A Genome-Wide Screen for Large-Effect Alloimmunization Susceptibility Loci among Red Blood Cell Transfusion Recipients with Sickle Cell Disease

Neil A. Hancharda, b  Joann M. Mouldsc  John W. Belmonta, b  Alice Chend

aDepartment of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; bARS/USDA/Children’s Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; cScientific Support Services, LifeShare Blood Centers, Shreveport, LA, USA; dDepartment of Pathology, Texas Heart Institute, Baylor St. Luke’s Medical Center, Houston, TX, USA

Introduction

Transfusion of allogenic blood products is an essential part of modern medicine; however, this process is also an immunologic challenge to the recipient, who is exposed to multiple foreign antigens as a consequence. In the setting of red cell transfusions, this exposure sets the stage for the development of alloantibodies, the presence of which creates the considerable challenge of providing compatible blood products while avoiding the attendant problem of hemolytic transfusion reactions [1]. Given that some 5 million patients receive blood products every year in the USA, including those who receive multiple transfusions as a consequence of inherited or acquired hematological disorders, post-transfusion alloimmunization is a problem of major public health significance [2].

In clinical practice, it is well established that some individuals will form multiple antibodies with almost each transfusion; such ‘responders’ appear immunologically distinct from ‘non-responders’ for whom the development of antibodies is a rare and unusual occurrence. This phenotypic dichotomy of ‘responders’ and ‘non-responders’ is commonly observed in persons with sickle cell disease (SCD), who often require red cell transfusions both in the acute setting of sickle crises and for their long-term disease management. The rate of red cell alloimmunization in SCD patients is empirically quoted in the range of 30–50% [3, 4] – several orders of magnitude higher than the ~5% figure observed in the general population [2, 5]. Alloimmune responders with SCD are also at increased risk of hyperhemolysis [6], an uncommon but potentially devastating type of delayed transfusion reaction.

The reason some individuals become responders while others do not is not well-understood, but is likely to be multifac-
torial, involving donor, recipient, and external factors [5–7]. For instance, there is support for genetic ancestry contributing to alloimmunization in SCD patients – the genetic ancestry of individuals with SCD is predominantly African American, whereas blood donors of similar ethnicity are underrepresented in the national donor pool [7–10]. This potentiates the mismatches between donor and recipient red cell antigens that are necessary to stimulate an alloimmune response. Accordingly, expanded red cell antigen phenotyping and closer matching of common antigens between donor and recipient has decreased the incidence of alloimmunization in susceptible groups [11, 12], but a significant rate of alloimmunization remains, strongly implicating other mechanisms in the pathogenesis [13].

Among the potential additional factors is the well-known genetic variation that exists between individuals; for instance, possession of human leukocyte antigen (HLA) types HLA-B35 and HLA-Cw4 alleles have been linked to ‘responder’ status in persons with SCD [14, 15], and a single nucleotide polymorphism (SNP) in the erythrocyte antigen Ro52 (TRIM21) has been associated with the induction and efficiency of alloimmunization in both a transfusion-dependent and age-dependent fashion in patients with SCD [16]. These and other similar reports [17] validate a genetic underpinning to the alloimmune response, but they do so in a disparate, ad hoc fashion, and the primary results have not been replicated in other cohorts.

Advances in microarray technologies designed to interrogate SNPs across the genome have made it possible to perform genome-wide association studies (GWAS). The GWAS approach has emerged as a powerful way of agnostically identifying disease susceptibility loci and has facilitated the identification of biologically important loci in multiple diseases across a range of disciplines [18, 19] including disease severity in SCD [20, 21]. Many of the associations with the greatest magnitudes of effect have occurred in diseases with an immune or autoimmune basis [22–25] or in disorders with a strong gene-environment component [26–28]. These associations have fundamentally changed our understanding of the underlying disease processes [29] and in some cases have facilitated clinically useful genetic testing [30, 31]. Therefore, GWAS are an attractive and robust way of identifying biologically relevant genes and gene pathways contributing to the alloimmune responder phenotype in individuals with SCD; this in turn holds the ultimate promise of new therapeutic targets for those at higher risk. Further, because the GWAS approach includes the identification of both susceptibility and protection alleles, relevant loci could directly impact the ability of blood centers to preemptively identify both susceptible recipients for whom extended phenotyping is especially indicated as well as likely non-responders to whom random units can be provided, thereby increasing the number of donors who are available to supply phenotype-matched blood to SCD patients.

As a first foray into alloimmunization genomics, we undertook a pilot case-control GWAS of alloimmune responder status in a cohort of multiply transfused participants with SCD, with the express goal of identifying large-effect susceptibility loci. The genome-wide coverage achieved also allowed us to evaluate previously associated loci and to perform subgroup analyses of allo-antibody accumulation.

### Material and Methods

#### Subjects

The study protocol was approved by the Institutional Review Board of St Luke’s Episcopal Hospital and of Baylor College of Medicine. DNA samples used were those taken from patients with a diagnosis of SCD referred to LifeShare for blood group genotyping prior to receiving a red cell transfusion. DNA remaining after completion of the clinical test was retained and de-identified. As these clinical refuse samples were fully de-identified, only limited clinical data was available and the total (or lifetime) number of transfusion units for each individual is unknown; however, all patients had to have received a minimum of two transfusions in the past to be included. Each patient’s hemoglobinopathy status was confirmed by genotyping. Samples were obtained from within the LifeShare service area which includes Southeast Texas, the northern half of Louisiana, and Southern Arkansas. Patients were also screened for antibodies using routine serologic techniques. This included incubations at room temperature and 37 °C using Lo-Ion (Immucor, Norcross, GA, USA) and an AHG phase using polyspecific antiglobulin reagent. Patients were characterized as ‘responders’ (N = 198) if they developed clinically significant alloantibodies after two or more packed red cell transfusions, whereas blood donors of similar ethnicity are underrepresented.

#### Table 1. Cohort (n = 384) and pilot study demographics

<table>
<thead>
<tr>
<th>Genotyped individuals</th>
<th>104</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alloimmune category</td>
<td></td>
</tr>
<tr>
<td>Non-responder (males)</td>
<td>54 (23)</td>
</tr>
<tr>
<td>Responder (males)</td>
<td>50 (20)</td>
</tr>
<tr>
<td>Alloantibodies, median (min, max, IQR)</td>
<td>4 (1, 11, 2–5)</td>
</tr>
<tr>
<td>Sickle genotype</td>
<td></td>
</tr>
<tr>
<td>HbSS</td>
<td>104</td>
</tr>
</tbody>
</table>

IQR = Interquartile range.
fused RBCs [32]. ‘Non-responders’ (N = 186) were those with no alloantibodies after a similar exposure (table 1). We obtained funding to analyze 109 of these DNA samples in the first phase pilot GWAS – 55 ‘responders’ and 54 ‘non-responders’ homozygous for the sickle cell (HbS) mutation. The responder group also included 17 persons with a history of hyperhemolysis syndrome, defined as a post-transfusion drop in hemoglobin below pre-transfusion levels without evidence of bleeding.

**DNA Extraction and Genotyping**

DNA was extracted from leukocyte buffy coats or whole blood using either a manual method (PureGene®; Qiagen, Venlo, the Netherlands) or the QIAcube® (Qiagen). The quantity of double-stranded DNA was assessed in each sample using the PicoGreen (Invitrogen, Carlsbad, CA, USA) assay. All assays were performed according to manufacturer’s specifications. Five samples had a DNA concentration below the 50 μg/μl concentration recommended for optimal microarray genotyping and were omitted from further study.

Genotyping was performed according to manufacturer’s instructions using the Bead-Station system and HumanOmni1-Quad BeadChips (Illumina, San Diego, CA, USA), which was specifically designed to provide improved SNP coverage in people of West African heritage, as might be expected for participants with SCD. The GenomeStudio software (Illumina) was used to interpret normalized fluorescent intensities as genotypes. HumanOmni1-Quad BeadChips contain approximately 1,140,419 SNPs. SNPs with genotyping efficiency < 90% or out of Hardy-Weinberg equilibrium in controls (at the p < 10^{-5} level) were removed. The final analysis incorporated genotype data from 1,008,655 SNPs.

**Sample Quality Control**

Known siblings or other related samples were excluded from the initial sample selection. In addition, samples with more than 5% missing data (genotyping efficiency < 0.95) or with evidence of excessive inbreeding (inbreeding coefficient F > 0.1) were removed. Pairs of samples with excess allele sharing suggestive of close familial relationships (parent-offspring, siblings, or 1st degree relatives; PI_HAT > 0.1) were identified by identity-by-descent (IBD); in such cases one member of the pair was removed from analysis (fig. 1).

**Fig. 1.** Sample quality control (QC); Samples were removed from analysis if they i) failed to meet the minimum DNA concentration for the assay according to manufacturer’s recommendations (DNA conc); ii) were successfully genotyped at less than 95% of included SNPs (Genotyping); iii) had an inbreeding coefficient (F) suggestive of an excess of homozygotes; iv) had evidence of close-familial pairing (identity-by-descent (IBD) or were ancestry population outliers on principal components analysis (see fig. 2).

**Fig. 2.** Multidimensional scaling (MDS) cluster plot of study participants for principal components C1 and C2. A Clustering of SCD pilot cohort with HapMap African Americans from Oklahoma and in relation to major continental HapMap groups (YRI = Yoruba from Nigeria; CEU = Ceph from Utah). B MDS clustering of cases (black circles) and controls (red triangles) in the pilot cohort. PCA study outliers are circled.
Identification of Population Stratification

Prior to analysis, the remaining individuals were evaluated for shared ancestry using multidimensional scaling (MDS). The principle of MDS is to cluster individuals of similar ancestry by repeatedly adjusting individual genotypes by the amounts attributable to ancestry. The resulting dimensions can be plotted to visualize individuals whose inferred ancestry is outside that of the sample. Clustering was performed with a subset of SNPs pruned using a sliding window to exclude SNPs with strong non-random association (linkage disequilibrium; LD) with surrounding SNPs ($r^2 > 0.2$). Five outliers were identified using this approach (fig. 2). Exclusion of these outliers resulted in a final cohort of 48 ‘responder’ cases and 46 ‘non-responder’ controls.

Statistical Analyses

MDS clustering, IBD analysis, and F statistics (inbreeding) were carried out using PLINK [33], as were basic statistical analyses ($\chi^2$ and permutation). After primary analysis, label swapping adaptive permutation was performed for each SNP. In label swapping, the case/control label is swapped for each individual, and the association statistic is derived multiple times to create a new null distribution with which to compare with the observed data. The ‘adaptive’ mode essentially curtails the run length by removing SNPs that are highly unlikely to attain statistically significance at each permutation, until a user-defined maximum number of permutations (in this case 1,000,000) is reached for SNPs that continue to be statistically significant. Genome-wide inflation was estimated using the genABEL package [34] in R. Manhattan and MDS plots were generated using a web-based R source code (http://GettingGeneticsDone.com) and LocusZoom [35]. Shapeit2 [36] and Impute-2 [37] were used to impute genotype data from the HapMap [38] and the 1000 Genomes Project Consortium [39]. Imputation uses derived patterns of population-specific LD between SNPs to infer genotypes at sites that have not directly been genotyped. This can be a particularly useful approach in populations of African ancestry, where genotyped SNPs may not provide true ‘genome-wide’ coverage.

In subgroup analyses of the cumulative number of antibodies developed among responders, the distribution of alloantibodies was slightly skewed to the right. A power transformation ($x^{1/4}$) was performed to better fit a normal distribution. Unadjusted linear regression models were fit to the data to derive beta statistics and p values.

Quanto 1.2.4 [40] was used to estimate study power. The study sample size employed had 80% power to determine genetic risk ratios (GRRs) greater than 10 at genome-wide significance ($p < 0.5 \times 10^{-8}$) using SNPs with minor allele frequency (MAF) of 0.15 or greater under an additive disease model.

Results

No Large-Effect Responder Loci, but Suggestive Association in ARAP1/STARD10

For our primary analysis, each SNP with MAF > 0.10 was tested for association ($\chi^2$) under an allelic trend model using 48
responders as cases and 46 non-responders as controls. The genome-wide inflation factor in this analysis was 0.985, suggesting little residual inflation of the test statistic from genotyping error, cryptic relatedness of individuals, or population stratification. We did not observe any SNPs or clusters of SNPs at or below the accepted threshold for statistical significance in genome-wide association studies (p < 0.5 × 10⁻⁸) (fig. 3a).

Next, we sought to evaluate SNPs with p values that were below the ‘background’ association of most SNPs, but did not meet genome-wide significance. We focused on regions of the genome that had more than one SNP within 100 kb with p < 1 × 10⁻⁴; this threshold was deemed as ‘suggestive’ association. One region met this criteria (fig. 3a) – a cluster of three SNPs (lowest p = 1.2 × 10⁻⁵) on the long arm of chromosome 11 at the 3' end of ARAP1, extending 5' of STARD10 (fig. 3c). This modest association persisted after permutation testing (lowest p = 1.4 × 10⁻⁴). Manual inspection of the genotyping clusters and the relative local LD pattern in the HapMap YRI population from Yoruba, Nigeria [38], suggested that technical (genotyping) error was unlikely to be a significant contributing factor. Having observed moderately suggestive association with SNPs across this region, we then performed statistical imputation of ~17,983 variants within in a 5 megabase (MB) region centered upon the SNP with the best association in the original genotyping. We then filtered for common variants (MAF > 0.05) with imputation certainty greater than 95%. This resulted in a final test set of 3,795 SNPs. Using a χ² test for allelic trend under the same case-control model, five additional SNPs showed similar and, in some cases, stronger association than the genotyped SNPs at this candidate locus (fig. 3b,c), although none reached genome-wide significance.

Next, we focused on two regions that have been previously associated with the development of alloantibodies in other studies – TRIM21 on chromosome 11 and the major histocompatibility complex (MHC) on chromosome 6. At TRIM21, none of the genotyped SNPs had a p value less than 10⁻³ (fig. 4a). A similar pattern was observed across the 4 MB of

![Fig. 4. Local Manhattan plots for previously associated loci TRIM21 (A) and the MHC (B).](image)

In each figure the highlighted SNP (purple diamond) is the strongest association in the region. In A pairwise LD (r²) is derived from the Hapmap YRI population. In B the gene list is restricted for space considerations.
the MHC; most SNPs had p values greater than $1 \times 10^{-3}$ (fig. 4b), with the strongest SNP association ($p < 0.0004$) being found in c6orf10, upstream of the MHC class II region some distance from the major HLA types previously associated with the phenotype.

**Subgroup Analysis of Alloantibody Accumulation Reveals an Additional Suggestive Locus**

As an exploratory subgroup analysis, we incorporated a quantitative approach that focused on SNPs influencing the number of antibodies formed by each responder ($N = 48$) using unadjusted linear regression models. In this analysis there were no SNPs with evidence of association at or below genome-wide significance; however, using the same criteria for suggestive loci employed in our primary analysis, three regions had more than two SNPs with $p < 1 \times 10^{-4}$ within 100 kb. The most convincing of these included a cluster of eight SNPs in an 11 kb region on chromosome 4p involving the Toll-receptor 10 (TLR10) gene (fig. 5).

**Discussion**

We did not observe any SNP associations with responder status at or below genome-wide significance in our cohort. In the small sample size of this pilot study, genome-wide significance represents a conservative cut-off p value below which false-positive results are highly unlikely. Thus, we can conclude that in this SCD cohort, becoming an alloimmune responder is not influenced by a locus with very large effect size (GRR > 10). Like many other complex traits, the contribution of common genetic variation to responder status is more likely to be the result of loci with more moderate effects. This class of effects is difficult to confidently discern in a pilot GWAS such as this one, and is further compounded by the African ancestry of the cohort. African populations have greater genetic diversity and less LD between SNPs. Because genome-wide genotyping will rarely, if ever, genotype the true causal variant, GWAS studies are dependent upon LD for genotyped SNPs to act as ‘proxies’ for ungenotyped SNPs. The relative lack of LD in African and African-derived populations thus makes finding more moderate association signals using conventional microarray genotyping platforms more dif-
variants in suggest that there is little evidence to support a major role for or known covariates of alloimmunization (e.g. age or transfusion). The study was not designed to look for more subtle determinants larger than that employed in previous studies; however, our human genome reference sequence (www.ensembl.org) but also maps to several sites in the current version of the human genome reference sequence. The fact that rs660 is not represented on the HumanOmni1-quad chip, fact that rs660-flanking sequences do not map uniquely in the current genome assemblies raises the possibility that technical issues with the Goodpasture antigen and is known to play a role in the autoimmune response seen in Goodpasture syndrome; there is thus a potential, albeit speculative, biological role for STARD10 in the processing and presentation of red cell membrane antigens. Larger, well-powered replication studies, utilizing more comprehensive genotyping platforms that are better suited to the African American population, are necessary to validate these findings.

In keeping with recent murine studies, our findings suggest that there is little evidence to support a major role for variants in TRIM21 in responder status. Our sample size was larger than that employed in previous studies; however, our study was not designed to look for more subtle determinants or known covariates of alloimmunization (e.g. age or transfusion dependency) and thus we could not confirm the postulated role of the rs660 SNP in this cohort. It should be noted that rs660 is not represented on the HumanOmni1-quad chip, but also maps to several sites in the current version of the human genome reference sequence (www.ensembl.org). The preliminary results presented here.

This pilot GWAS of alloimmune responder status and alloantibody number did not reveal evidence of a large-effect locus in this SCD cohort, including previously identified loci TRIM21 and the MHC. However, suggestive association at the putative candidate loci ARAPI/STARD10 and TLR10 advocate for future genomics studies that will not only evaluate these preliminary results but also engender novel approaches to understanding differential alloantibody susceptibility in SCD. A genomics approach to the problem has recently been endorsed as a viable framework in which to study various aspects of the SCD responder phenotype. The challenge demonstrated by this study is that success in such a venture will require larger cohorts built around a broad expertise and strong collaboration.

Acknowledgements

This work was funded in part by a grant from Gulf Coast Regional Blood Center to AC. NH is supported by a Clinical Scientist Development Award from the Doris Duke Charitable Foundation. We would also like to thank Kristine Bucasas for help with implementing the statistical imputation.

Disclosure Statement

JMM serves is currently an employee of Grifols Inc. and previously served as the principal investigator for grants and research agreements between LifeShare and Novartis Diagnostics or BioArray/Immucor. No other authors have conflicts to declare.
References


12 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Mallard J, Sklar P, de Bakker PJ, Daly MJ, Skam PC: Plink: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575.


A Genome-Wide Screen for Large-Effect Alloimmunization Susceptibility Loci among Red Blood Cell Transfusion Recipients with Sickle Cell Disease


A Genome-Wide Screen for Large-Effect Alloimmunization Susceptibility Loci among Red Blood Cell Transfusion Recipients with Sickle Cell Disease

Trans Med Hemother 2014;41:453–461