Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: McCormack M, Alfirevic A, Bourgeois S, et al. HLA-A*3101 and carbamazepine-induced hypersensitivity reactions in Europeans. N Engl J Med 2011;364:1134-43.

Supplementary Appendix

Methods

1) **Recruiting centres**
2) **Definition of Stevens Johnson Syndrome**
3) **Data access for healthy controls**
4) **Details on cleaning of Illumina data**
5) **Details of imputation of HLA alleles in EPIGEN Consortium samples**
6) **HLA-A *3101 SSP genotyping**
7) **Details of population stratification**

Results

1) **Table 1. High resolution sequence based HLA-A typing**
2) **Figure 1. Population stratification detection in participants with hypersensitivity syndrome**
3) **Figure 2. Population stratification detection in participants with maculopapular exanthema**
4) **Figure 3. Manhattan plot for logistic regression on CBZ-induced maculopapular exanthema**
5) **Figure 4. SSP PCR for HLA-A*3101**

Methods

Recruiting centres

University of Liverpool collaborators included the UK national prospective SANAD study, the EUDRAGENE study and Universitätsspital Bern (Switzerland). The EPIGEN Consortium includes Duke University (North Carolina), University College London (UK), Université Libre de Bruxelles (Brussels, Belgium) and the Royal College of Surgeons and Beaumont Hospital in Dublin, Ireland.

Definition of Stevens Johnson Syndrome

SJS was defined as skin detachment 1-10% (SJS), 10-30% (overlap syndrome) and >30% (TEN), the presence of atypical target lesions predominately on the trunk and face, severe, often haemorrhagic, erosions of mucous membranes. Other manifestations indicating systemic involvement included fever, liver chemistry elevations, intestinal and pulmonary manifestations, presence of lymphopenia).

Data access for healthy controls

Individual-level genotype data (Illumina 1.2M (custom chip) were accessed for both sets of healthy controls by application to the Wellcome Trust Case Control Consortium Data Access Committee (https://www.wtccc.org.uk/ccc1/access_to_data_samples.shtml).

Details on cleaning of Illumina data

Illumina 1.2M on 22 HSS cases: The Illumina gender markers were utilized to compare clinical data with the DNA sample gender. Hardy-Weinberg equilibrium was estimated and SNPs which failed ($p< 0.0001$) in control population were excluded from further analyses. In addition, SNPs with MAF <1% and SNPs with genotype success rate < 97% were excluded. Final Manhattan plot cleansing step was based on low intensities and regional LD estimates. Accurate annotation of the variant names and strand orientation by explicit specification of the version of the human genome assembly and of dbSNP ensured compatibility of different datasets for comparison in pooled analysis of case-control studies.

Illumina 610k on 49 EPIGEN **MPE cases**: All quality control measures were performed using PLINK. All SNPs with a genotyping success rate < 90%, those with MAF < 2% and subjects with genotyping success rate < 95% were excluded from the dataset. SNPs that were outside a Hardy-Weinberg Equilibrium threshold of $p<0.0001$ in control subjects were also removed. To correct for potential cryptic relatedness, the dataset was thinned to SNPs in linkage equilibrium (SNPs with $r^2<0.25$) and identity-by-descent of overall genome was calculated with a cut-off threshold of 12% shared with another individual. Population structure was determined by principal components analysis and outliers were removed.

Details of imputation of HLA alleles in EPIGEN Consortium samples:

We imputed HLA types across 5 loci (HLA-A, B, C, DR, DQ) to the level of 4 digits maximum. To achieve this, our full GWAS dataset was reduced to SNPs within the HLA locus (NCBI Build 36, base positions 25760204 to 33529555 on chromosome 6). All MPE cases and CBZ-tolerant controls were imputed together as one population. . Background European HLA allele frequencies were obtained from www.allelefrequencies.net [21] and through the Irish Blood Transfusion Service Bone Marrow Registry [22]. The quality of the imputation by MACH is described by an r^2 value for each HLA type imputed. The quality threshold was set to $r^2>0.9$. Summary information and quality scores per HLA type imputed are available upon request.

Background European HLA allele frequencies were obtained from www.allelefrequencies.net[21] and through the Irish Blood Transfusion Service Bone Marrow Registry[22]

HLA-A *3101 SSP genotyping:

In order to validate the association with the imputed HLA-A*3101 (in EPIGEN samples), we performed an established wet lab technique to confirm the presence of HLA alleles by amplification of genomic DNA with sequence specific primers (SSP) in a polymerase chain reaction. Sequence-specific primers for HLA-A*3101 were obtained with kind permission from Mary N. Carrington of the National Cancer Institute- Frederick (Maryland, USA) for detecting HLA-A*3101 across the entire EPIGEN cohort. The forward primer sequence is 5'-GATAGAGCAGGAGAGGCCT-3' and the reverse primer sequence is 5'-AGCGCAGGTCCTCGTTCAA-3'. Visualisation of confirmatory bands was possible by placing an electric current across a 2% agarose gel containing the samples (see Figure 4 below).

All samples were correctly imputed as each HLA-A*3101 allele carrier was positively identified by the SSP typing method.

Details of population stratification

Principle component analysis was performed to assess genetic markers for ethnicity. All participants were self reported Caucasans. Only individuals with genetically matching Caucasian ethnicity were included into the association analysis. We were working with Phase 1 1958 controls (n=1500), we removed 204 outliers from PCA plot leaving us with 1296 samples for case/control analysis. PCA plots for Liverpool HSS (n=22) in conjunction with healthy controls (n=2291) and EPIGEN MPE and EPIGEN controls are shown in Supp Fig 1 and Supp Fig 2 below respectively.

Results

Table 1. High resolution sequence-based HLA-A typing in patients with carbamazepine (CBZ)-induced hypersensitivity syndrome and CBZ-tolerant patients.

HLA allele‡	CBZ-tolerant controls	Allelic Frequency	Hypersensitive patients	Allelic Frequency	P-value	Odds Ratio (95% CI)
A*0101g§	17	0.19	12	0.27	0.3729	1.57 (0.6703-3.6595)
A*0201g	21	0.24	8	0.18	0.511	0.71 (0.2855-1.7605)
A*0301g	11	0.13	0	0	0.0601	0.08 (0.0044-1.3157)
A*1101g	11	0.13	3	0.07	0.3839	0.51 (0.1352-1.9399)
A*2301g	2	0.02	0	0	1	0.39 (0.0183-8.2745)
A*2402g	3	0.03	5	0.11	0.1162	3.63 (0.8263-15.969)
A*2403g	0	0	1	0.02	0.2576	6.10 (0.2436-152.9367)
A*2501g	5	0.06	1	0.02	0.6631	0.39 (0.0437-3.4093)
A*2601g	6	0.07	2	0.05	0.7182	0.65 (0.1259-3.365)
A*2608	1	0.01	0	0	1	0.66 (0.0262-16.419)
A*3001g	3	0.03	0	0	0.6622	0.27 (0.0139-5.4331)
A*3002	1	0.01	0	0	1	0.66 (0.0262-16.419)
A*3101g	**1**	**0.01**	**8**	**0.18**	**0.00064**	**19.33 (2.3325-160.2439)**
A*3201	3	0.03	1	0.02	1	0.66 (0.0665-6.5245)
A*3303g	0	0	2	0.05	0.1078	10.41 (0.489-221.697)
A*6801	0	0	1	0.02	0.2576	6.10 (0.2436-152.9367)
A*6801g	3	0.03	0	0	0.6645	0.27 (0.0139-5.4331)
Total alleles	88		44			

‡*Antigen recognition site allele*

§*Alleles bearing suffix 'g' in A locus have identical sequences in exon 2 and exon 3 antigen recognition sites.*

An association between HSS and HLA-A*3101 presence was identified (OR 24.6; 95% CI 2.8, 214.0); P=0.00004). No other HLA-A alleles showed statistically different allelic frequencies.

Figure 1. Population stratification detection using GWAS in WTCCC and patients with hypersensitivity syndrome (HSS)

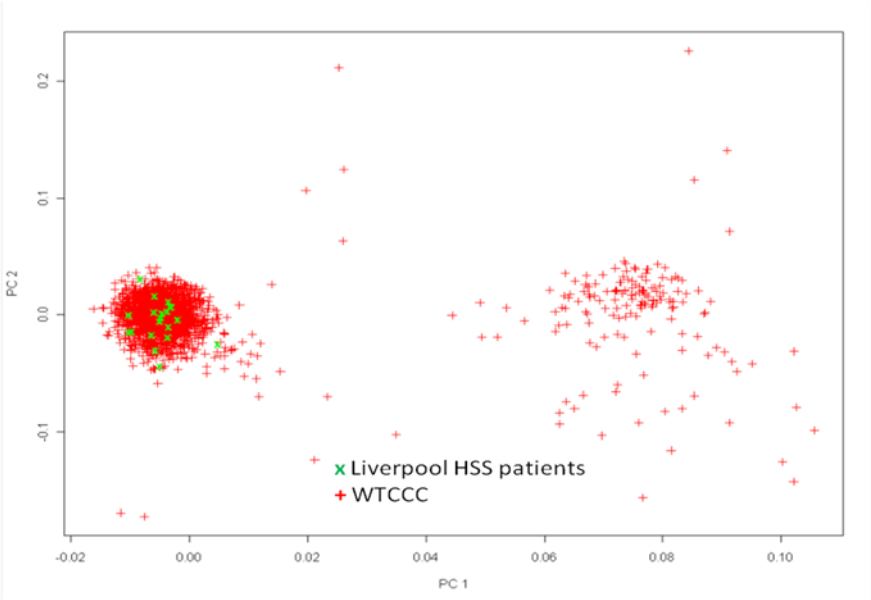


Figure 2. Population stratification detection using GWAS within the cohort of maculopapular exanthema (MPE) cases from the EPIGEN consortium and healthy controls from the 1958 cohort. All participants were self reported Europeans. Ethnicity was reported orally to clinician during investigations and recorded in clinical notes.

Principal Components Analysis was performed to reduce subjects to a homogenous cluster of apparent northern European ethnicity. The final cohort is shown below.

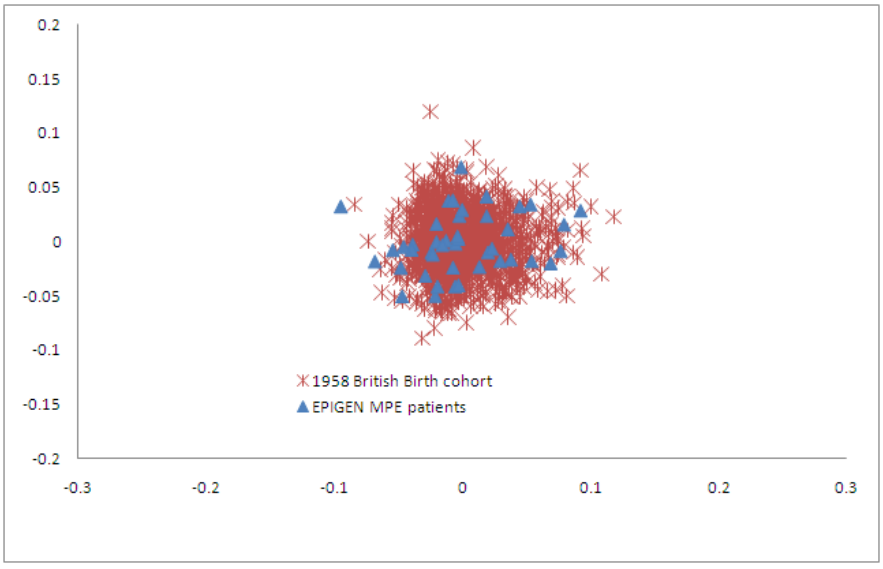
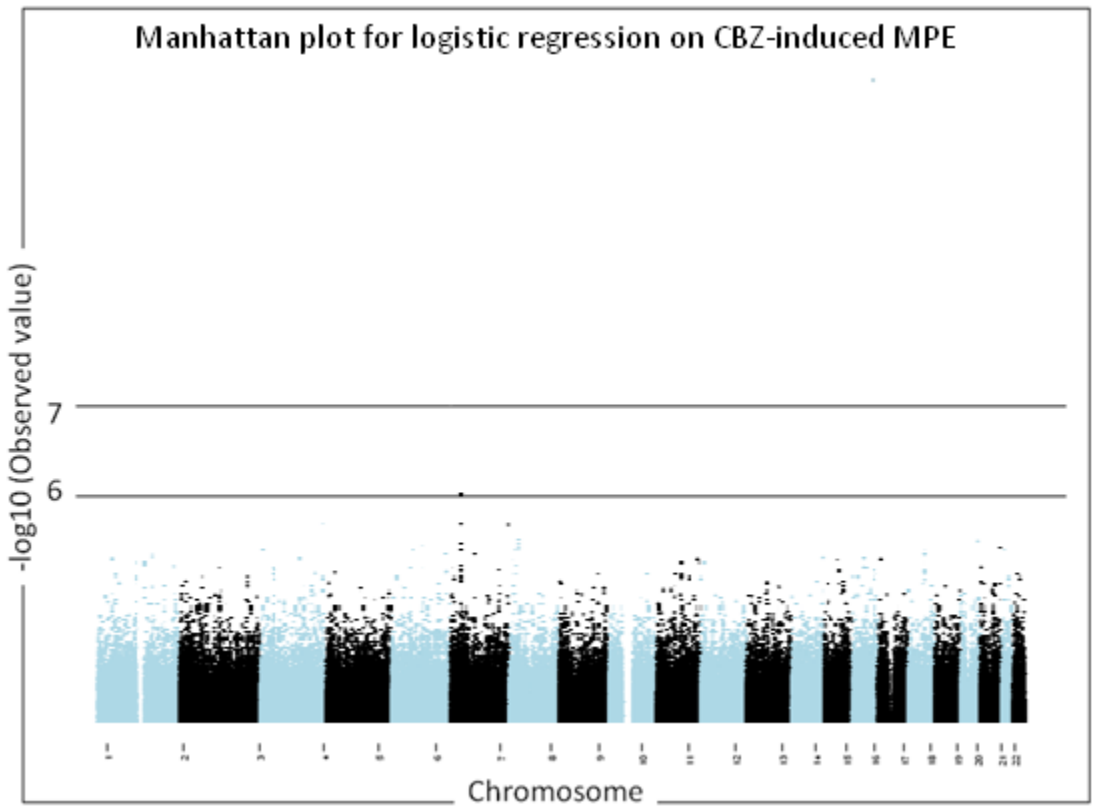
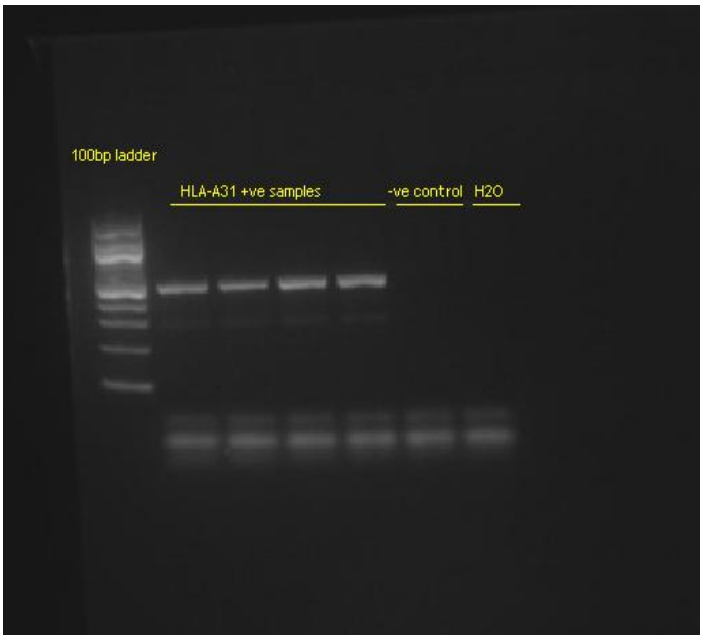


Figure 3. Results of MPE GWAS analysis



GWAS analysis involving MPE cases (n=43) and healthy controls from the 1958 British Birth Cohort (n=1296). Significant EIGENSTRAT covariates are included as covariates in logistic regression model. The top hit on chromosome six is HLA*3101 ($P = 1.11 \times 10^{-6}$).

Figure 4: Selection of imputed HLA-A*3101 samples following PCR amplification with sequence specific primers.



Imputed positive sample lanes 1-4 contained a strong gel band as confirmation whereas the imputed negative sample shown here in lane 5 contained no band.