

ORIGINAL ARTICLE

Comprehensive genome-wide evaluation of lapatinib-induced liver injury yields a single genetic signal centered on known risk allele *HLA-DRB1*07:01*LR Parham¹, LP Briley¹, L Li¹, J Shen¹, PJ Newcombe², KS King¹, AJ Slater¹, A Dilthey³, Z Iqbal³, G McVean³, CJ Cox², MR Nelson¹ and CF Spraggs²

Lapatinib is associated with a low incidence of serious liver injury. Previous investigations have identified and confirmed the Class II allele *HLA-DRB1*07:01* to be strongly associated with lapatinib-induced liver injury; however, the moderate positive predictive value limits its clinical utility. To assess whether additional genetic variants located within the major histocompatibility complex locus or elsewhere in the genome may influence lapatinib-induced liver injury risk, and potentially lead to a genetic association with improved predictive qualities, we have taken two approaches: a genome-wide association study and a whole-genome sequencing study. This evaluation did not reveal additional associations other than the previously identified association for *HLA-DRB1*07:01*. The present study represents the most comprehensive genetic evaluation of drug-induced liver injury (DILI) or hypersensitivity, and suggests that investigation of possible human leukocyte antigen associations with DILI and other hypersensitivities represents an important first step in understanding the mechanism of these events.

The Pharmacogenomics Journal (2016) **16**, 180–185; doi:10.1038/tpj.2015.40; published online 19 May 2015

INTRODUCTION

Abnormal liver chemistry elevations observed during drug therapy may require treatment discontinuation and may result in poorer clinical outcomes. Identification of risk factors for such events may improve safety management and overall patient outcomes. Robust human leukocyte antigen (HLA) allele associations have been detected for a range of serious drug-induced adverse events.¹ For abacavir² and carbamazepine,³ such associations have been translated into clinical safety management using specific HLA allele screening to identify patients with higher risk of severe hypersensitivity reactions as a consequence of receiving the drug. In addition, specific HLA alleles have been associated with drug-induced liver injury (DILI) for a range of widely used medicines, including the antibiotics, flucloxacillin⁴ and amoxicillin-clavulanic acid.⁵ Furthermore, DILI has led to drug withdrawals from the market with termination of their development programmes, including ximelagatran⁶ and lumiracoxib.⁷ Further analysis of amoxicillin-clavulanic acid-induced liver injury has suggested additional associations with immune-related, non-HLA genes.⁵

Lapatinib (Tykerb/Tyverb, GlaxoSmithKline), a human epidermal growth factor receptor 2 and epidermal growth factor receptor inhibitor, is approved for administration in combination with other targeted chemotherapy agents for the treatment of patients with advanced or metastatic breast cancer whose tumors overexpress human epidermal growth factor receptor 2, and who have received prior cancer therapy, including anthracycline, a taxane and trastuzumab. Although extensive clinical experience shows an acceptable safety profile, lapatinib has been associated with liver injury and a low incidence of National Cancer Institute Common Terminology Criteria for Adverse Events grade 3 isolated alanine aminotransferase (ALT) elevation (3%) and potentially life-

threatening concurrent ALT elevation with hyperbilirubinemia (Hy's Law, 0.6%).⁸ Previous investigations of patients with breast cancer treated with lapatinib found that the Class II allele *HLA-DRB1*07:01* (and *DQA1*02:01*, which is in strong linkage disequilibrium) is associated with lapatinib-induced liver injury.⁹ These results were extended to show the association of this allele with lapatinib-induced Hy's Law cases.¹⁰ In addition to these retrospective analyses, this association was prospectively validated in a large, well-controlled, randomized clinical trial, where *HLA-DRB1*07:01* allele carriage was enriched in on-treatment grade 3 ALT elevation cases.¹¹ This Class II HLA association implicates immune mechanisms and identifies a patient subgroup at increased risk of this safety event. Awareness of these findings and the established widespread availability of HLA testing could potentially support clinical management of liver injury during lapatinib treatment. However, despite *HLA-DRB1*07:01* typing producing a high negative predictive value, suitable for safety biomarker prediction, this is accompanied by a moderate positive predictive value for identifying individual patients at higher risk.¹¹ This presents limitations in its use as a prospectively applied predictive biomarker, as the majority of *HLA-DRB1*07:01* allele-positive patients will not experience liver injury during 1 year of treatment and may receive benefit from lapatinib for the treatment of their breast cancer.

It is conceivable that additional genetic variants located within the major histocompatibility complex (MHC) locus or elsewhere in the genome may influence lapatinib-induced liver injury risk and potentially lead to a genetic test with improved predictive qualities. For example, there is evidence that in addition to an *HLA* association, a variant in *PTPN22*, a gene associated with a variety of autoimmune diseases, is also associated with amoxicillin-clavulanic

¹GlaxoSmithKline Research & Development, Research Triangle Park, NC, USA; ²GlaxoSmithKline Research & Development, Stevenage, UK and ³Department of Statistics, University of Oxford, Oxford, UK. Correspondence: Dr CF Spraggs, GlaxoSmithKline Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, UK. E-mail: colin.f.spraggs@gsk.com

Received 19 November 2014; revised 13 February 2015; accepted 26 March 2015; published online 19 May 2015

Table 1. Summary of analyses performed

Analysis type	End point	Number of subjects analyzed	Variants	Methods
GWAS	CC	34 cases (baseline ALT $\leq 1 \times$ ULN and max ALT $> 3 \times$ ULN) 810 controls (baseline and all ALT measures $\leq 1 \times$ ULN)	1 085 881 SNPs 478 HLA 4-digit classic alleles 1967 polymorphic HLA amino-acid residues	Logistic regression
WGS	QTA CC	1192 26 hepatologist adjudicated, possible Hy's cases/extreme ALT cases (19 HLA carriers) 19 matched controls (13 HLA carriers)	11 688 195 SNVs (5 703 813 SNVs with ≥ 4 copies observed and 98 565 lower frequency variants aggregated in HLA carriers) 1 270 946 indels (652 732 indels in HLA carriers)	Linear regression Burden test frequency weighted burden test C-alpha test

Abbreviations: ALT, alanine aminotransferase; GWAS, genome-wide association studies; HLA, human leukocyte antigen; QTA, quantitative trait analysis; SNP, single-nucleotide polymorphism; ULN, upper limit of normal; WGS, whole-genome sequencing.

acid-induced liver injury.⁵ To address this, we have taken two approaches: (i) a genome-wide association study (GWAS), assessing common genetic variants in a single, large, lapatinib monotherapy clinical trial of patients previously treated with chemotherapy for early-stage human epidermal growth factor receptor 2-positive breast cancer (EGF105485, TEACH, NCT00374322)^{11,12} and (ii) a whole-genome sequencing (WGS) study, assessing common and rare genetic variants in a limited number of possible Hy's cases and extreme ALT elevation cases and matched controls identified from multiple clinical trials.¹⁰ This represents one of the most thorough investigations of the genetics of a drug-induced adverse event to date.

MATERIALS AND METHODS

Patients and clinical trials

EGF105485 (TEACH: Tykerb evaluation after chemotherapy). EGF105485 was a placebo-controlled, multicenter, randomized Phase III trial, where women with human epidermal growth factor receptor 2-positive early-breast cancer were assigned to receive lapatinib or placebo daily for 12 months.¹² A total of 3146 patients were available for safety analyses of the TEACH study: 1572 and 1574 patients were randomized to lapatinib and placebo, respectively, and were well-balanced for baseline characteristics. The pharmacogenetic experiment included 1194 and 1259 patients randomized to lapatinib and placebo, respectively, who consented to provide a blood sample for pharmacogenetic analysis.¹¹ DNA was isolated from whole blood.

To assess genetic variants across the genome for association with ALT elevations, ALT end points, including both case-control status using analysis definitions described previously⁹ and maximum on-treatment ALT, to allow for greater power to detect genetic effects of smaller size, were analyzed. For quantitative trait analysis, maximum on-treatment ALT (expressed as a fraction of the laboratory-specific upper limit of normal (ULN)) was derived from values from initiation of lapatinib treatment and up to 30 days after last dose of lapatinib treatment. All values were log-transformed (base 10) to account for the highly skewed distribution of this quantitative measure and make it suitable for parametric model fitting.

Possible Hy's law and extreme ALT cases and matched controls. Possible Hy's law and extreme ALT cases during lapatinib treatment, and matched controls with consented pharmacogenetic samples were selected for analysis. Twenty patients from seven clinical trials were identified as possible Hy's Law cases and were adjudicated by hepatologist experts as having 'probable' ($n=13$) or 'possible' ($n=6$) DILI or 'inadequate information' ($n=1$).¹⁰ In addition, six extreme ALT cases ($>10 \times$ ULN) from EGF105485 were identified. Possible controls were identified as individuals with normal baseline and on-treatment liver chemistry measures for ALT and total bilirubin, from 13 previously analyzed clinical trials.⁹ The DILI cases were identified by serious adverse event reporting in ongoing clinical trials. At the time of matching, some studies that provided DILI cases were ongoing, randomised, blinded clinical trials, where the corresponding data to identify matched controls from these same studies were not yet available. If no well-matched control was available from an alternate study at the time, these cases were genotyped without a matched control.

Controls were matched to cases by assessing the following variables of interest: *HLA-DQA1*02:01*, *HLA-DRB1*07:01*, *HLA-DQB1*02:02* and *TNXB* (rs12153855)⁹ and *UGT1A1*28* (ref. 10) genotypes, clinical trial, country and comparable genetic ancestry (using principal components generated from available GWAS data). This resulted in 19 cases ($n=15$ possible Hy's and $n=4$ extreme ALT) having a best-matched control assigned to them, and seven cases without any well-matched controls. A total of 26 cases and 19 matched controls were selected for further analyses. Primary analyses were conducted in cases and matched controls who carried the known, validated biomarker, *HLA-DRB1*07:01* to reduce case misclassification.

Note that all of the extreme ALT cases ($n=6$) and some of the possible Hy's Law cases ($n=5$) selected from EGF105485 were included in both GWAS and WGS studies presented herein.

Liver chemistry

Serum ALT, aspartate-aminotransferase, total bilirubin and alkaline phosphatase measurements were performed at multiple laboratories. Original laboratory values were converted to the multiples of ULN by dividing by laboratory-specific ULN values. The frequency of liver chemistry testing and monitoring was variable across clinical trials.

Data generation

GWAS. GWAS SNP data were generated for all subjects who consented to genetic analysis using the Illumina Human1M-Duo or the Illumina HumanOmni1-Quad BeadChip arrays by Expression Analysis (Durham, NC, USA). Classical 4-digit HLA genotyping was performed by Laboratory Corporation (Burlington, NC, USA) and Histogenetics (Ossining, NY, USA). *UGT1A1*28* was assessed using GeneScan fragment analysis by sequencing at GlaxoSmithKline (Philadelphia, PA, USA).

Candidate gene variants included the following four categories: (i) classical HLA alleles for *HLA-A*, *-B*, *-C*, *-DQA1*, *-DQB1*, *-DPB1* and *-DRB1*, (ii) functional variants from DILI genes with potential lapatinib metabolite interaction (*ABCB11* and *UGT1A1*), (iii) functional variants from lapatinib metabolism and disposition genes (*CYP3A4*, *CYP3A5*, *CYP2C8*, *CYP2C19*, *SLCO1B1*, *ABCB1*, *ABCG2*, *NR1I2* and *NR1I3*) and (iv) 213 autoimmune disease associated markers with reported *P*-values at genome-wide significance levels from the literature.⁵ The variants in categories (ii) to (iv) were present as SNPs on the GWAS platform.

WGS. WGS was performed at BGI (Shenzhen, China). DNA was fragmented and size selected to create 2×180 base pair libraries and 2×800 bp libraries. These libraries were interrogated on an Illumina HiSeq 2000 to generate 90 base pair paired end reads. For each sample ~ 100 gigabases of sequence was generated, resulting in at least $30 \times$ sequence coverage. Individual genotypes in the form of variant call format files were generated by aligning the short reads to GRCh37 using Stampy¹³ and variants were called using Platypus.¹⁴ All variants were annotated using the Variant Effect Predictor developed by Ensembl.¹⁵

Statistical methods

Table 1 summarises the analyses that were conducted.

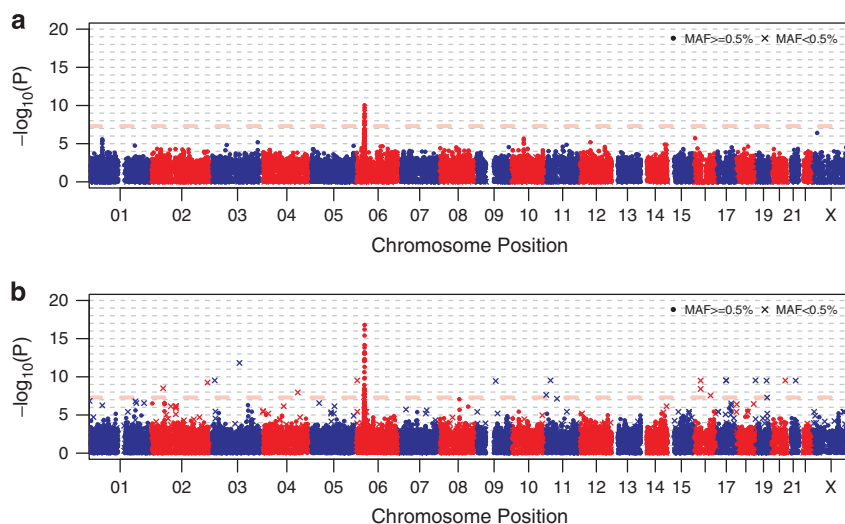


Figure 1. Genome-wide association analyses of (a) cases and controls and (b) maximum on-treatment ALT. The dashed red line indicates the Bonferroni-adjusted significance threshold (5×10^{-8}).

GWAS. To identify common genetic variants (minor allele frequency (MAF) $> 0.5\%$) associated with on-treatment ALT elevation across the genome, both quantitative trait and case-control analyses were carried out in all patients in the lapatinib-treated arm, under an additive genetic model, using linear or logistic regression. Analyses were conducted using the R statistical software. The known biomarker, *HLA-DRB1*07:01*, is highly correlated with *HLA-DQA1*02:01*;¹¹ however, for the purpose of this experiment, *HLA-DRB1*07:01* was chosen for conditioning. Regression analysis adjusting for *HLA-DRB1*07:01* genotype was carried out to assess whether associated variants were due to linkage disequilibrium with the known biomarker. Key covariates, including age, baseline ALT and genetic ancestry (as principal component 1–6 scores derived using GWAS data) were adjusted for in all analyses. Additional sensitivity analyses were conducted in principal component analysis-defined White ($N = 775$) and East Asia ($N = 236$) subgroups for quantitative traits adjusting for age and baseline ALT to check the consistency of top results by race.

HLA analyses at the resolution of amino-acid residues may provide additional statistical power and insight into biological mechanism compared with analysis at the level of classical HLA alleles and MHC region SNPs. By following standard definitions for HLA allele sequences from the EMBL-EBI Immunogenetics HLA Database,¹⁶ we extracted the amino-acid sequences and residues for each observed four-digit HLA classic allele.^{17,18} Multi-allelic amino-acid positions were encoded as multiple di-allelic amino acids, each corresponding to a specific amino acid. Association analysis for each di-allelic amino acid was carried out using the same association analysis method as for the bi-allelic SNP.

For declaring experiment-wide statistical significance, Bonferroni-threshold criteria 1×10^{-4} (0.05/500) and 5×10^{-8} (0.05/1 000 000) were applied to candidate gene and GWAS results, reflecting the number of variants tested, respectively.

WGS. To reduce power loss due to case misclassification, the main analysis was restricted to *HLA-DRB1*07:01* carriers only. Additional sensitivity analyses were conducted using only the five cases with ALT $> 20 \times \text{ULN}$ and total bilirubin $> 2 \times \text{ULN}$ —a much more stringent definition. As all statistical analyses were limited to *HLA-DRB1*07:01* carriers, any associations identified would be independent of the known HLA biomarker. To avoid fitting too many parameters, with such a small sample size, none of the analyses adjusted for covariates in the first instance, resulting in a generally liberal test. Any statistically interesting results ($P < 1 \times 10^{-7}$) would have been further interrogated by inspecting distributions of key factors such as age and genetic principal component scores.

All variants, including SNPs and small insertions and deletions (indels), with four or more copies observed were analyzed one at a time for association with liver injury using Fisher's exact test. In the analysis of the 32 HLA carriers (19 cases and 13 controls) this corresponds to including all variants with a sample MAF $> 6\%$. In addition to the single-variant tests for common variants in this sample, we conducted gene-wide tests on the

rarer variants. Within each gene, three classes of variants with less than four copies were aggregated, according to increasing belief in likelihood of functional impact: (i) all exonic variants, and intronic variants annotated to reside in a splice site or to be highly conserved, (ii) all non-synonymous variants and (iii) variants predicted to be 'probably damaging' by Polyphen-2 (ref. 19) and/or 'deleterious' by SIFT (Sorting Intolerant From Tolerant, a tool to predict whether an amino-acid substitution affects protein function).²⁰ For each aggregate test, three statistical methods were explored: (i) burden test, (ii) frequency weighted burden test and (iii) c-alpha test (see Supplementary Information for more details). With all tests, to avoid violation to asymptotic assumptions due to the small sample size, permutation was used to generate *P*-values. The PLINK/SEQ software suite (<http://pngu.mgh.harvard.edu/~purcell/plink/>, <http://atgu.mgh.harvard.edu/plinkseq/>) was used to perform all statistical tests on the WGS data.

RESULTS

GWAS

1 085 881 GWAS SNPs were successfully genotyped and were polymorphic in the current sample, among which 1 010 443 (93%) were common (MAF $> 0.5\%$). Candidate gene variants included the *UGT1A1*28* allele, 216 common SNPs from GWAS and 478 HLA 4-digit classic alleles (248 alleles with MAF $> 0.5\%$). A total of 1967 polymorphic HLA amino-acid variants at 1411 unique polymorphic positions were identified based on HLA classic allele types, where 1349 (69%) amino-acid variants were common in the analyzed data.

In the ALT case-control analyses, only variants in the MHC region were statistically significant (Figure 1a, *HLA-DRB1*07:01*, $P = 7.8 \times 10^{-11}$). However, no statistically significant associations remained upon conditioning on *HLA-DRB1*07:01* carriage (Supplementary Figure 1a). After conditioning on *HLA-DRB1*07:01* carriage, there is no enrichment of association results above what would be expected under the null distribution (Supplementary Figure 2).

In the analysis of ALT as a quantitative trait, again only genetic variants in the MHC region were statistically significant (Figure 1b, *HLA-DRB1*07:01*, $P = 2.0 \times 10^{-18}$). Upon conditioning on *HLA-DRB1*07:01* carriage, one intronic variant in *TPD52* (rs7828135, MAF = 5%, $P = 4.5 \times 10^{-8}$) passed the pre-specified GWAS significance threshold (Supplementary Figure 1b) and *HLA-DQA1*03:03* (MAF = 6%, $P = 6.5 \times 10^{-5}$) passed the pre-specified candidate gene significance threshold. Neither of these genetic variants remained statistically significant at the pre-specified levels after conditioning on *HLA-DRB1*07:01* carriage in the case-control

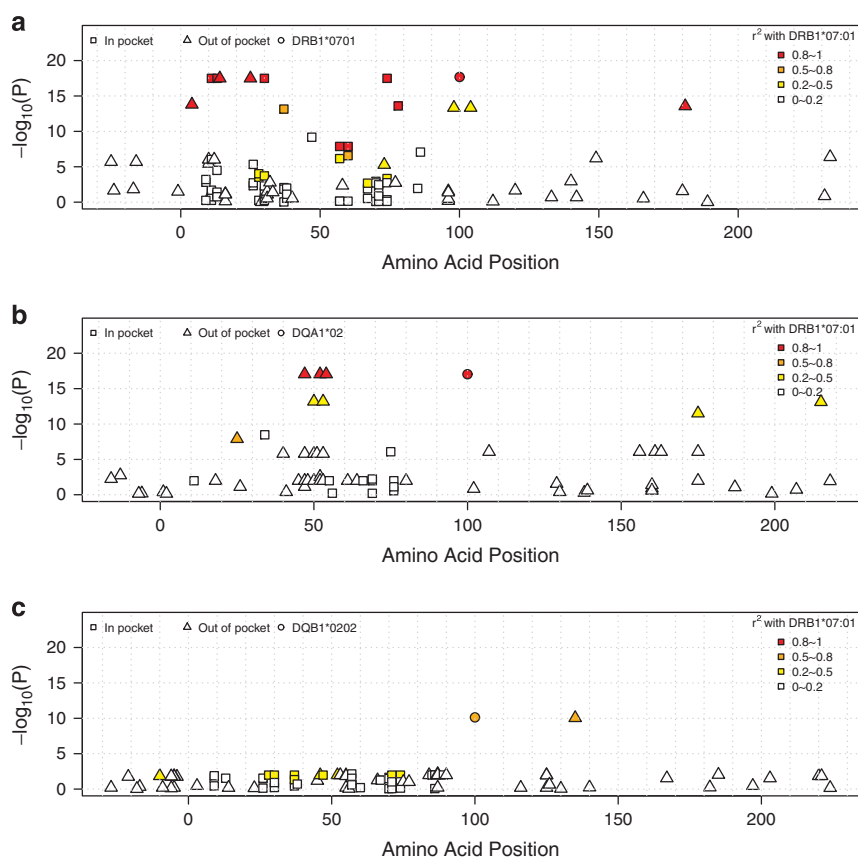


Figure 2. Association of HLA amino acids with maximum on-treatment ALT for (a) *HLA-DRB1*, (b) *HLA-DQA1* and (c) *HLA-DQB1*. Amino acids are differentiated between those that are located inside and outside of the antigen-binding pocket.

analyses. Sensitivity analyses within principal component analysis-defined White and East Asia subgroups showed similar effects for these top variants (Supplementary Figure 3).

The combined analysis of GWAS SNPs, HLA alleles and derived HLA amino acids showed that the association signal centers around the previously validated biomarker *HLA-DRB1*07:01* (Supplementary Figure 4). To attempt to discriminate functionality between the correlated alleles, regression analysis of the amino acids was conducted and identified several amino acids with equivalent statistical associations. Of the most strongly associated amino acids, several are located within the antigen-binding pocket of *HLA-DRB1* (Figure 2a), as compared with none of those within the antigen-binding pocket of *HLA-DQA1* (Figure 2b). We did not see any evidence of an association with *HLA-DQB1*-specific amino acids either within or outside the antigen-binding domain pocket (Figure 2c).

Review of binding pocket amino-acid specificity and heterodimer formation of DR and DQ suggest that two HLA-DQ heterodimers (DQ2.2 and DQ9.2) would be required to account for the association with lapatinib-induced ALT elevation, whereas for HLA-DR, a single heterodimer (DR7) is sufficient.¹¹ These results support the causal role of *HLA-DRB1*07:01* in lapatinib-induced liver injury.

WGS

After filtering on missingness, 11 688 195 SNPs and 1 270 946 indels were available for analysis across all 45 subjects.

For the association analysis among *HLA-DRB1*07:01* carriers, ~5.7 million SNPs and 653 000 indels with four or more copies observed were analyzed individually.

No individual variants analyzed showed a statistically significant association (Figure 3a). Both the QQ plot for 6.4 million tests conducted genome-wide, and for a subset of 44 000 tests conducted in the exome (Supplementary Figure 5), where our prior probability of causal variant discovery is higher, showed fewer tests with small P -values than expected under the null distribution. This observation was not surprising, given the small sample size. The sensitivity analysis in the extreme cases and controls also did not yield any statistically significant variants (Figure 3b). In addition, we reviewed the WGS data from the five extreme cases for potentially pathogenic variations, focusing on variants associated with immune response, drug metabolism and liver injury, through the Ingenuity Variant Analysis software package (Qiagen Redwood City, CA, USA). No additional variants of interest, outside the *HLA* genes, were identified through this process.

Despite the huge number of tests, this experiment was powered to find a perfect genetic biomarker of liver injury that would be present in all cases and none of the controls. A variant that perfectly discriminates cases from controls among HLA carriers would achieve $P = 2.9 \times 10^{-9}$ under Fisher's exact test, and therefore reach statistical significance (Supplementary Table 1). As we have not seen such a predictor, we can conclude that one is unlikely to exist within the regions of the genome that were successfully sequenced.

DISCUSSION

This comprehensive genetic evaluation has resulted in one clear signal—*HLA-DRB1*07:01*—being associated with lapatinib-induced liver injury. This HLA association is consistent with a

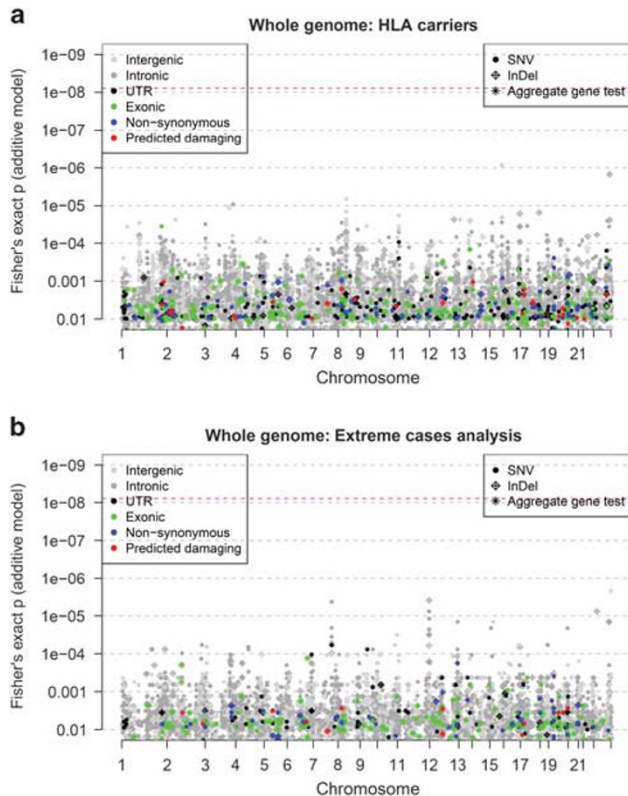


Figure 3. Genome-wide sequence analysis in *HLA-DRB1*07:01* carriers in (a) all 19 sequenced cases and (b) only the most extreme 5 cases versus 13 controls. The dashed red lines indicate the Bonferroni significance threshold ($P=8 \times 10^{-9}$). The functional predictions for each variant are indicated by color and the variant type by symbol, as shown within the figure.

number of other GWAS for immune-related safety events, where key genetic risk factors are captured in HLA variation.^{1,21}

In the GWAS, the genetic signal centered around *HLA-DRB1*07:01* remained the strongest association. Moderately significant results of two genetic variants found in the conditional quantitative trait analyses were not significant in case-control analyses, and did not lend themselves to be clinically useful predictors, with no substantial increase in the positive predictive values of the known biomarker. Although *HLA-DRB1*07:01* is in strong linkage disequilibrium with other genetic variants in the MHC region, results from investigation of HLA heterodimer formation and review of binding pocket amino-acid specificity suggest that *HLA-DRB1*07:01* is the most likely causal allele. Reasonable power was available to detect genetic associations of a reasonable effect size in the quantitative (Supplementary Figure 6) and case-control (Supplementary Figure 7) GWAS analyses. Small genetic effects that may have been underpowered in this study, and subsequently not detected, would not have resulted in a significant impact on improving clinical utility.

The additional WGS experiment added little, if any, value to the overall conclusions. No additional rare variants were identified that were statistically significant or had a strong biological rationale. Furthermore, given the challenges of interrogating the MHC region accurately using WGS data,²² and the opportunity to accurately impute HLA genotypes using GWAS data,²³ one could argue that the GWAS approach is more likely to identify potential causative MHC variants than WGS approaches, questioning the added value of using WGS and whole-exome sequencing to investigate relatively common adverse events, as recently proposed,²⁴ over the use of traditional GWAS arrays.

Focusing on the clinical utility of *HLA-DRB1*07:01*, the marker's high negative predictive value and moderate positive predictive value,¹¹ describe how essentially all individuals who experience serious lapatinib-induced liver injury carry this biomarker, but the majority of biomarker carriers do not develop serious liver injury. This suggests that there may be additional triggers or agents contributing to this safety event and is consistent with other HLA associations related to drug-induced hypersensitivity.^{1,21} Although this comprehensive genetic evaluation of a relatively common event has ruled out the possibility of a large, clinically meaningful effect existing, additional rare genetic factors of smaller effect size may still exist that impact the biology of this event, but would not significantly improve the clinical utility. Furthermore, additional, non-genetic components may exist that determine propensity to produce this adverse event among *HLA-DRB1*07:01* allele carrying patients.²¹

The present comprehensive genetic evaluation did not yield any additional genetic predictors to enable improved clinical risk prediction. However, the single, robust association of *HLA-DRB1*07:01* provides important insight into mechanism of action of this safety event and represents an important first step in understanding the mechanism of these events. As a result of this comprehensive genetic evaluation, it is unlikely that further genetic marker identification experiments would be fruitful and suggest that future work should focus on immune cell biology^{25,26} to determine additional biological restrictions to immune activation and continued understanding around the potential clinical utility of this known biomarker for lapatinib-induced liver injury.

CONFLICT OF INTEREST

LRP, LPB, LL, JS, KSK, AJS, CJC, MRN and CFS are employees of and hold stocks/stock options in GlaxoSmithKline. PJN was an employee of GlaxoSmithKline at the time of the study. GM has received research support from GlaxoSmithKline. AD and ZI declare no conflict of interest.

ACKNOWLEDGMENTS

This study was funded by GlaxoSmithKline. The manuscript was written by the authors with no third party writing support. We acknowledge the excellent support of GlaxoSmithKline colleagues in Genetics, Statistical Genetics, Oncology Medicines Development, Molecular Discovery Research and the GlaxoSmithKline Hepatotoxicity Board. We also thank the study investigators for their diligence in providing study and case narrative data and patients for generously providing their clinical data and DNA for this investigation. Funding for the study was provided by GlaxoSmithKline.

REFERENCES

- Phillips EJ, Mallal SA. Pharmacogenetics of drug hypersensitivity. *Pharmacogenomics* 2010; **11**: 973–987.
- Mallal SA, Phillips EJ, Carosi G, Molina JM, Workman C, Tomazic J et al. HLA-B*5701 screening for hypersensitivity to abacavir. *New Engl J Med* 2008; **358**: 568–579.
- Chen P, Lin JJ, Lu CS, Ong CT, Hsieh PF, Yang CC et al. Carbamazepine-induced toxic effects and HLA-B*1502 screening in Taiwan. *New Engl J Med* 2011; **364**: 1126–1133.
- 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.
- Lucena MI, Molokhia M, Shen Y, Urban TJ, Aithal GP, Andrade RJ et al. Susceptibility to amoxicillin-clavulanate-induced liver injury is influenced by multiple HLA class I and II alleles. *Gastroenterology* 2011; **141**: 338–347.
- Kindmark A, Jawaid A, Harbron CG, Barratt BT, Bengtsson OF, Andersson TB et al. Genome-wide pharmacogenetic investigation of a hepatic adverse event without clinical signs of immunopathology suggests an underlying immune pathogenesis. *Pharmacogenomics J* 2008; **8**: 186–195.
- Singer JB, Lewitzky S, Leroy E, Yang F, Zhao X, Klickstein L et al. A genome-wide study identifies HLA alleles associated with lumiracoxib-related liver injury. *Nat Genet* 2010; **42**: 711–714.
- Moy B, Rappold E, Williams L, Kelly T, Nicolodi L, Maltzman JD et al. Hepatobiliary abnormalities in patients with metastatic cancer treated with lapatinib. *J Clin Oncol* 2009; **27**: 1043.

- 9 Spraggs CF, Budde LR, Briley LP, Bing N, Cox CJ, King KS *et al*. HLA-DQA1*02:01 is a major risk factor for lapatinib-induced hepatotoxicity in women with advanced breast cancer. *J Clin Oncol* 2011; **29**: 667–673.
- 10 Spraggs CF, Parham LR, Hunt CM, Dollery CT. Lapatinib-induced liver injury characterized by Class II HLA and Gilbert's syndrome genotypes. *Clin Pharmacol Ther* 2012; **91**: 647–652.
- 11 Schaid DJ, Spraggs CF, McDonnell SK, Parham LR, Cox CJ, Ejlersen B *et al*. Prospective validation of HLA-DRB1*07:01/DQA1*02:01 allele carriage as a predictive risk factor for lapatinib-induced liver injury. *J Clin Oncol* 2014; **32**: 2296–2303.
- 12 Goss PE, Smith IE, O'Shaughnessy J, Ejlersen B, Kaufmann M, Boyle F *et al*. Adjuvant lapatinib for women with early-stage HER2-positive breast cancer: a randomised, controlled, phase 3 trial. *Lancet Oncol* 2013; **14**: 88–96.
- 13 Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res* 2011; **21**: 936–939.
- 14 Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRWGS500 Consortium *et al*. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet* 2014; **46**: 912–918.
- 15 McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 2010; **26**: 2069–2070.
- 16 Robinson J, Waller MJ, Fail SC, McWilliam H, Lopez R, Parham P *et al*. The IMGT/HLA database. *Nucleic Acids Res* 2009; **37**: D1013–D1017.
- 17 Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X *et al*. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* 2012; **44**: 291–296.
- 18 Jia X, Han B, Onengut-Gumuscu S, Chen W-M, Concannon PJ, Rich SS *et al*. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One* 2013; **8**: e64683.
- 19 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P *et al*. A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248–249.
- 20 Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; **4**: 1073–1081.
- 21 Bharadwaj M, Illing P, Theodossis A, Purcell AW, Rossjohn J, McCluskey J. Drug hypersensitivity and human leukocyte antigens of the major histocompatibility complex. *Annu Rev Pharmacol Toxicol* 2012; **52**: 401–431.
- 22 Dilthey A, Cox C, Iqbal Z, Nelson MR, McVean G. A population reference graph for diploid genome inference. 2014. Available at <http://biorxiv.org/content/early/2014/07/08/006973>. Last accessed 30 September 2014.
- 23 Zheng X, Shen J, Cox C, Wakefield JC, Ehm MG, Nelson MR *et al*. HIBAG—HLA genotype imputation with attribute bagging. *Pharmacogenomics* 2014; **14**: 192–200.
- 24 Urban TJ, Goldstein DB. Pharmacogenetics at 50: genomic personalization comes of age. *Sci Transl Med* 2014; **6**: 220ps1.
- 25 Chessman D, Kostenko L, Lethborg D, Purcell AW, Williamson NA, Chen Z *et al*. Human leukocyte antigen class I-restricted activation of CD8+ T cells provides the immunogenetic basis of a systemic drug hypersensitivity. *Immunity* 2008; **28**: 822–832.
- 26 Monshi M, Faulkner L, Gibson A, Jenkins RE, Farrell J, Earnshaw CJ *et al*. Human Leukocyte Antigen (HLA)-B*57:01- restricted activation of drug-specific T cells provides the immunological basis for flucloxacillin-induced liver injury. *Hepatology* 2013; **57**: 727–739.



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

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