.32 (1.18–1.47)</td>
<td>1.18x10^-6</td>
<td>1.50 (1.16–1.97)</td>
<td>0.00367</td>
<td>1.34 (1.21–1.49)</td>
<td>2.15x10^-6</td>
</tr>
<tr>
<td>15q26.1</td>
<td>SEMA4B</td>
<td>90,734,426</td>
<td>rs34560261</td>
<td>T</td>
<td>C</td>
<td>0.22</td>
<td>0.17</td>
<td>1.52 (1.32–1.76)</td>
<td>8.47x10^-9</td>
<td>1.51 (1.03–2.21)</td>
<td>0.03257</td>
<td>1.52 (1.22–1.74)</td>
<td>8.12x10^-10</td>
</tr>
</tbody>
</table>

Each SNP was tested for association by logistic regression using an additive regression model; total N = 5161 biologically independent subjects (Ncases = 1044 and Ncontrols = 4145).
healthy matched controls. Differences in transcript abundance between cases and controls in these bulk tissue samples was observed for 117 genes with a greater representation of transcripts from 80 genes in affected tissue and 37 genes with reduced representation of their respective transcripts (Supplementary Tables 7–9; Supplementary Figure 7). Of these, only C2, within the MHC at 6p21.33, is located within 1 Mb of any of the FFA associated loci and only two of the 117 genes (CCL19 and EPSTI1) are located within 1 Mb of any variant with moderate evidence of association ($P < 5 \times 10^{-5}$) with FFA. Investigation of the enrichment of gene sets and pathways indicate that immune genes are over-represented amongst the differentially expressed genes (DEGs) (Supplementary Tables 8–12). Notably, four of the 10 most extreme DEGs are genes that have an established role in the interferon gamma (IFNγ) pathway.
Discussion

We have identified four genomic loci, at which genetic variation is robustly associated with the lichenoid inflammatory and scarring dermatosis FFA in two independent cohorts of European ancestry.

The strongest effect on FFA susceptibility is observed at 6p21.1 which is located within the MHC region. Through imputation of classical HLA alleles we implicate the Class I allele HLA-B*07:02 as conferring a five-fold increase in risk of FFA. The highly polymorphic HLA genes and their encoded proteins play a key role in...
role in self and non-self immune recognition and are known to
determine susceptibility to numerous infectious and auto-
immune disorders23. HLA-B^07:02 itself has previously been
reported to be associated with HIV progression but has not been
implicated in the susceptibility to human disease24. The hair
follicle bulge region and the outer root sheath express low levels
of HLA-A, HLA-B and HLA-C and these are key to rendering
immune privilege25,26. HLA-B^07:02 may facilitate the process of
hair follicular autoantigen presentation culminating in the auto-
inflammatory lymphocytic destruction of the hair follicle bulge
and its resident epithelial hair follicle stem cells. Investigation of
differentially expressed genes between affected and unaffected
tissue further highlights the importance of genes encoding the
components of innate and adaptive immunity and, notably,
the IFNγ pathway, which is an important regulator of antigen
presentation. Also relevant to a putative role of T cell dysfunction
in FFA, the lead variant at the 8q24.22 locus is located in intron 1
of ST3GAL1, which encodes a membrane bound sialyltransferase.
Changes in cell surface glycan structures have been implicated in
human T cell lymphocyte activation and maturation27,28 and
ST3GAL1 itself has been implicated in immunity by home-
ostatically controlling CD8^+ T cells6,17.

At 2p22.2 we observe strong evidence that the causal variant
underlying the association at this locus is a missense variant in
CYP1B1, a ubiquitously expressed gene encoding the Cytochrome
P450 1B1 microsomal enzyme, also known as xenobiotic mono-
oxynogenase and aryl hydrocarbon hydroxylase. This enzyme con-
tributes to the oxidative metabolism of oestradiol and oestrogen
to their corresponding hydroxylated catechol oestrogen29–30.
Functional investigation of allelic variation in CYP1B1 has shown
that the FFA protective p.Asn453Ser allele increases the rate of
CYP1B1 degradation leading to reduced intracellular CYP1B1
levels15. Given the established role of CYP1B1 in sex hormone
metabolism, alongside the female preponderance of FFA and its
rapid and recent increase in incidence, it is plausible to speculate
that an increase in exposure of females to a CYP1B1 substrate,
whether endogenous or exogenous, may contribute to the
development of FFA. The temporal relationship between
the introduction of oral contraceptives in the 1960s and the
appearance of FFA in the published literature in the 1990s should
be fully explored with a well-designed gene-environment inter-
action study. Nevertheless, no striking differences in such sub-
strates nor any other metabolites were identified in our
metabolomic study although this may reflect the limited power of
this initial investigation to observe more subtle disruption of this
metabolic pathway. It should be noted that CYP1B1 has also been
implicated in human immune cell regulation31,32 and the
potential that FFA susceptibility at the 2p22.2 is mediated
through immune pathways cannot be excluded.

In summary, this exploration of the molecular genetics of
FFA susceptibility we identify common alleles at four genomic
loci that contribute to disease risk. The putative biological impact
of this genetic variation indicates that the disease is a complex
auto-immuno-inflammatory trait underpinned by risk alleles in MHC
Class I molecule-mediated antigen processing and T cell home-
ostasis and function. The insight into the pathobiology of FFA
from the genetic susceptibility loci combined with the observation
that there is an increase in transcripts encoding components of
the IFNγ pathway in affected scalp tissue suggest that drugs such as
JAK inhibitors, some of which have already proved effective in
alopecia areata33 and trialled in lichen planopilaris34, may prove
to be efficacious for FFA.

Methods

Clinical resource. Ethical approval was granted by the Northampton NRES
Committee, UK (REC 15/EM/0273) and the study was conducted in accordance
with the Declaration of Helsinki (https://www.wma.net/policies-post/wma-
declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-
participants/) and the Declaration of Helsinki—extended cohort of highly clinically
consistent female cases of classic FFA, diagnosed by specialist dermatology clinics
in the UK and Spain. All recruited cases were of European ancestry and diagnosed with
FFA on the basis of the following clinical and histopathological features (recently
proposed as diagnostic criteria)35: (1) cicatricial areolar involvement of the
frontal, temporal/parietal hair margin; (2) bilateral eyebrow loss; (3) clinical, tri-
choscopic (or histological) evidence of lichenoid perifollicular inflammatory pre-
sence; (4) facial or body hair loss; (5) absence of multifocal scalp involvement
and other signs suggestive of classic LPP or its Graham-Little-Piccardi-Lasseur
variant.

All research participants provided written informed consent for participation in
the study. The individual depicted in Fig. 1 provided informed consent for
publication of her clinical images.

Genotyping and genome-wide association analysis. For the UK cohort,
genome-wide genotyping of cases was undertaken using Infinium Ex-
pressExome BeadChip array (Illumina) and an unselected female control
cohort from the English Longitudinal Study of Aging (ELSA) project (http://www.pro-
jectoinma.ac.uk), genotyped on the Infinium Omni2.5M BeadChip array (Illumina).
We retained variants that were assayed with the same probe design on both
genotyping arrays and excluded variants with a call rate of <99% or which deviated
from Hardy–Weinberg Equilibrium (P < 10^–8). Individuals with a call rate of <99%
or a genotyping concordance heterozygosity within the SNP was excluded. A subset
in linkage equilibrium (r^2 < 0.2 between each pair) was used to evaluate relatedness
between individuals using the KING software package (KING; version 2.1.1).

Individuals were thus removed from the study such that no two individuals had
estimated relatedness closer than third degree (Kinship coefficient > 0.0442).

In the Spanish cohort, FFA case genotyping was performed using the Infinium
OmniExpressExome BeadChip array (Illumina). Genotypic data for unaffected
case controls were obtained from 189 individuals from the ENMA Ambiente
(INMA) project (Valencia, Sabadell and Menorca, Spain http://www.proyectoinma.
org) genotyped on the Omni-Quad BeadChip array (Illumina). Genotype calling,
quality control and imputation were undertaken following the same protocol as
described for the UK cohort across all variants that are assayed on both genotyping
arrays with the same probe design.

Following quality control, genome-wide imputation was performed for both
cohorts using the Michigan Imputation Server, with the Haplotype Reference
Consortium (HRC) reference haplotype panel (www.haploype-reference-
corpus.org). All variants with an imputation info score < 0.7 or a minor allele
frequency of <0.05 were excluded from downstream analysis. This process of data
generation and QC resulted in a combined total of 7,039,930 variants successfully
imputed or genotyped in a combined total of 1,316 cases and 4,145 controls.

Genome-wide association testing was performed on all variants with MAF >
0.005 using a logistic Wald association test (EPACTS), including the first five
components as covariates. Association testing was performed based on
46,789 variants performed and individuals outlying the main cluster (implying non-European ancestry) were also
excluded from further analysis.

In the Spanish cohort, FFA case genotyping was performed using the Infinium
OmniExpressExome BeadChip array (Illumina). Genotypic data for unaffected
case controls were obtained from 189 individuals from the ENMA Ambiente
(INMA) project (Valencia, Sabadell and Menorca, Spain http://www.proyectoinma.
org) genotyped on the Omni-Quad BeadChip array (Illumina). Genotype calling,
quality control and imputation were undertaken following the same protocol as
described for the UK cohort across all variants that are assayed on both genotyping
arrays with the same probe design.

Following quality control, genome-wide imputation was performed for both
cohorts using the Michigan Imputation Server, with the Haplotype Reference
Consortium (HRC) reference haplotype panel (www.haploype-reference-
corpus.org). All variants with an imputation info score < 0.7 or a minor allele
frequency of <0.05 were excluded from downstream analysis. This process of data
generation and QC resulted in a combined total of 7,039,930 variants successfully
imputed or genotyped in a combined total of 1,316 cases and 4,145 controls.

Genome-wide association testing was performed on all variants with MAF >
0.005 using a logistic Wald association test (EPACTS), including the first five
components as covariates. Association testing was performed based on
46,789 variants performed and individuals outlying the main cluster (implying non-European ancestry) were also
excluded from further analysis.

Causal variant identification and evaluation. For the UK and Spanish case-
cohorts, imputation of classical HLA alleles to two- and four-digit reso-
lution was performed with the SNP2HLA tool, based on the genotypes of 1,297
single nucleotide polymorphisms (SNPs) from phase 3 of the 1000 Genomes
dataset after conditioning on all independently-associated HLA alleles.85

In the UK cohort, dosage-based association testing was performed in PLINK v1.9
for all 208 alleles that were well imputed (r^2 > 0.9), using a logistic regression
framework that included the same covariates as the full GWAS39. Replication of
specific variants of interest was undertaken in the Spanish replication cohort in
the same test. Association meta-analysis was performed using the
meta package in R30. To test for multiple independent association signals, stepwise
conditional analysis was performed: at each round of testing, the dosage of the
HLA allele achieving the lowest discovery phase association p-value in the previous
round was added to the list of covariates. This process was iterated until no allele
HLA allele achieving the lowest discovery phase association p-value in the previous
round was added to the list of covariates. This process was iterated until no allele
association with the same probe design.

Causal variant identification and evaluation. For the UK and Spanish case-
cohorts, imputation of classical HLA alleles to two- and four-digit reso-
lution was performed with the SNP2HLA tool, based on the genotypes of 1,297
single nucleotide polymorphisms (SNPs) from phase 3 of the 1000 Genomes
dataset after conditioning on all independently-associated HLA alleles.85

In the UK cohort, dosage-based association testing was performed in PLINK v1.9
for all 208 alleles that were well imputed (r^2 > 0.9), using a logistic regression
framework that included the same covariates as the full GWAS39. Replication of
specific variants of interest was undertaken in the Spanish replication cohort in
the same test. Association meta-analysis was performed using the
meta package in R30. To test for multiple independent association signals, stepwise
conditional analysis was performed: at each round of testing, the dosage of the
HLA allele achieving the lowest discovery phase association p-value in the previous
round was added to the list of covariates. This process was iterated until no allele
association with the same probe design.
locus, we constructed a credible set of variants considered most likely to be causal based on evidence for association as quantified by their Bayes factor\(^5\).

In order to explore the correlation between genetic variation and tissue expression, eQTL colocalization analysis was performed between the observed FFA association signals and sun-exposed skin cis-eQTL data from the Gene-Tissue Expression (GTEX) project database\(^4\). Candidate skin eQTLs were identified by looking into whether any variant within the FFA risk loci was associated with varied expression of genes. Bayesian testing for colocalization between the FFA association signal and the skin eQTL signal was undertaken using a set of overlapping variants for the two datasets, employing the R package coloc tool\(^4\), with a defined prior probability of colocalization of \(P = 10^{-5}\).

Heritability estimation. FFA heritability estimated by genome-wide SNPs (MAF > 0.01) was estimated using the genomic relationship matrix (GREML) approach, implemented in the software tool package GCTA\(^45\). Heritability estimates were expressed on the liability scale using an estimated prevalence of FFA of one in five thousand.

Transcriptomic analysis. We performed transcriptome profiling with RNA-seq data from scalp skin from seven cases of European ancestry and seven matched controls (Supplementary Table 1). All seven cases were clinically evaluated prior to obtaining skin biopsy from actively inflamed parietal scalp skin for histologic confirmation of FFA. Samples from cases were only subjected to downstream processing if they were confirmed to be actively inflamed upon histological evaluation. Macroscopically unremarkable parietal scalp skin was also harvested from healthy controls undergoing plastic facial surgery and all control tissue specimens were also examined macroscopically and confirmed to be histologically normal (Supplementary Figure 1). Total RNA was isolated from each tissue sample using the RNeasy Plus Universal kit (Qiagen, Valencia, CA, USA) as per the manufacturer’s protocol and instructions. Samples with RNA Integrity Number (RIN) < 8 were rejected from further processing.

Whole transcriptome RNAseq libraries were prepared using the Agilent SureSelect strand-specific RNA library preparation kit (Agilent, Santa Clara, CA, USA) and multiplexed sequencing was performed on the HiSeq 2500 platform (Illumina, San Diego, CA, USA). The clean read of the raw transcriptomic data files were conducted using an established analytical pipeline (Supplementary Figure 2). EdgeR software package in the (R-based) Bioconductor was utilized to undertake differential expression analysis, following the trimmed mean of M-values (TMM) normalization protocol and instructions. Samples with RNA Integrity Number (RIN) < 8 were rejected from further processing.

Plasma metabolomic analysis. Fifty-two treatment-naive post-menopausal female FFA cases of European ancestry (median age 64; mean BMI 24.7) and 35 matched controls (median age 58; mean BMI 24.5) were recruited. Peripheral venous blood was collected and centrifuged (at 1300 g for 15 min) to separate plasma, which was aliquoted and stored at \(-80^\circ\text{C}\) until required for further analysis. Metabolomic profiling of samples was undertaken by Metabolon (Durham, NC, USA) by subjecting plasma samples to methanol extraction prior to splitting into aliquots for analysis by ultra-high pressure liquid chromatography/mass spectrometry (UHPLCMS)\(^46\). Metabolites were identified by automated comparison of mass to reference library of chemical standards followed by visual inspection for quality control\(^45\). Missing values were presumed to be below the limits of detection and were therefore imputed to the compound minimum. Metabolomic data analysis was undertaken, having accounted for medicinal drug-related by-products and discarded unnamed molecules. We constructed a heat map illustrating group differences at individual and group level. Using MetaboAnalyst 4.0\(^4\) and univariate comparison of abundance of 947 named metabolites (<500Da) metabolites between cases and controls was performed using the Mann-Whitney U test.

Data availability

Full meta-analysis summary statistics are available at the European Genome phenome archive (EGA) under the collection ID EGAS00001003460. All raw and processed transcriptomic data are available at the Gene Expression Omnibus (GEO) under the collection ID GSE125733. All other data that supports the findings of this study are available from the corresponding author upon reasonable request.

References

Acknowledgements

We acknowledge financial support from the Department of Health via the National Institute for Health Research (NIHR) Rare Diseases Translational Research Collaboration (NIHR-RD TBC) and by the NIHR Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust and King’s College London. Samples and data from Spanish patients included in this study were provided by the Hospital Universitario Ramón y Cajal-IRYCIS Biobank integrated in the Spanish Hospital Platform Biobanks Network and were processed following standard operation procedures with appropriate approval of the Ethical and Scientific Committees. We thank the study participants in the UK and Spain for their help. Data used for replication in this research was provided by the INMAC—Inflam Sp´a de Ambiente [Environment and Childhood] Project (www.proyectoinma.org), which is supported in part by funds. We are thankful to the English Longitudinal Study of Ageing (ELSA) for providing population control genotyping data. We thank the NIHR Rare Genetic Disease Research Consortium Agreement Team, especially Gillian Bothwick, for their help with setting up multiple UK research participating sites. We thank Dr. Sophia Karagianis, Dr. Francesca Capan, Prof Jemma Mellerio and Dr. Ian White for their advice and guidance. We are thankful to Prof Kathryn Lewis, Dr. Raquel Iniesta, Dr. Ken Hanscombe and Dr. Leo Bottolo for their input and advice with the statistical analysis aspects of the metabolomic exploration. We thank the numerous research assistants and nurses, especially Teena Mackenzie, Sophie Devine, Ruth Joslyn, Siona Bayschpolec, Anne Thomson, Pauline Buchanan and Caroline White for their help with recruitment. Very special thanks go to Dr. Dimitra Drista for her continuous support and input throughout this work.

Author contributions


Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-09117-w.

Competing interests: The authors declare no competing interests.

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

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019

Christos Tzitzios, Christos Petridis, Nick Dand, Chrysanthi Ainali, Jake R. Saklatvala, Venu Pullabhatla, Alexandros Onoufriadis, Rashida Pramanik, David Baudry, Sang Hyuck Lee, Kristie Wood, Lu Liu,

1St. John’s Institute of Dermatology, King’s College London, London, Guy’s Hospital, London SE1 9RT, UK. 2Department of Medical and Molecular Genetics, King’s College London, Guy’s Hospital, London SE1 9RT, UK. 3National Institute for Health Research Barts Health Biomedical Research Centre, St. John’s Institute of Dermatology, King’s College London, London SE1 9RT, UK. 4NIHR GSTF/KCL Comprehensive Biomedical Research Centre, Guy’s & St. Thomas’ NHS Foundation Trust, London SE1 9RT, UK. 5NIHR Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust (SLaM) & Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King’s College London, Denmark Hill Campus, London SE5 8EF, UK. 6National Diagnostic EB Laboratory, St. Thomas Hospital, London SE1 7EH, UK. 7Metabolon, Inc., Morrisville, NC 27560, USA. 8Alan Llyell Centre for Dermatology, Queen Elizabeth University Hospital, Glasgow G51 4TF, UK. 9The Dermatology Centre, The University of Manchester, Salford Royal NHS Foundation Trust, Salford M6 8HD, UK. 10Department of Dermatology, Barnet General Hospital, Royal Free Foundation Trust, London EN5 3DJ, UK. 11Department of Dermatology, Solihull Hospital, Solihull B91 2JL, UK. 12Department of Dermatology, Brighton and Sussex University Hospitals NHS Trust, Brighton BN2 3EW, UK. 13Department of Dermatology, Royal Hallamshire Hospital, Sheffield S10 2JF, UK. 14Department of Dermatology, Royal Free Hospital, London NW3 2QG, UK. 15Department of Dermatology, Royal London Hospital, Barts Health NHS Trust, London E1 1BB, UK. 16Department of Dermatology, Imperial College Healthcare NHS Trust, London W12 0HS, UK. 17Department of Dermatology, University Hospitals of Leicester, Leicester LE3 9QP, UK. 18Department of Dermatology, St. John’s Institute of Dermatology, St. Thomas Hospital, London SE1 7EH, UK. 19Department of Dermatology, University Hospitals of Leicester, Leicester LE3 9QP, UK. 20Department of Dermatology, University Hospitals of Leicester, Leicester LE3 9QP, UK. 21Department of Dermatology, The Churchill Hospital, Oxford OX3 7LP, UK. 22Trichology Unit, Dermatology Department, Ramon Y Cajal Hospital, University of Alcalá, IRYCS, Madrid 28034, Spain. 23Hospital Universitario Fundacion Jimenez Diaz, Madrid 28040, Spain. 24Department of Dermatology, Hospital Donostia, San Sebastian 20014, Spain. 25Hywel Dda University Health Board, Cardiff SA31 3BB, UK. 26Department of Dermatology, Norfolk & Norwich University Hospitals NHS Foundation Trust, Norwich NR4 7UY, UK. 27Department of Dermatology, Ninewells Hospital, Dundee DD1 9SY, UK. 28Department of Dermatology, Queen Margaret Hospital, Dunfermline KY12 0SU, UK. 29Department of Dermatology, County Durham and Darlington NHS Foundation Trust, Darlington DL3 6HX, UK. 30Department of Dermatology, Royal Victoria Hospital, Belfast BT12 6BA, UK. 31Department of Dermatology, Whiteabbey Hospital, Northern Health & Social Care Trust, Co Antrim BT37 9RH, UK. 32Department of Dermatology, Chapel Allerton Hospital, Leeds teaching Hospitals NHS Trust, Leeds LS7 4SA, UK. 33Department of Dermatology, Surrey and Sussex Healthcare NHS Trust, Surrey RH1 5RH, UK. 34Department of Dermatology, University Hospitals Bristol NHS Foundation Trust, Bristol BS2 8WH, UK. 35Department of Dermatology, Gloucestershire Hospitals NHS Foundation Trust, Gloucester GL1 3NN, UK. 36Department of Dermatology, King’s College London, King’s College London NHS Foundation Trust, King’s College London, London SE1 9RT, UK. 37Department of Dermatology, Royal Victoria Infirmary, Newcastle upon Tyne NE1 4LP, UK. 38Department of Dermatology, University of Melbourne, Melbourne VIC 3010, Australia. 39The London Skin and Hair Clinic, London WC1V 7DN, UK. 40Department of Dermatology, School of Clinical Medicine, Nelson R Mandela School of Medicine, Durban 4001, South Africa. 41Craniofacial Surgery Unit, Chelsea & Westminster Hospital, London SW10 9NH, UK. 42Genetics Unit, Dexeus Women’s Health, Barcelona 08028, Spain. 43Department of Dermatopathology, St. John’s Institute of Dermatology, St. Thomas’ Hospital, London SE1 7EH, UK. 44The Department of Twin Research & Genetic Epidemiology, King’s College London, London SE1 7EH, UK. 45Centre for Stem Cells and Regenerative Medicine, King’s College London, Guy’s Hospital, London SE1 9RT, UK. These authors contributed equally: Michael A. Simpson, John A. McGrath.