The Pharmacogenomics Journal (2012) 12, 96–104

www.nature.com/tpj

ORIGINAL ARTICLE

Genome-wide association study of serious blistering skin rash caused by drugs

Y Shen[1,2,13], P Nicoletti[1,3,13], A Floratos[1], M Pirmohamed[4], M Molokhia[5], P Geppetti[3], S Benemei[3], B Giomi[3], D Schena[6], A Vultaggio[7], R Stern[8], MJ Daly[9,10], S John[11], MR Nelson[12] and I Pe'er[1,2], the International Serious Adverse Events Consortium (SAEC)

[1]*Center for Computational Biology and Bioinformatics, Columbia University, New York, NY, USA;* [2]*Department of Computer Science, Columbia University, New York, NY, USA;* [3]*Department of Clinical Pharmacology, University of Florence, Florence, Italy;* [4]*MRC Centre for Drug Safety Sciences, Department of Pharmacology, The University of Liverpool, Liverpool, UK;* [5]*London School of Hygiene and Tropical Medicine, University of London, London, UK;* [6]*Department of Dermatology, University of Verona, Verona, Italy;* [7]*Department of Internal Medicine, Immunoallergology Section, Careggi Teaching Hospital, University of Florence, Florence, Italy;* [8]*Department of Dermatology Beth Israel, Deaconess Medical Center, Harvard Medical School, Boston, MA, USA;* [9]*Program in Medical and Population Genetics, Broad Institute of Harvard University and Massachusetts Institute of Technology, Cambridge, MA, USA;* [10]*Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA;* [11]*Pfizer, New London, CT, USA and* [12]*GlaxoSmithKline, Research Triangle Park, NC, USA*

Correspondence:
Dr I Pe'er, Department of Computer Science, Columbia University, 1214 Amsterdam Avenue, Mail code 0401, New York, NY 10027, USA.
E-mail: itsik@cs.columbia.edu and Dr MR Nelson, GlaxoSmithKline, 5 Moore Drive, MAI.A1227, Research Triangle Park, NC 27709, USA.
E-mail: matthew.r.nelson@gsk.com

[13]These authors contributed equally to this work.

Received 14 January 2010; revised 7 October 2010; accepted 27 October 2010; published online 11 January 2011

Stevens–Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) are rare but severe, potentially life threatening adverse drug reactions characterized by skin blistering. Previous studies have identified drug-specific and population-specific genetic risk factors with large effects. In this study, we report the first genome-wide association study (GWAS) of SJS/TEN induced by a variety of drugs. Our aim was to identify common genetic risk factors with large effects on SJS/TEN risk. We conducted a genome-wide analysis of 96 retrospective cases and 198 controls with a panel of over one million single-nucleotide polymorphisms (SNPs). We further improved power with about 4000 additional controls from publicly available datasets. No genome-wide significant associations with SNPs or copy number variants were observed, although several genomic regions were suggested that may have a role in predisposing to drug-induced SJS/TEN. Our GWAS did not find common, highly penetrant genetic risk factors responsible for SJS/TEN events in the cases selected.
The Pharmacogenomics Journal (2012) **12,** 96–104; doi:10.1038/tpj.2010.84; published online 11 January 2011

Keywords: pharmacogenetics; GWAS; serious skin rash; Stevens–Johnson syndrome; SNP; drug adverse reactions

Introduction

Stevens–Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) are the extreme clinical manifestations of mucocutaneous eruptions. SJS/TEN are the rarest varieties of severe cutaneous adverse reactions, mostly induced by a variety of medications through a dose-independent immune response.[1–2] SJS/TEN are characterized by mortality rates between 1 and 30%, depending on the extent of blistering, and up to 35% of serious sequelae in survivors.[2]

Over the past few decades, a series of studies addressed the epidemiology of SJS/TEN in terms of variability according to ethnicity and causative drugs.[3] Each year, SJS/TEN affects eight Asian,[4] and only one to two ethnically European individuals per million persons. Antibacterial sulfonamides (for example, co-trimoxazole), allopurinol and antiepileptic drugs (phenytoin, carbamazepine and lamotrigine) are consistently the most frequent SJS/TEN-inducing drugs.[5] Drug specific relative risk differs significantly among ethnic groups. For instance, carbamazepine accounts for 25–33% of SJS cases in east Asians, but only 5–6% in Europeans.[6]

Heritable predisposition to SJS/TEN was recognized by several case–control candidate gene studies, but this is likely to be drug and often ethnicity specific. Indeed, in a broader east Asian population[6–8] Human leucocyte antigen (*HLA*)-*B*1502* is a predictive marker for carbamazepine-related severe cutaneous adverse

reaction with a positive predictive value of 7%. SJS/TEN caused by phenytoin also shows a weaker association with *HLA-B*1502*.[9] This allele is rare in Europeans, which could be the reason why similar HLA risk factors have not been identified.[10] In contrast, ethnicity does not seem to influence the incidence of allopurinol severe cutaneous adverse reaction,[11] in which *HLA-B*5801* has been shown to be strongly associated in both European[12] and Asian populations.[13] Finally, a candidate gene study in Japanese patients with SJS/TEN to a number of drugs showed an association with *HLA-A*0260*.[14] Although there have been a number of studies investigating the role that candidate genes may have in drug-induced SJS/TEN, no highly reproducible genetic predictors other than *HLA-B*1502* and *HLA-B*5801* have emerged.

Here, we report the first genome-wide association study (GWAS) of SJS/TEN. In this retrospective study, we used standard phenotypic criteria to identify SJS/TEN cases that were due to a number of drugs. Our aim was to discover highly penetrant and common genetic factors that increase the risk of developing SJS/TEN, which may not necessarily be drug specific.

Materials and methods

Subjects

PGX40001 collection. Cases were defined by Pirmohamed *et al*.[15] Briefly, cases were retrospectively enrolled and screened by expert review based on SJS/TEN inclusion and exclusion criteria. The controls were collected at the site of case matching for age, gender and ethnicity. Six of the cases were defined as SJS–TEN spectrum, that is, they were within the spectrum of the blistering skin reactions (SJS, overlap syndrome, TEN), but incomplete clinical data meant that we were unable to categorize accurately. In total, 71 cases and 135 controls were genotyped using Illumina 1M in our study (Illumina, http://www.illumina.com).

LAM30004 collection. Cases were defined by Kazeem *et al*.[16] All cases and controls were epilepsy patients treated with lamotrigine and recruited retrospectively. Cases were patients who developed SJS/TEN or hypersensitivity reactions (although only SJS/TEN cases are included in this study), whereas controls were patients exposed to lamotrigine without developing SJS/TEN. In addition to matching for the drug of interest, the controls were also chosen on the basis of other factors, such as age, ethnicity and concurrent valproic acid usage. SJS/TEN was defined by dermatologists using standard phenotypic criteria. In total, 6 cases and 63 controls were included in this study.

Italian cases. Between November 2007 and March 2008, 19 retrospective Caucasian cases of SJS because of the number of drugs were collected from Dermatology Department at the University of Florence and Dermatology Department at University of Verona. Cases were defined on the basis of three major clinical criteria: pattern of skin lesions, distribution of lesions and percentage of epidermal detachment during the course of the disease. A diagnosis of SJS was considered if blistering did not exceed 10%, whereas TEN cases had $>30\%$ of the body surface area blistered, with SJS–TEN overlap cases occupying an intermediate position.[3] Etiology assessment was performed on all cases, and only cases with drug as the most likely cause were included in the study. Exclusion criteria were (a) concomitant human immunodeficiency virus infection, and (b) concomitant immunosuppressant drugs. Ethical approval was provided by the University of Florence Ethics committee.

Control selection

The controls from PGX40001 were collected at the same site as the cases, matched on the basis of ethnicity and gender. The controls in LAM30004 were matched on drug treatment as well as age and ethnicity. No specific control matching was carried out for the Italian cases. We obtained the genotypes of 88 HapMap3 TSI[17] subjects as the controls for the Italian cases. We also included 659 POPRES (population reference sample) controls genotyped by Serious Adverse Events Consortium as the basis for separating sub-populations within Europeans using principal components analysis (PCA; see below). In addition, about 4837 WTCCC2 subjects were analyzed and 4108 of these were chosen to match the northwestern Europeans (nw-EU) cases.

Genotyping

All subjects from PGX40001 and LAM30004 and POPRES controls and Italian collection were genotyped at Expression Analysis (Research Triangle Park, NC, USA). PGX40001 and LAM30004 and were genotype together using Illumina 1M chip, whereas the Italian cases by Illumina 1M-duo chip. The POPRES subjects were genotyped by Illumina 1M or 1M-Duo for another Serious Adverse Events Consortium study[18] The chips contain about 1.07 and 1.2 million markers of single-nucleotide polymorphisms (SNPs) and copy number variant (CNV) probes, respectively. The genotypes of HapMap TSI were downloaded from http://hapmap.org public release no. 27. The genotypes of WTCCC2 subjects were downloaded from The European Genome–phenome Archive (http://www.ebi.ac.uk/ega/)—1958 British Birth Cohort and United Kingdom Blood Service Group (only Illumina 1M data).

Genotype quality control

For each set of genotype data, we applied a series of quality control steps. Specifically, any marker that did not pass any of the following criteria was discarded:

1. Call rate $>95\%$;
2. Minor allele frequency $>1\%$;
3. A *P*-value for Hardy–Weinberg equilibrium $>10^{-7}$ in controls (if applicable).

After applying these criteria, 837 175 SNPs were left in the PGX40001 and LAM30004 subjects. The Italian collection was merged with HapMap TSI subjects. There were 781 191 SNPs left after removing those SNPs that had a significant allele frequency difference between HapMap TSI and

POPRES Italians. All subjects had <10% of missing genotyping calls.

In addition, we identified subjects that were highly related on the basis of estimated identity-by-descent using PLINK v1.05[19,20] There were 21 samples with identity-by-descent sharing values >0.2 and <0.9 with at least one other sample. Removing 12 samples resolved the issue (each of removed samples had lower overall SNP call rate than its related sample). There were four pairs of samples with almost identical genotypes (identity-by-descent sharing >0.99). These near-identical samples were further investigated by comparing the genotypes from this study with those from a previous experiment that included Affymetrix 500K (Affymetrix, http://www.affymetrix.com) SNP genotyping. Three samples (two were cases) were found to have significant inconsistency and were excluded. The fourth pair was also identical according to the Affymetrix 500K result and regarded as a true sample duplicate. The sample with the higher SNP call rate was retained. In addition, two cases were removed because of the inconsistency between reported ethnicity and that inferred by PCA (see next section). In total, 18 subjects were discarded, including 5 cases and 13 controls.

When combining data from external sources (such as WTCCC2 and HapMap), we removed SNPs that had significantly different allele frequencies compared with the control set that was genotyped in the same facility as the cases (Fisher's exact test *P*-value $<10^{-4}$). We found that all SNPs with potential strand annotation issues were also discarded by this approach.

Principal components analysis

We used the smartPCA program from the EIGENSTRAT package (version 3.0)[21] to conduct PCA in order to expose population structure. SNPs from known regions[22] of long-range linkage disequilibrium were removed before conducting PCA. We first combined the study genotype data with HapMap data to identify major ancestry groups (Europeans, Asians and Africans). We then conducted PCA on Europeans (self-reported non-Hispanic white or European) to separate sub-populations among Europeans.

Statistical analysis

We conducted statistical tests through PLINK v1.05. For the overall European cases and controls, we tested association using logistic regression with gender and the first four eigen scores as covariates under an additive model. For the nw-EU group, we tested associations using Fisher's exact test under an additive model.

Power simulation

The simulation conditions are listed in Supplementary Table S1. The procedure was implemented in R (version 2.6–2.9).[23]

CNV analysis

The CNV calls were generated using PennCNV[24] software (April 2009 version). To ensure the accuracy of CNV calling, we applied stringent sample and CNV filtering procedures. We studied the relationship among the mean and standard deviation of log *R* ratio (normalized signal intensity from BeadStudio by Illumina) and the number of CNV calls (Supplementary Figure S1). We included all sample that had a log *R* ratio standard deviation <0.23, maximum number of total CNV calls <200, maximum number of 100 kb CNV call <20, bioaccumulation factor median >0.55 or <0.45, bioaccumulation factor drift >0.002 or waviness factor >0.04 or <−0.04. Additionally, to ensure high-confidence CNVs, we excluded individual CNVs if:

1. PennCNV-generated confidence score was <10;
2. Called based on fewer than 10 SNPs/CNV probes; and
3. Spanned within 1 Mb from centromeres or telomeres.

We performed burden and common CNVs association analysis. Any CNV that was present in at least three subjects was considered to be common.[25] Associations were tested using two tails permuted (100 000 times) Fisher's exact analysis using PLINK software, by considering duplications and deletions separately.[16] We also investigated singleton CNVs >100 kb[26] to find evidence for individual predisposition to SJS/TEN. For this analysis, we adopted a coverage cutoff excluding all CNVs that had coverage >20 genetic markers/CNVs. All analyses were carried out in the nw-EU subjects that passed the genotyped quality control procedure, excluding Italian cases, as they did not have an ethnically matched control group.

Results

Initial GWAS

This study was based on three SJS/TEN collections: PGX40001,[15] LAM30004[16] and an Italian collection (Materials and methods). Both PGX40001 and Italian cases were induced by multiple drugs (Tables 1a and b), whereas all LAM30004 case were because of lamotrigine. In total, 294 subjects were genotyped, of which 276 (including 91 cases and 185 controls) passed quality control steps (Materials and methods). Genetic structure was investigated by PCA

Table 1a Causal drug summary of SJS/TEN collections

Drugs	*Sub-population*			*Total*
	nw-EU	*s-EU*	*e-EU*	
Co-trimoxazole	12	—	—	12
Lamotrigine	9	—	—	9
Amoxicillin	3	2	—	5
Phenytoin	7	—	1	8
Moxifloxacin	2	—	1	3
Carbamazepine	3	—	—	3
Allopurinol	1	2	—	3
Clarithromycin	1	2	—	3
Others	21	16	—	34

Abbreviations: e-EU, eastern Europeans; nw-EU, northwestern Europeans; s-EU, Italians; SJS, Stevens–Johnson syndrome; TEN, toxic epidermal necrolysis.
Six subjects experienced SJS/TEN due to more than one causal drug.

Table 1b Demographic and clinical characteristics of the enrolled patients summarized by cohort

Collection	*Number of subjects*	*% Female*	*European cohort*			*Diagnosis*		
			Cases	*Controls*	*Case/control ratio*	*SJS*	*Overlap*	*TEN*
PGX40001	206	67	48	107	2.2	23	22	3
LAM30004	69	51	5	51	10.20	5	—	—
Italian collection	19	42	19	—	Case only	14	3	2
TSI	88	50	—	88	Controls only	—	—	—
WTCCC2	4837	49	—	4108	Controls only	—	—	—
POPRES	659	53	—	648	Controls only	—	—	—

Abbreviations: POPRES, population reference sample; SJS, Stevens–Johnson syndrome; TEN, toxic epidermal necrolysis.

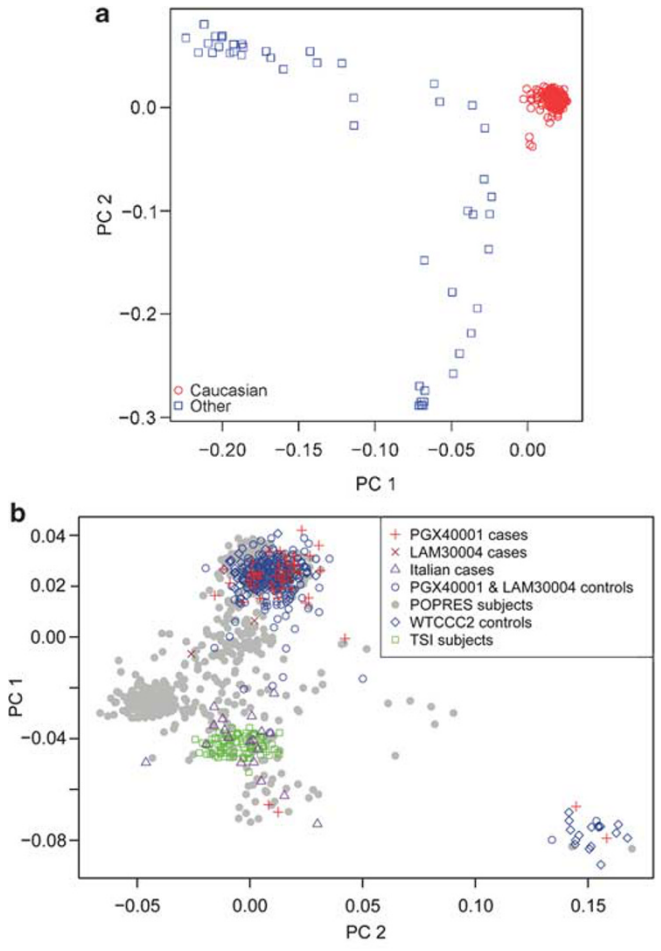

Figure 1 Population structure of SJS/TEN cohorts. (**a**) Population structure of all subjects from three collections. The red circles represent Caucasian subjects, blue squares represent subjects of other ethnicities. (**b**) Population structure of Caucasian subjects. The first two eigen vectors separate the Europeans into nw-EU cluster (top), Italian cluster (lower center) and eastern Europeans (lower right). The gray cluster on the lower left are POPRES of Spanish origin. PC, principal component.

(Figure 1a). We selected 72 European cases and 162 controls for further analysis. To improve the power of the study and the ability to control for population stratification within Europeans, we combined the genotypes of European subjects with those of 88 HapMap3 TSI and 659 POPRES[27] subjects, which represent representative sub-populations among Europeans. The first two eigen vectors from PCA separated the cohort into United Kingdom (nw-EU), Italian (s-EU) and east Europeans clusters (Figure 1b). For each case, we selected up to seven closest controls based on eigen scores of the first four vectors, resulting in 461 controls that match 72 cases. We tested the association of single SNPs using logistic regression with first four eigen scores and gender as covariates under an additive model (Figure 2). We found no SNPs with a *P*-value $<10^{-6}$. The top associated SNP was rs6016348 on chromosome 20 (*P*-value $=1.3\times10^{-6}$; odds ratio (OR) $=2.9$, 95% confidence interval: 1.9–4.6; Table 2).

We then tested the association of the SNPs in s-EU group and the nw-EU group separately.

s-EU group. There were 21 cases and 107 controls. Among the controls, 88 were from HapMap TSI, and 18 were from POPRES of Italian origin. To remove the SNPs that may have genotyping errors, we discarded all SNPs with significant different allelic frequencies between the TSI and POPRES controls of Italian origin ($P<10^{-4}$). We tested associations with 781 191 SNPs that passed quality control and found no SNP with a *P*-value $<10^{-6}$.

nw-EU group. We first analyzed the nw-EU subjects from PGX40001 and LAM30004 collections only. There were 46 cases and 143 controls, including 59 males and 130 females. We tested associations with 837 070 SNPs that passed quality control. We found no SNPs with *P*-values $<10^{-6}$.

Improving power by expanding the control set

Because of the rarity of SJS/TEN, it is extremely difficult to collect more cases to improve the power of the study in a timely manner. Therefore, we expanded the control set[15] with WTCCC2 publicly available data set,[28] genotyped by Illumina 1M chip. We combined the cleaned WTCCC2 set with the nw-EU group. In the combined set, there were 46 cases and 4251 controls. We applied Fisher's exact test on the data set (Supplementary Figure S2). The top associated SNP is rs17137412 (chromosome 7, position 7767212, *P*-value $=1.2\times10^{-8}$, OR: 4.0, 95% confidence interval of OR: 2.5–6.2; minor allele frequency (MAF) in cases versus

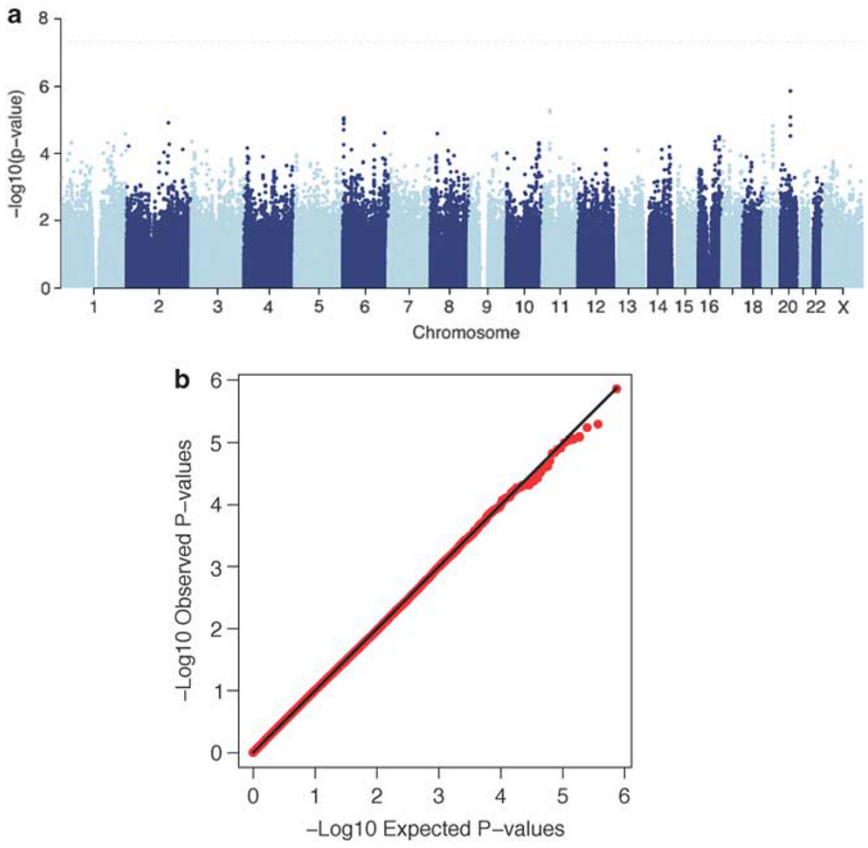

Figure 2 GWAS result from overall European cases and controls. Results of the analysis of 72 European cases exposed to multiple drugs and 461 selected controls. (**a**) The 'Manhattan' plot showing each SNP, the –log10 of the *P*-value from logistic regression test was plotted against chromosome position. The dashed horizontal line denotes genome-wide significance threshold. (**b**) The quantile–quantile plot of –log10 of *P*-values against the expected values under the null model. The bulk of the values (red) closely follows the expectation under the null model (black line).

controls: 0.33 versus 0.11). This is an intronic SNP within the *AC006465.3* hypothetical gene, and close to *RPA3*. There were no other SNPs with a *P*-value $<10^{-6}$.

Drug-specific groups

Previous studies with carbamazepine and allopurinol indicate that, it is possible to detect drug-specific risk alleles with large effects based on small numbers of cases.[6,13,29] To investigate this, we looked at two drug-specific groups: co-trimoxazole (12 European cases) and lamotrigine (9 European cases).

Co-trimoxazole SJS/TEN cases. No SNPs with a *P*-value $<10^{-6}$ were identified when comparing the 12 cases to 143 nw-EU controls or 2251 WTCCC2 controls.

Lamotrigine SJS/TEN cases. No SNPs with a *P*-value $<10^{-6}$ were identified when comparing the 9 cases with 143 nw-EU controls (including 52 drug-matched controls). Interestingly, the top associated SNP in an additive model (rs12019361, $P=7\times10^{-6}$, OR = 12, MAF in all controls: 0.06, MAF in 52 lamotrigine-exposed controls: 0.067) is intronic in the *ADAM22*, which is a gene involved in epilepsy,[30] a primary indication for this drug.

Association with CNVs

Among the 192 nw-EU individuals who passed the SNP quality control checks, 134 individuals (37 cases and 97 controls) passed stringent quality-control criteria for CNV calling (Materials and methods). A total of 1062 duplications and 2171 deletions were predicted. The numbers and average size of both deletions and duplications were not significantly different between cases and controls. After multi-test correction, none of the common CNVs had a significant association. There was no significant difference in the number of singleton CNVs that are >100 kb, although the average size of deletions were larger in cases than in controls (458 versus 203 kb, permuted *P*-value = 0.03). We found 11 unique oversize (>500 kb) CNVs with three of them being >1 Mb (Table 3).

Power calculations

We conducted *post hoc* power calculations to better understand the power and the potential benefit from expanding control sets in the context of our study.

We assumed a certain OR value under an additive genetic model, and a range of MAFs in the population (MAF_{Popu}). The MAF in cases (MAF_C) was determined by the OR and MAF_{Popu}. The number of cases and controls were chosen to

Table 2 GWAS result: top associated SNPs

Marker Chromosome	SNP	Type	Closest gene	Overall cases (72 versus 461) Logistic regression		nw-EU (46 versus 4251) Fisher's exact test				s-EU cohort (21 versus 97) Logistic regression			
				P value	OR (95% CI)	MAF		P value	OR (95% CI)	MAF		P value	OR (95% CI)
						Case	Control			Case	Control		
6	rs981946	Intronic	*SLC22A23*	8.9×10^{-6}	2.3 (1.6–3.4)	0.52	0.36	0.0021	1.9 (1.2–2.9)	0.43	0.31	0.0063	2.9 (1.3–6.2)
6	rs1079284	Intronic	*SLC22A23*	9.3×10^{-6}	2.3 (1.6–3.4)	0.52	0.36	0.0021	1.9 (1.2–2.9)	0.43	0.31	0.0063	2.9 (1.3–6.2)
7	rs17137412	Intronic	*RPA3*	0.0001	2.4 (1.5–3.7)	0.34	0.11	1.3×10^{-8}[a]	4.0 (2.5–6.2)	0.07	0.08	0.87	0.9 (0.2–3.1)
7	rs12019361	Intronic	*ADAM22*	0.0042	2.2 (1.2–3.8)	0.13	0.07	0.019	2.1 (1.1–3.9)	0.12	0.08	0.36	1.6 (0.5–5.2)
20	rs6016348	Intergenic	*MAFB*	1.4×10^{-6}	2.9 (1.9–4.5)	0.29	0.13	4.2×10^{-5}	2.7 (1.7–4.3)	0.26	0.13	0.022	3.1 (1.1–8.3)
20	rs6016358	Intergenic	*MAFB*	8.2×10^{-6}	2.9 (1.8–4.7)	0.24	0.09	1.3×10^{-5}	3.2 (2.0–5.3)	0.19	0.10	0.079	2.4 (0.9–6.5)

Abbreviations: CI, confidence interval; GWAS, genome-wide association study; MAF, minor allele frequency; nw-EU, northwestern Europeans; OR, odds ratio; s-EU, Italians; SNP, single-nucleotide polymorphism.
[a]Exclusion of the six probable Stevens–Johnson syndrome cases did not change the association with rs17137412 in nw-EU group ($P=2.4\times10^{-8}$; OR = 4.3, 95% CI: 2.7–6.8).

reflect the scenarios in our study (Supplementary Table S1). All controls were assumed to be population controls. The prevalence of SJS/TEN was set at 0.001, which is likely to be larger than the true value in the European population. The number of latent cases in the population controls was calculated as the product of the prevalence and the total number of controls. The numbers of minor alleles in cases and controls were determined by sampling independently from binomial distributions, with *P*-value at MAF_C and MAF_{Popu}, respectively. For each combination of the condition, we repeated the sampling procedure 10 000 times, and each time calculated the *P*-values using Fisher's exact test assuming additive models. We then estimated the power as the fraction when the *P*-value is $<5\times10^{-8}$ (genome-wide significance).

As expected, the power always increases with the increase in the number of controls. With 46 cases, assuming an OR of 4.0 and MAF_{Popu} of 0.1, which are the conditions that are similar to the top associated SNP in nw-EU SJS/TEN group, the increase in power of detecting genome-wide significant (*P*-value $<5\times10^{-8}$) markers reaches a plateau of 0.5 at around 2000 controls. The power is 0.18 from the original sample size of 143 controls. Therefore, using publicly available external population controls increases the power by about threefold (Figure 3). In all, 80 cases are required to achieve a comparable increase in power given the original 143 controls.

Discussion

SJS and TEN are life threatening adverse drug reactions, in which genetic predisposition has been shown to be important with drugs, such as carbamazepine and allopurinol. In this study, we have evaluated patients who we have genotyped previously using a candidate gene approach,[16] and combined them with new patient cohorts exposed to lamotrigine and an Italian group exposed to various drugs. We have then undertaken a GWAS, the first GWAS on SJS/TEN patients, with the aim to identify common variants that exert large effects on the risk of developing drug-induced SJS/TEN. The strengths of the study were as follows: (a) it is the largest set of SJS–TEN patients collected to identify genetic predisposition; (b) all patients were clinically phenotyped consistently using the criteria laid down by Bastuji-Garin *et al*;[3] (c) only patients, in whom there was plausible drug-related etiology were recruited; and (d) it was undertaken using the most up-to-date genotyping technology. Many different drugs are known to cause SJS/TEN, and this is reflected in the drug groups represented in our study. On the one hand it provides power to identify a common predisposing locus (or loci), but on the other hand it is limited by the fact that the number of patients in each drug group is low. A limitation of this study is that although all cases were deemed as prominently drug caused, they were retrospectively recruited when the etiology assessment process was not standardized comparing with current

Table 3 CNV analysis result

Collection	*CNV type*	*Genotype*	*Chromosome*	*Start position*	*End position*	*Length (bp)*	*Involved genes*
PGX40001	DUP	Het	1	103776619	106583135	2806517	*RNPC3, AMY2B, AMY2A, AMY1A, AMY1C, AMY1B*
LAM30004	DUP	Het	2	124743745	125785851	1042107	*CNTNAP5*
PGX40002	DEL	Het	7	124892028	125395114	503087	*GRM8*
PGX40001	DEL	Het	13	111553204	114114639	2561436	*SOX1, AK055145, C13orf28, TUBGCP3, C13orf35, ATP11A, MCF2 L, F7, F10, PROZ, PCID2, CUL4A, LAMP1, GRTP1, ADPRHL1, DCUN1D2, TMCO3, TFDP1, ATP4B, GRK1, BC034570, GAS6, DQ866763, FAM70B, RASA3, CDC16*
PGX40001	DUP	Het	20	45902724	46559948	657225	Between *SULF2* and *PREX1*

Abbreviations: CNV, copy number variant; DEL, deletion; DUP, duplication; Het, heterozygous.

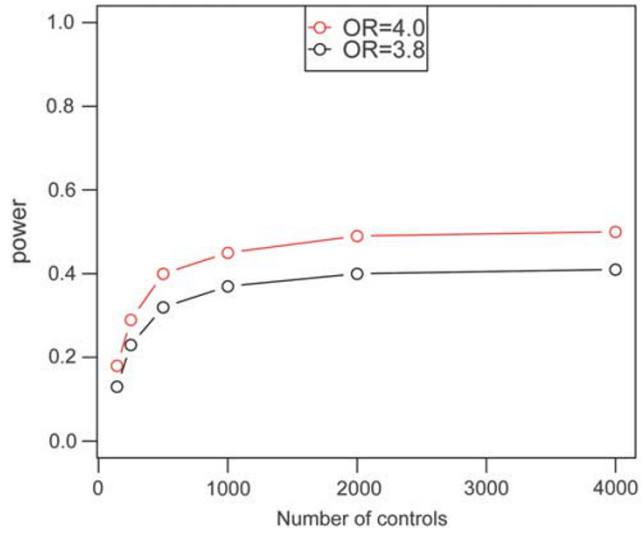

Figure 3 Improving power by expanding control set. The power was defined as the proportion of simulations, wherein *P*-values were smaller than the cutoff value (5×10^{-8}) with 46 cases and the number of controls in the *x*-axis, assuming the odds ratio of the associated SNP was 4.0 and the minor allele frequency was 0.1 (conditions similar to the top associated SNP from the nw-EU group).

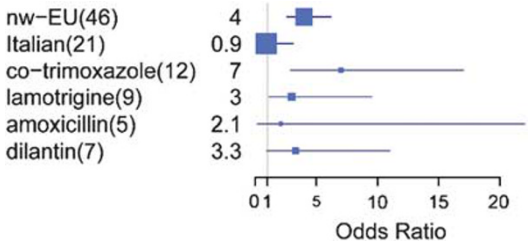

Figure 4 The effect of rs17137412 in different groups. The numbers in parentheses are the numbers of cases for each group. The horizontal blue lines mark the 95% confidence interval of odds ratios.

prospective studies, leaving the possibility of including some non-drug-induced cases.

We found no SNP with genome-wide significance using the standard definition (*P*-value $<5 \times 10^{-8}$) from the overall set of European subjects. From the nw-EU group, we did find a SNP in chromosome 7 with a *P*-value of about 1×10^{-8} and an odds ratio of 4, but it was not replicated in the s-EU group, in which it had an odds ratio <1, even though the MAFs of controls are similar in the nw-EU and s-EU groups. The discrepancy of the effect of this SNP in these two groups may reflect the fact that different drugs are involved (Table 1a), and the effect of the SNP could be drug-specific (Figure 4). Therefore, it seems unlikely that there are common and highly penetrant genetic risk factors that are responsible for most of the SJS/TEN events across drugs. Given that many drugs responsible for the SJS/TEN cases, these results do not exclude the possibility that there may be highly penetrant risk alleles with specific drugs in particular ethnic populations.

Our study highlighted a few biologically relevant SNPs (Table 2), which may contribute to the risk of the event and deserve further investigations. Two SNPs from one of the top associated regions (rs981946 and rs1079284 in chromosome 6) were intronic to gene *SLC22A23*. This gene encodes a human transporter with structural similarity to the drug transporter *SLC22* family.[31] Experimental data suggested that *SLC22s* are involved in drug transport in kidneys, liver and brain. The interaction of *SLC22s* with several classes of drugs is well documented, but which of transporters is the main factor for each drug has not been accessed.[32] We investigated the association of *SLC22A23* SNPs and β-lactam caused SJS/TEN, as β-lactam antibiotics are substrates for transporters. The result is negative, comparing 16 β-lactam cases and 96 genetically matched controls, the *P*-value of the SNPs is about 0.08 and OR is about 2.

We found a few singleton CNVs that cover genes with suggestive functions relevant to SJS/TEN. One of them is a 600 kb heterozygous duplication on 12q13, flanked by *PREX1* and *SULF2* genes. *PREX1* encodes for a protein that regulates GTP-mediated responses to inflammatory signals in mouse macrophage.[33,34] Another interesting CNV is a 2 Mb heterozygous deletion of the telomeric portion of the long arm of chromosome 13 (13q34). The region contains several genes, including *LAMP1* (or *CD107a*). *LAMP1* encodes for a lysosome-associated transmembrane

glycoprotein, which is a marker of both CD8+ lymphocytes[35] and skin mast cell activation.[36] The glycoprotein also seems to be implicated in antigen presentation through major histocompatibility complex.[37,38]

We demonstrated that the power of an association study is greatly improved by expanding the set of controls using publicly available genotypes from individuals of the same ethnicity. On the basis of *post hoc* power calculation, we found that under similar conditions to the top associated SNP from the nw-EU group, the power of detecting genome-wide significant SNPs was increased threefold by expanding the control set to 40 times of the number of cases. This is especially important for studies in which the adverse drug reactions is relatively rare, wherein the collection of a large number of cases is difficult, both in terms of time and cost. Furthermore, wherein the adverse drug reaction is rare, the use of drug-exposed controls does not produce an appreciable difference in power over that obtained through the use of population controls.[29]

In conclusion, this GWAS highlighted a few interesting genomic regions that may increase the risk of developing SJS/TEN from exposure to a variety of drugs. The absence of SNPs with genome-wide significant association indicates that common risk factors with large effect for most SJS/TEN reactions from different drugs are not likely. To study SJS/TEN caused by individual drugs, more cases than were available in our study will need to be identified and recruited to achieve reasonable power. This underscores the importance of systematic efforts of collecting drug-specific SJS/TEN cases. It is also consistent with the concept of deep phenotyping, in which patients with apparently similar clinical manifestations may in fact have different pathogenetic biological processes. Stratification on the basis of more precisely defined phenotypes could theoretically increase statistical power. To this end, we will be launching a consortium termed ITCH (International Consortium on Drug Hypersensitivity), whose remit will be to recruit cases of serious cutaneous adverse drug reactions globally. Another future direction is to study rare variants through sequencing. Recent progress on next-generation sequencing technologies[39–41] has made it feasible to conduct large scale whole-genome or whole-exome resequencing, which has already been demonstrated to be a powerful tool to discover causal genes for Mendelian traits.[42–45] Thus, applying whole-genome or whole-exome sequencing to the study of rare drug adverse reactions and common diseases is the logical next step.

Conflict of interest

The author declares no conflict of interest.

Acknowledgments

We thank Arthur L Holden (Serious Adverse Events Consortium) Brian Spear (Abbott), Joe Walker (Daiichi-Sankyo), Lon Cardon (GlaxoSmithKline), Vincent Mooser (GlaxoSmithKline), Nadine Cohen (Johnson & Johnson), Joanne Meyer (Novartis) Klaus Lindpaintner (Roche), Robert Dix (Sanofi-Aventis), Leonardo Sahelijo (Takeda), Michael Dunn (Wellcome Trust), and Michael E Burczynski (Wyeth) for their support. We also thank Bernd Jagla for his early contribution to the project. We also thank Julian Arbuckle and Clive Bowman for their contribution in recruiting the SJS–TEN cohort from study PGX40001, and all the clinicians who recruited cases and controls to the study, and all the patients who contributed DNA for the study. This work was funded by the International Serious Adverse Events Consortium.

References

1 Roujeau JC, Guillaume JC, Fabre JP, Penso D, Fléchet ML, Girre JP. Toxic epidermal necrolysis (Lyell syndrome). Incidence and drug etiology in France, 1981–1985. *Arch Dermatol* 1990; **126**: 37–42.
2 Roujeau JC, Stern RS. Severe adverse cutaneous reactions to drugs. *N Engl J Med* 1994; **331**: 1272–1285.
3 Bastuji-Garin S, Rzany B, Stern RS, Shear NH, Naldi L, Roujeau J-C. Clinical classification of cases of toxic epidermal necrolysis, Stevens-Johnson syndrome, and erythema multiforme. *Arch Dermatol* 1993; **129**: 92–96.
4 Lin MS, Dai YS, Pwu RF, Chen YH, Chang NC. Risk estimates for drugs suspected of being associated with Stevens-Johnson syndrome and toxic epidermal necrolysis: a case-control study. *Intern Med J* 2005; **35**: 188–190.
5 Mockenhaupt M, Viboud C, Dunant A, Naldi L, Halevy S, Bouwes Bavinck JN *et al.* Stevens-Johnson syndrome and toxic epidermal necrolysis: assessment of medication risks with emphasis on recently marketed drugs. The EuroSCAR-study. *J Invest Dermatol* 2008; **128**: 35–44.
6 Chung W-H, Hung S-I, Hong H-S, Hsih M-S, Yang L-C, Ho H-C *et al.* Medical genetics: a marker for Stevens-Johnson syndrome. *Nature* 2004; **428**: 486.
7 Lonjou C, Thomas L, Borot N, Ledger N, de Toma C, LeLouet H *et al.* A marker for Stevens-Johnson syndrome: ethnicity matters. *Pharmacogenomics J* 2006; **6**: 265–268.
8 Hung SI, Chung WH, Jee SH, Chen WC, Chang YT, Lee WR *et al.* Genetic susceptibility to carbamazepine-induced cutaneous adverse drug reactions. *Pharmacogenet Genomics* 2006; **16**: 297–306.
9 Man CB, Kwan P, Baum L, Yu E, Lau KM, Cheng AS *et al.* Association between HLA-B*1502 allele and antiepileptic drug-induced cutaneous reactions in Han Chinese. *Epilepsia* 2007; **48**: 1015–1018.
10 Roujeau JC, Huynh TN, Bracq C, Guillaume JC, Revuz J, Touraine R. Genetic susceptibility to toxic epidermal necrolysis. *Arch Dermatol* 1987; **123**: 1171–1173.
11 Halevy S, Ghislain PD, Mockenhaupt M, Fagot JP, Bouwes Bavinck JN, Sidoroff A *et al.* Allopurinol is the most common cause of Stevens-Johnson syndrome and toxic epidermal necrolysis in Europe and Israel. *J Am Acad Dermatol* 2008; **58**: 25–32.
12 Lonjou C, Borot N, Sekula P, Ledger N, Thomas L, Halevy S *et al.* A European study of HLA-B in Stevens-Johnson syndrome and toxic epidermal necrolysis related to five high-risk drugs. *Pharmacogenet Genomics* 2008; **18**: 99–107.
13 Hung SI, Chung WH, Liou LB, Chu CC, Lin M, Huang HP *et al.* HLA-B*5801 allele as a genetic marker for severe cutaneous adverse reactions caused by allopurinol. *Proc Natl Acad Sci USA* 2005; **102**: 4134–4139. Epub 2005 Mar 4132.
14 Ueta M, Sotozono C, Tokunaga K, Yabe T, Kinoshita S. Strong association between HLA-A*0206 and Stevens-Johnson syndrome in the Japanese. *Am J Ophthalmol* 2007; **143**: 367–368.
15 Pirmohamed M, Arbuckle JB, Bowman CE, Brunner M, Burns DK, Delrieu O *et al.* Investigation into the multidimensional genetic basis of drug-induced Stevens-Johnson syndrome and toxic epidermal necrolysis. *Pharmacogenomics* 2007; **8**: 1661–1691.
16 Kazeem GR, Cox C, Aponte J, Messenheimer J, Brazell C, Nelsen AC *et al.* High-resolution HLA genotyping and severe cutaneous adverse reactions in lamotrigine-treated patients. *Pharmacogenet Genomics* 2009; **19**: 661–665.

17 The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; **467**: 52–58.
18 Daly AK, Donaldson PT, Bhatnagar P, Shen Y, Pe'er I, Floratos A *et al.* HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nat Genet* 2009; **41**: 816–819.
19 Purcell S. PLINK v 1.05.
20 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
21 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–909.
22 Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A *et al.* Genes mirror geography within Europe. *Nature* 2008; **456**: 98–101.
23 R Development Core Team. *R. A Language and Environment for Statistical Computing*, vol. 1. Vienna Austria R Foundation for Statistical Computing, 2009. Available from http://www.r-project.org.
24 Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 2007; **17**: 1665–1674.
25 Need AC, Ge D, Weale ME, Maia J, Feng S, Heinzen EL *et al.* A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet* 2009; **5**: e1000373.
26 Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008; **320**: 539–543.
27 Nelson MR, Bryc K, King KS, Indap A, Boyko AR, Novembre J *et al.* The population reference sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am J Hum Genet* 2008; **83**: 347–358.
28 Wellcome trust case control consortium—phase 2, downloaded from http://www.ebi.ac.uk/ega/page.php.
29 Nelson MR, Bacanu SA, Mosteller M, Li L, Bowman CE, Roses AD *et al.* Genome-wide approaches to identify pharmacogenetic contributions to adverse drug reactions. *Pharmacogenomics J* 2009; **9**: 23–33.
30 Fukata Y, Adesnik H, Iwanaga T, Bredt DS, Nicoll RA, Fukata M. Epilepsy-related ligand/receptor complex LGI1 and ADAM22 regulate synaptic transmission. *Science* 2006; **313**: 1792–1795.
31 Jacobsson JA, Haitina T, Lindblom J, Fredriksson R. Identification of six putative human transporters with structural similarity to the drug transporter SLC22 family. *Genomics* 2007; **90**: 595–609.
32 Rizwan A, Burckhardt G. Organic anion transporters of the SLC22 family: biopharmaceutical, physiological, and pathological roles. *Pharm Res* 2007; **24**: 450–470.
33 Dinauer MC. Regulation of neutrophil function by Rac GTPases. *Curr Opin Hematol* 2003; **10**: 8–15.
34 Zhao T, Nalbant P, Hoshino M, Dong X, Wu D, Bokoch GM. Signaling requirements for translocation of P-Rex1, a key Rac2 exchange factor involved in chemoattractant-stimulated human neutrophil function. *J Leukoc Biol* 2007; **81**: 1127–1136.
35 Aktas E, Kucuksezer UC, Bilgic S, Erten G, Deniz G. Relationship between CD107a expression and cytotoxic activity. *Cell Immunol* 2009; **254**: 149–154.
36 Grutzkau A, Smorodchenko A, Lippert U, Kirchhof L, Artuc M, Henz BM. LAMP-1 and LAMP-2, but not LAMP-3, are reliable markers for activation-induced secretion of human mast cells. *Cytometry A* 2004; **61**: 62–68.
37 Dani A, Chaudhry A, Mukherjee P, Rajagopal D, Bhatia S, George A *et al.* The pathway for MHCII-mediated presentation of endogenous proteins involves peptide transport to the endo-lysosomal compartment. *J Cell Sci* 2004; **117**: 4219–4230.
38 Delamarre L, Pack M, Chang H, Mellman I, Trombetta ES. Differential lysosomal proteolysis in antigen-presenting cells determines antigen fate. *Science* 2005; **307**: 1630–1634.
39 Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008; **456**: 53–59.
40 Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005; **437**: 376–380.
41 Wang J, Wang W, Li R, Li Y, Tian G, Goodman L *et al.* The diploid genome sequence of an Asian individual. *Nature* 2008; **456**: 60–65.
42 Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009; **461**: 272–276.
43 Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010; **42**: 30–35.
44 Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DCY, Nazareth L *et al.* Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* 2010; **362**: 1181–1191.
45 Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 2010; **328**: 636–639.

Supplementary Information accompanies the paper on the The Pharmacogenomics Journal website (http://www.nature.com/tpj)