Agranulocytosis is a serious, although rare, adverse reaction to sulfasalazine, which is used to treat inflammatory joint and bowel disease. We performed a genome-wide association study comprising 9,380,034 polymorphisms and 180 HLA alleles in 36 cases of sulfasalazine-induced agranulocytosis and 5,170 population controls. Sulfasalazine-induced agranulocytosis was significantly associated with the HLA region on chromosome 6. The top hit (rs9266634) was located close to HLA-B, odds ratio (OR) 5.36 (95% confidence interval (CI) (2.97, 9.69) P = 2.55 × 10⁻⁸). We HLA-sequenced a second cohort consisting of 40 cases and 142 treated controls, and confirmed significant associations with HLA-B*08:01, OR = 2.25 (95% CI (1.02, 4.97) P = 0.0439), in particular the HLA-B*08:01 haplotype HLA-DQB1*02:01-DRB1*03:01-B*08:01-C*07:01, OR = 3.79 (95% CI (1.63, 8.80) P = 0.0019), and with HLA-A*31:01, OR = 4.81 (95% CI (1.52, 15.26) P = 0.0077). The number needed to test for HLA-B*08:01 and HLA-A*31:01 to avoid one case was estimated to be 1,500. We suggest that intensified monitoring or alternative treatment should be considered for known carriers of HLA-B*08:01 or HLA-A*31:01.

Study Highlights

**WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?**
Agranulocytosis is a serious, although rare, adverse reaction to sulfasalazine—a drug used to treat inflammatory joint and bowel disease. The risk is most pronounced during the first 3 months of treatment, and agranulocytosis should be suspected in patients with unexplained fever or nonspecific illness during this period. There are no good predictors for sulfasalazine-induced agranulocytosis.

**WHAT QUESTION DID THIS STUDY ADDRESS?**
The aim was to identify genetic variants predisposing to sulfasalazine-induced agranulocytosis in a European population.

**WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?**
We performed a genome-wide association study (GWAS) and identified genetic variants in the HLA region that increased the risk of sulfasalazine-induced agranulocytosis significantly.

**HOW THIS MIGHT CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?**
We suggest that known carriers of HLA-B*08:01 or HLA-A*31:01 should be placed under intensified monitoring when starting sulfasalazine or be offered an alternative drug.

Agranulocytosis is a blood dyscrasia characterized by an absolute neutrophil count below 0.5 × 10⁹/L (500/µL). The vast majority of cases are drug-related, either through direct dose-dependent cytotoxicity or idiosyncratic reactions to at least 125 nonchemotherapy drugs. Idiosyncratic agranulocytosis leads to a rapid decline in neutrophils that causes a high risk of overwhelming infection, with an average mortality rate of about 5%. The onset of drug-induced idiosyncratic agranulocytosis is...
usually within months of starting therapy, and the reaction is generally believed to be immune-mediated. In the few instances where genetic studies have been performed, associations with human leukocyte antigen (HLA) genes or other genes associated with the immune response have been found. This provides further evidence for an immune-mediated mechanism for drug-induced idiosyncratic agranulocytosis.

Sulfasalazine (Salazopyrin, Pfizer, Groton, CT), a drug developed in the 1930s for the treatment of inflammatory joint and bowel disease, was reported to cause fatal idiosyncratic agranulocytosis already in 1942. In nonfatal cases, patients usually recover within weeks of drug withdrawal. The incidence was estimated to be 1 in 2,400 patients during the first month of treatment, 1 in 700 during the second and third months (0.14%), and 1 in 11,200 after more than 3 months of treatment in a Swedish population. In the same population, the background incidence of agranulocytosis was approximately 1 in 150,000 inhabitants.

Sulfasalazine-induced agranulocytosis is believed to be a host immune response to a reactive metabolite of sulfasalazine. Sulfasalazine is first split into 5-amino salicylic acid and sulfapyridine by intestinal bacteria. Sulfapyridine is in contrast to 5-amino salicylic acid rapidly absorbed into the bloodstream. It is further metabolized in the liver into N-acetyl-sulfapyridine by the enzyme N-acetyltransferase 2 (NAT2). N-acetyl-sulfapyridine is primarily bioactivated by the enzyme CYP2C9 into a cytotoxic hydroxylamine that is subsequently excreted as a glucuronic acid conjugate. Inherited variants of the drug-metabolizing enzymes NAT2 or CYP2C9 could in theory shunt the metabolism towards more toxic metabolites.

In a previous study, we investigated whether coding variants in the NAT2 gene influenced the risk of sulfasalazine-induced agranulocytosis. A higher proportion of slow acetylators was seen among 39 cases of sulfasalazine-induced agranulocytosis (69%) compared with 75 sulfasalazine-tolerant patients (45%). However, no difference in slow acetylator frequency was seen between the 39 cases and 448 population-based control subjects. We further genotyped these cases and treated controls for CYP2C9*2 and *3, but no association between agranulocytosis and genotype was seen (unpublished results). A subsequent study investigated the possible association between sulfasalazine-induced agranulocytosis and mutations in the human neutrophil elastase gene (ELANE) that causes severe congenital neutropenia (SCN). None of the mutations previously reported to cause SCN were found in 36 patients with agranulocytosis induced by sulfasalazine. No other studies on candidate genes for sulfasalazine-induced agranulocytosis have to our knowledge been published. The European Drug-induced Agranulocytosis Consortium (EuDAC, see Acknowledgments) was initiated with the aim to perform a genome-wide association study (GWAS) to identify novel genetic variants predisposing to agranulocytosis. We here present the results on sulfasalazine-induced agranulocytosis.

RESULTS

Characteristics of 36 cases of agranulocytosis induced by sulfasalazine and 5,170 population controls collected in Sweden, Germany, France, and Spain are shown in Supplementary Table S1. Comparisons of demographic and clinical factors between the 30 cases recruited in Sweden and their 183 matched controls are shown in Table 1.

Genome-wide association analyses

Sulfasalazine-induced agranulocytosis was associated with the major histocompatibility complex (MHC) region (HLA region) on chromosome 6 on a genome-wide level after adjusting according to sex and genetic principal components 1–4 (Figure 1, Supplementary Figure S1). The single nucleotide polymorphism (SNP) with the best evidence for association was rs9266634, which is located in an intergenic region 22,000 bases from HLA-B, odds ratio (OR) 5.36 (95% confidence interval (CI) 2.97, 9.69) P = 2.55 × 10⁻⁸ (Table 2, Supplementary Table S2, Supplementary Figure S1). There was also a significant association with the intergenic insertion rs202233001 positioned in a repeat region close to the paired box 1 gene (PaX1) on chromosome 20, OR = 6.76 (95% CI 3.41, 13.38), P = 4.19 × 10⁻⁸. When the data were adjusted for the top hit rs9266634, the chromosome 20 SNP rs202233001 was no longer significantly associated.

Follow-up analyses were performed adjusting for each top hit in a sequential manner (Table 2). After adjusting for rs9266634 on chromosome 6, rs111876221 on chromosome 5 reached genome-wide significance. This SNP is located in intron 1 of the serine incorporator 5 gene (SERINC5). After adjusting for the two independently associated rs9266634 and rs111876221, the SNP rs113887891 on chromosome 1 was significant on a genome-wide level. This intron variant is in a repeat region in the gene for the nonprotein-coding transcript LINC01762. After adjusting also for rs113887891, the SNP rs12082628 on chromosome 1 was significant on a genome-wide level. This variant is in a repeat region in an intron of the cholinergic receptor, muscarinic 3 gene (CHRM3). All associated variants outside the HLA region were infrequent, and none of them had strong evidence of being functional in primary B cells from peripheral blood.

Analysis of classical HLA alleles at second field resolution

The HLA alleles HLA-C*14:02 and HLA-B*08:01 were associated with sulfasalazine-induced agranulocytosis on an HLA-wide level (P < 2.8 × 10⁻⁸), OR = 7.72 (95% CI 2.90, 20.58) P = 4.36 × 10⁻⁵) and 2.92 (95% CI 1.71, 4.98) P = 8.37 × 10⁻³, respectively (Table 2, Supplementary Table S3). The top genome-wide SNP rs9266634 was in linkage disequilibrium (LD) with both HLA-C*14:02 (D² = 0.87, r² = 0.01, P < 2.22 × 10⁻¹⁶) and HLA-B*08:01 (D² = 0.99, r² = 0.16, P < 2.22 × 10⁻¹⁶). A multiple analysis, including all associated HLA alleles and the top SNP, shows that rs9266634 had a small independent effect (Supplementary Figure S2). Top HLA types obtained by sequential adjustment for each significantly associated type are
shown in Table 2. The association with HLA-B*08:01 was strengthened after adjusting for HLA-C*14:02, OR = 3.24 (95% CI (1.98, 5.57) \( P = 2.26 \times 10^{-3} \)). HLA-A*31:01 consistently had odds ratios around 4, but \( P \)-values were above the defined HLA-wide significance level.

We corrected for possible confounding by indication and population stratification by considering the Swedish cohort separately (Table 1). The odds ratio for HLA-C*14:02 was 11.08 (95% CI (4.10, 29.89) \( P = 2.05 \times 10^{-2} \)) when Swedish cases and population controls were compared, and 10.30 (95% CI (2.44, 43.52) \( P = 1.51 \times 10^{-3} \)) when Swedish cases were compared with matched controls. The corresponding odds ratios for HLA-B*08:01 were 3.15 (95% CI (1.78, 5.57) \( P = 7.83 \times 10^{-3} \)), and 4.97 (95% CI (1.54, 16.01) \( P = 7.25 \times 10^{-3} \)) (Figure 2).

**Replication of SNPs and HLA alleles**

An independent cohort recruited in Sweden was used for replication of the association with HLA (Table 3). We sequenced HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DQB1, and HLA-DPB1 in 40 cases of sulfasalazine-induced agranulocytosis and 142 controls treated with sulfasalazine for at least 90 days. Due to correction for multiple testing of the five top HLA types shown in Table 2, the cutoff for significance was \( P < 0.01 \). HLA-B*08:01 was nominally associated with agranulocytosis, OR = 2.25 (95% CI (1.02, 4.97) \( P = 0.0439 \), C-statistic = 0.57) (Table 4). Thirteen cases (32.5%) were heterozygous for HLA-B*08:01, as compared to 25 controls (17.6%), and none were homozygous. We hypothesized that these patients carried the most frequent HLA-B*08:01 haplotype, HLA-DQB1*02:01-
DRB1*03:01-B*08:01-C*07:01, which is present in 7.4% of Europeans. Notably, all 13 cases with HLA-B*08:01 carried HLA-C*07:01, DQB1*02:01, and DRB1*03:01, and thus probably had the common haplotype. In comparison, only 16 of 25 sulfasalazine-treated controls with HLA-B*08:01 carried this combination of HLA alleles. The odds ratio for carrying HLA-DQB1*02:01-DRB1*03:01-B*08:01-C*07:01 was 3.79 (95% CI (1.63, 8.80) \( P = 0.0019, \) C-statistic = 0.61), which was statistically significant (Table 4). Furthermore, the association with HLA-A*31:01 was statistically significant, OR = 4.81 (95% CI (1.52, 15.26) \( P = 0.0077, \) C-statistic = 0.57) (Table 4). The allele frequency of HLA-A*31:01 was high among cases (9%) compared with the expected frequency 2.7% in Europeans. HLA-C*14:02 was only present in one control, and therefore too rare to be assessed.

**Predictive ability and clinical implications**

The specificity to predict risk of agranulocytosis increased from 82.4% to 88.7% when using HLA-DQB1*02:01-DRB1*03:01-B*08:01-C*07:01 instead of HLA-B*08:01. This indicates that fewer patients would be falsely predicted as being at risk when using the full haplotype. The sensitivity was 32.5% with either HLA-B*08:01 or the full haplotype, but when combining HLA-B*08:01 with HLA-A*31:01, the sensitivity increased to 47.5%. With a sensitivity of 47.5% and an incidence of 0.0014 (1 in 700 patients receiving treatment), we could theoretically reduce the incidence of agranulocytosis to 0.00075 (0.0014−0.475\*0.0014) by screening for HLA-B*08:01 and HLA-A*31:01. The number needed to genotype (NNG) for HLA-B*08:01 and HLA-A*31:01 to avoid one case was estimated to be about 1,500, which is the reciprocal of the absolute risk reduction, 1/(0.0014\*0.475).

**DISCUSSION**

The sulphamidine sulfasalazine is known to carry a high risk of agranulocytosis, the relative risk is estimated to be almost 10 times higher than for the sulphamidine sulfamethoxazole. The characteristics of our cases (53% women, median age 55–59 years, median treatment time to onset 50 days) were similar to those found in a previous study on the epidemiology of sulfasalazine-induced agranulocytosis in Sweden. In the present study, sulfasalazine-induced agranulocytosis was mainly associated with the MHC region on chromosome 6 that encodes HLA genes. Agranulocytosis induced by other drugs has previously been associated with HLA. We found an association between sulfasalazine-induced agranulocytosis and HLA-B*08:01 that was replicated with similar ORs (2.92 vs. 2.25). All HLA sequenced cases carrying HLA-B*08:01 also carried DQB1*02:01, DRB1*03:01, and HLA-C*07:01, and the full haplotype conferred a higher OR (3.79). This indicates that the HLA-
There is no apparent chemical structural similarity between sulfasalazine and carbamazepine. We did not replicate our association with the rare HLA-C*14:02 allele in the HLA sequenced cohort. HLA-C*14:02 has an expected allele frequency of only 1.3% in Europeans, but is more common in Asia, with allele frequencies up to 8.2%.22

Notably, HLA-B*08:01 and HLA-A*31:01 have both been shown to increase the risk of carbamazepine hypersensitivity.12,23,24 Several studies have shown that HLA-A*31:01 increases the risk of carbamazepine hypersensitivity.28

There are three main proposed hypotheses through which HLA can interact with a drug to trigger an inappropriate immune response.26 According to the first model, the drug binds covalently to a peptide fragment that is presented on the MHC
molecule, while the second model proposes that the drug binds directly to the T-cell receptor or MHC molecule. In the third model, the drug binds noncovalently to the antigen-binding cleft of the HLA protein, which alters the repertoire of peptides it can bind. This last model has been demonstrated for abacavir, a drug strongly associated with acute hypersensitivity syndrome among carriers of the HLA-B*57:01 allele. It is not known whether any of the proposed mechanisms explain agranulocytosis induced by sulfasalazine. There is evidence that another sulphamamide, sulfamethoxazole, triggers an immune response by binding both covalently to proteins including MHC, and noncovalently to MHC/peptide complexes.

Although we found associations between sulfasalazine-induced agranulocytosis and classical HLA alleles, we cannot exclude influence from nearby noncoding regulatory variants. In the HLA-region, it is generally difficult to identify an actual causative variant due to a high degree of gene density and polymorphism, and extended linkage disequilibrium. Apart from our top hit close to HLA-B, a genome-wide significant association was observed in an intergenic region on chromosome 20 close to PAXI, a member of the paired box family of transcription factors. This infrequent variant is located in a repeat region and has no evidence of being functional in primary B cells from peripheral blood. After correcting for the top HLA hit, we found an association with a rare intronic variant in SERINC5, which encodes a protein that restricts human immunodeficiency virus (HIV) infectivity.

Based on imputed data from the Roadmap Epigenome project, the SERINC5 SNP is located in a weak enhancer region in primary B cells from peripheral blood characterized by histone 3 being monomethylated at lysine 4 (H3K4me1) and acetylated at lysine 27 (H3K27ac). Lastly, after sequentially adjusting for the other top hits, two infrequent SNPs on chromosome 1 were independently associated on a genome-wide level. However, these associations were likely to be spurious since both SNPs are in repeat regions and have no evidence of being regulatory variants in B-cells.

There are several limitations of this study. First, sulfasalazine-induced agranulocytosis is a rare event, and it is difficult to identify a large enough number of cases. For the present study, cases were collected over a period of 20 years, which resulted in a total of 76 cases. Second, HLA sequencing could not be performed in the discovery phase since we did not have access to DNA from the controls. In the replication phase all cases and controls were HLA-sequenced, thus making the results more reliable. Third, several polymorphisms in the HLA loci are known to confer genetic susceptibility to inflammatory arthritis or bowel disease. Due to the small sample size we did not stratify by disease; however, the distribution of diseases was similar among cases and matched controls. We therefore believe that there was no confounding by indication. Fourth, the most frequent disease in our cohort, rheumatoid arthritis, is strongly associated with a subgroup of HLA-DRB1*04 alleles, and to a lesser extent seronegative rheumatoid arthritis is associated with HLA-B*08. We did not discriminate between seropositive and seronegative rheumatoid arthritis, and could not test whether serotype influenced the detected association with HLA-B*08:01. Fifth, the number needed to screen (NNG) for HLA-B*08:01 and HLA-A*31:01 to avoid one case was high, 1,500. HLA-A*31:01 also increases
### Table 3  Characteristics of the patients in the replication cohort collected in Sweden

<table>
<thead>
<tr>
<th></th>
<th>Cases (n = 40)</th>
<th>Controls treated ≥ 90 days (n = 142)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, male [%]</td>
<td>18 [45.0]</td>
<td>64 [45.1]</td>
<td>1.00</td>
</tr>
<tr>
<td>Age, years at agranulocytosis [%]</td>
<td>at enrolment [%]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;25</td>
<td>4 [10.3]</td>
<td>8 [5.6]</td>
<td>0.21</td>
</tr>
<tr>
<td>25-29</td>
<td>0</td>
<td>7 [4.9]</td>
<td></td>
</tr>
<tr>
<td>30-34</td>
<td>1 [2.6]</td>
<td>14 [9.9]</td>
<td></td>
</tr>
<tr>
<td>35-39</td>
<td>2 [5.1]</td>
<td>11 [7.7]</td>
<td></td>
</tr>
<tr>
<td>40-44</td>
<td>2 [5.1]</td>
<td>14 [9.9]</td>
<td></td>
</tr>
<tr>
<td>45-49</td>
<td>6 [15.4]</td>
<td>16 [11.3]</td>
<td></td>
</tr>
<tr>
<td>50-54</td>
<td>4 [10.3]</td>
<td>20 [14.1]</td>
<td></td>
</tr>
<tr>
<td>55-59</td>
<td>2 [5.1]</td>
<td>14 [9.9]</td>
<td></td>
</tr>
<tr>
<td>60-64</td>
<td>4 [10.3]</td>
<td>11 [7.7]</td>
<td></td>
</tr>
<tr>
<td>65-69</td>
<td>5 [12.8]</td>
<td>7 [4.9]</td>
<td></td>
</tr>
<tr>
<td>70-74</td>
<td>6 [15.4]</td>
<td>8 [5.6]</td>
<td></td>
</tr>
<tr>
<td>&gt;74</td>
<td>3 [7.7]</td>
<td>12 [8.5]</td>
<td></td>
</tr>
<tr>
<td>Diagnosis(^a) [% of cases]</td>
<td>[% of controls]</td>
<td></td>
<td>1.25 × 10(^-4)</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>20 [50.0]</td>
<td>71 [50.0]</td>
<td>1.00</td>
</tr>
<tr>
<td>Crohn’s disease</td>
<td>1 [2.5]</td>
<td>3 [2.1]</td>
<td>1.00</td>
</tr>
<tr>
<td>Colitis, ulcerative or unspecified</td>
<td>10 [25.0]</td>
<td>5 [3.5]</td>
<td>1.00</td>
</tr>
<tr>
<td>Psoriasis arthritis</td>
<td>2 [5.0]</td>
<td>18 [12.7]</td>
<td>0.25</td>
</tr>
<tr>
<td>Ankylosing spondylitis</td>
<td>0</td>
<td>24 [16.9]</td>
<td>0.07</td>
</tr>
<tr>
<td>Arthritis unspecified</td>
<td>6 [15.0]</td>
<td>22 [15.5]</td>
<td>1.00</td>
</tr>
<tr>
<td>Plantar fasciitis</td>
<td>1 [2.5]</td>
<td>0</td>
<td>0.22</td>
</tr>
<tr>
<td>Total number of diagnosis</td>
<td>40</td>
<td>143</td>
<td></td>
</tr>
</tbody>
</table>

Controls were patients treated with sulfasalazine for at least 90 days without reacting with agranulocytosis.

\(^a\)One control had two diagnosis: ulcerative colitis and arthritis.

### Table 4  Replication of HLA types associated with sulfasalazine-induced agranulocytosis

<table>
<thead>
<tr>
<th>HLA allele</th>
<th>N</th>
<th>N Case</th>
<th>N Control</th>
<th>MAF Case</th>
<th>MAF Control</th>
<th>OR</th>
<th>95% CI</th>
<th>C Stat</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>C*14:02</td>
<td>182</td>
<td>40</td>
<td>142</td>
<td>0.00</td>
<td>0.00</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B*08:01</td>
<td>182</td>
<td>40</td>
<td>142</td>
<td>0.16</td>
<td>0.09</td>
<td>2.25</td>
<td>[1.02, 4.97]</td>
<td>0.574</td>
<td>0.0439</td>
</tr>
<tr>
<td>Full haplotype</td>
<td>182</td>
<td>40</td>
<td>142</td>
<td>0.16</td>
<td>0.06</td>
<td>3.79</td>
<td>[1.63, 8.80]</td>
<td>0.606</td>
<td>0.0019</td>
</tr>
<tr>
<td>C*02:02</td>
<td>182</td>
<td>40</td>
<td>142</td>
<td>0.09</td>
<td>0.09</td>
<td>0.99</td>
<td>[0.39, 2.50]</td>
<td>0.501</td>
<td>0.9876</td>
</tr>
<tr>
<td>B*07:02</td>
<td>182</td>
<td>40</td>
<td>142</td>
<td>0.19</td>
<td>0.12</td>
<td>1.67</td>
<td>[0.84, 3.28]</td>
<td>0.571</td>
<td>0.1406</td>
</tr>
<tr>
<td>A*31:01</td>
<td>182</td>
<td>40</td>
<td>142</td>
<td>0.09</td>
<td>0.02</td>
<td>4.81</td>
<td>[1.52, 15.26]</td>
<td>0.566</td>
<td>0.0077</td>
</tr>
</tbody>
</table>

The HLA types were selected by sequential adjustment for each significant association in the discovery cohort. Associated HLA types were compared between replication cases and controls treated 90 days or more using logistic regression. The effect was modeled per increase of one present HLA type. The full haplotype is HLA-DQB1*02:01-DRB1*03:01-B*08:01-C*07:01.

N = number, OR = odds ratio, 95% CI = 95% confidence interval, C Stat = C statistic, N/A = not available.
the risk of carbamazepine-induced hypersensitivity or skin reaction.28 Since HLA-A*31:01 is prevalent throughout the world, and the reactions are common, only 15–43 patients starting carbamazepine need to be screened for HLA-A*31:01 to avoid one case. There are other examples where preemptive genotyping is recommended despite a high NNG, although this is unusual.28,36 HLA-B*15:02 is a strong predictor of the serious skin reactions Stevens–Johnson’s syndrome and toxic epidermal necrolysis in East Asia, where this HLA type is prevalent.28 Before starting carbamazepine, patients originating from East Asia should be tested for HLA-B*15:02. Even in Asia, 10,000–24,665 patients need to be screened to avoid one case, due to the rarity of these serious skin reactions.

Strengths of this study are that it has a fairly large and homogeneous discovery cohort, and that the finding was replicated in a second cohort. In general, pharmacogenomic associations are robust and more likely to be translated into clinical practice than genetic associations with complex traits.37 There are several reasons. First, genetic variants associated with drug-related phenotypes usually have larger effect sizes than those associated with complex disease risk. Second, pharmacogenomic variants tend to have higher allele frequencies than variants associated with complex traits. This is possibly due to the fact that complex traits have been subject to negative selection over the course of human evolution, while pharmacogenomic variants only are deleterious in the presence of a certain drug, and drug treatment is a relatively new intervention. Third, pharmacogenomic variants are more likely to be actionable, since it is easier to select an alternative drug than to modify risk factors for complex disease, such as lifestyle. Fourth, it has been predicted that before long patients will have their pharmacogenome readily available in their medical records, thus removing the necessity to order the test.38

In conclusion, we have confirmed that sulfasalazine-induced agranulocytosis is associated with HLA-B*08:01 and HLA-A*31:01 in Europeans. The association was stronger for the HLA-DQB1*02:01-DRB1*03:01-B*08:01-C*07:01 haplotype than for HLA-B*08:01 alone, but the NNG did not change when using the full haplotype. Whether to use genetics in translational medicine to adapt drug treatment depends on the severity of the expected adverse drug reaction (ADR) and the existence of other treatment options.36 We believe that the drug sulfasalazine fulfills these criteria, although the high NNG makes testing challenging. The solution could be to include HLA-B*08:01 and HLA-A*31:01 into preemptive screening programs, where multiple tests are performed at a low cost. Known carriers of HLA-B*08:01 or HLA-A*31:01 starting sulfasalazine could then be placed under intensified monitoring or be offered an alternative drug. This individualization would be a further step towards personalized treatment of inflammatory disease.

METHODS

Ethical statement

The study was approved by the local Ethics Committees (2008/213 and 2010/231, Uppsala, Sweden; Dec. 22, 2014, Málaga, Spain; RTFI01, Barcelona, Spain; Charité-Universitätsmedizin Berlin, Germany; CPP Sud-Ouest et Outre-Mer I No. 1-09-24, Toulouse, France; and for recruitment of Swedish controls Stockholm Dnr 2007-644-31 and 2011/463-32). Research was carried out in accordance with the latest update of the Declaration of Helsinki. Written informed consent was obtained from all participants. The study protocol has been indexed in the European Network of Centres for Pharmacovigilance and Pharmacovigilance (ENCePP) register at www.encpp.eu ("The EuDAC Study"). Part of the replication cohort that had previously been collected for genetic studies of sulfasalazine-induced adverse reactions was approved by the Ethics Committee of the Medical Faculty, Uppsala University, Sweden (95-200 and 97-312).

Sample description

EuDAC consists of a network of investigators in Sweden, Spain, France, and Germany. The basis for case recruitment in Sweden (www.swedegenese) and France was through nationwide spontaneous ADR reports sent from healthcare professionals to the respective national drug regulatory authority. In Spain, cases were recruited both from spontaneous ADR reports and through active surveillance at 17 hospitals in Barcelona. German cases were recruited through active surveillance at 50 hospitals in Berlin, as described.28 Each included subject was at least 18 years of age and able to give informed consent. We defined cases as patients who had developed an absolute neutrophil count of 0.5 × 10^9/L or less (<500/µL) during drug therapy or within 7 days of stopping medication. Each case was required to exhibit complete recovery after cessation of the drug with an absolute neutrophil count >1.0 × 10^9/L (>1000/µL) or a compatible bone marrow aspirate or biopsy. Additional inclusion and exclusion criteria have been described previously.39 We collected clinical data (demographics, medical history, drug treatment history, laboratory data, and ancestry) through interviews using a standardized questionnaire, and by obtaining and reviewing medical records. At each center, cases were evaluated by at least one senior investigator, and a final adjudication of the complete dataset was performed by a specialist in hematology. Causality assessment was according to the WHO standard algorithm.40

Overall, 243 cases were collected for the study. Seven Swedish cases and two Spanish cases were excluded after adjudication due to either exposure to chemotherapy, negative rechallenge, unknown white blood cell count, diagnosis of chronic lymphatic leukemia, or missing clinical data. No German or French case was excluded. Out of the 234 cases of drug-induced agranulocytosis fulfilling all requirements, 36 were associated with sulfasalazine, which is the focus of this study. These cases originated from Sweden (n = 30), Germany (n = 3), and France (n = 3), while Spain had no case induced by sulfasalazine (Supplementary Table S1). Consenting population controls were available from Sweden, Spain, and Germany. In total, 5,170 controls were utilized: 4,891 unrelated individuals from the Swedish Twin Registry,41 183 Spanish, and 96 German individuals.28 Of the 183 Spanish controls, 147 had been recruited in a previous study of upper gastrointestinal bleeding,42 while the remaining 36 were healthy control subjects. Controls with a known history of agranulocytosis or neutropenia from any cause were excluded. Matching of controls from the Swedish Twin Registry was performed by linking with individual data from the Swedish National Patient Register, and the Swedish Prescribed Drug Register. Matched controls were those with a hospital diagnosis that is an indication for sulfasalazine treatment (data available from 1964). Diseases included were Crohn’s disease, ulcerative colitis, rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis, unspeciﬁed arthritis, and Sjögren’s syndrome. Some of these controls had also collected prescriptions of sulfasalazine (data available from July 2005). Patient characteristics were compared between discovery cases and controls using Fisher’s exact and chi² tests (Table 1).

For the replication, we used 40 cases of sulfasalazine-induced agranulocytosis and 142 controls who had been treated with sulfasalazine for a minimum of 3 months (90 days) without developing agranulocytosis (Table 3). Twenty-nine of the cases and all the controls had been collected previously in Sweden for genetic studies of sulfasalazine-induced adverse reactions.19,20 In addition, 11 new cases of sulfasalazine-induced
agranulocytosis collected in Sweden were included in the replication. All patients and controls had been treated with sulfasalazine for inflammatory joint disease, inflammatory bowel disease, or Sjögren’s syndrome. Patient characteristics were compared between replication cases and controls using Fisher’s exact, chi² and t-tests.

**Power calculation**

Given a genome-wide significance level of $5 \times 10^{-8}$, an ADR incidence of 0.14% (9), and using an additive genetic model, our sample size was powered to detect common genetic variants with effect sizes of clinical utility. For example, with 36 cases and 5,000 controls, we had 80% power to detect an OR of 4 for variants with a minor allele frequency (MAF) of 40%, and 80% power to detect OR slightly above 5 for variants with an MAF of 20% (Supplementary Figure S3).

**Genome-wide array data and analyses**

DNA was extracted from peripheral venous blood. Thirty-three cases recruited in Sweden and France, and 147 Spanish controls were genotyped with the Illumina HumanOmni 2.5M chip (Illumina, San Diego, CA; Figure 3). The remaining 36 Spanish controls had been genotyped with the Illumina HumanOmni1-Quad 1M chip. Three cases and 96 controls from Germany were genotyped with the Illumina HumanOmniExpress 700K array, as were 4,891 controls from the Swedish Twin Registry. Genotype calls were generated using the Genome Studio software from Illumina and the Genome Reference Consortium human assembly GRCh37.

GWAS quality control (QC) and data management was performed using PLINK v. 1.9. The resulting merged data included 596,010 SNPs. Imputation of genotypes was performed using PhaseIT and Impute v. 2.45 with the 1000 Genomes Project reference set (phase III, v. October 2014). The total number of SNPs after imputation was 9,380,034. All cases and controls were within the European cluster according to genetic principal component analysis (PCA) (Supplementary Figures S4, S5). Due to cases and controls being from different countries of Europe, sensitivity analyses were performed by reanalyzing each top finding in the largest group of cases and controls, which was from Sweden. See Supplementary Methods for additional details on QC, PCA, and imputation.

All genome-wide analyses were adjusted for sex and the first four principal components. SNP effects were modeled only as additive. The conventional genome-wide significance threshold $P < 5 \times 10^{-8}$ was used to correct for multiple testing. Results are presented as Manhattan plots. When genome-wide significant signals were found, analyses were performed sequentially by adjusting for each genome-wide significant signal until none were left. Logistic regression was used to estimate univariate and multiple models. Genome-wide statistical analyses was performed using PLINK v. 1.97,8 and individual SNPs were analyzed statistically, including LD calculation, using R 3.3.1 and the R packages rms and genetics (R Foundation for Statistical Computing, Vienna, Austria). The Q-Q plot is shown in Supplementary Figure S6.

**HLA allele imputation and analysis of the discovery cohort**

Imputation to first and second field resolution of 180 classical HLA alleles, amino acid residues, and individual SNPs was performed on the nonimputed merged and quality-controlled genome-wide data using the software SNP2HLA with a reference panel of 5,225 individuals. Logistic regression was used to test the association between HLA types and agranulocytosis. The HLA-wide significance level was set to 0.05. When HLA-wide significant signals were found, analyses were performed sequentially by adjusting for each HLA-wide significant signal until none remained. To avoid confounding by indication, top HLA signals were tested using a cohort of cases and controls matched for sulfasalazine treatment and/or inflammatory joint or bowel disease that was available from Sweden.

![Figure 3](http://www.cpt-journal.com) Study design. Design of the EuDAC study on sulfasalazine-induced agranulocytosis.

**HLA sequencing of the replication cohort**

For the replication cohort, HLA sequencing of HLA A, B, C, DP, DQ, and DR was performed on MiSeq (Illumina) as previously described by Cereb et al. The top HLA types were obtained by sequential adjustment for each significant HLA type in the discovery cohort. These five HLA types were tested for statistical associations, thus correction for five multiple tests was performed ($P < 0.01$).

**Analysis of predictive ability**

The predictive ability of logistic regression models for selected HLA analyses was expressed as the C-statistic (equivalent to the area under the receiver operating characteristic curve), which is a measure of the ability of the model to discriminate between cases and controls. A C-statistic of 0.5 indicates predictive ability similar to chance (a 50:50 likelihood of predicting the outcome), while 1 indicates perfect discrimination. The number needed to genotype (NNG) to avoid one case was estimated from the reciprocal of the predicted absolute risk reduction, where the absolute risk reduction is calculated as the prevalence (adverse drug reaction in the population) × the sensitivity of the diagnostic test.

**SUPPLEMENTARY MATERIAL** is linked to the online version of the article at http://www.cpt-journal.com

**ACKNOWLEDGMENTS**

EuDAC collaborators: Maryse Lapeyre-Mestre, Jean Louis Montastruc (Service de Pharmacologie Médicale et Clinique, Centre Hospitalier Universitaire, Faculté de Médecine de l’Université de Toulouse, UMR Inserm 1027, CIC 1436, Toulouse, France); Edeltraut Garbe (Leibniz-Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany); Lourdes Vendrell (Fundació Institut Català de Farmacologia, Hospital Universitari Vall d’Hebron, Barcelona, Spain); Ramon Puig Treserra (Fundació Institut Català de Farmacologia, Barcelona, Spain); Jose Luis Caro (Banc de Sang i Teixits, Barcelona, Spain); Maria Sainz Gil, Maria-Isabel Jimenez Semania, Inés Salado (Centro de Estudios sobre la Seguridad de los Medicamentos, Universidad de Valladolid, Spain); Paul McKeigue (University of Edinburgh Medical School, UK); Erik Eliasson (Clinical Pharmacology, Karolinska Institutet, Sweden); Håkan Melthius, Ulrica Ramqvist, Elisabet Stjemborg, Sofie Collin, Eva Prado Lopez, Agnes Wadelius, Martha Wadelius and Agnes Kataja Knight (Department of Medical Sciences, Clinical Pharmacology, Uppsala University, Uppsala, Sweden); Daniel Garwicz (Department of Medical Sciences, Clinical Chemistry, Uppsala University, Uppsala, Sweden); Bruno Stricker (Erasmus Medical Center, The Netherlands); Julia Ruiz-
AUTHOR CONTRIBUTIONS

The authors declare no conflicts of interest.

The authors declare no conflicts of interest.

22. Meuwissen, T.H., van Beusichem, M.L., Visscher, P.M. & Thompson, R.G. A fast method for calculating reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.