

UCSF

UC San Francisco Previously Published Works

Title

Bridging ImmunoGenomic Data Analysis Workflow Gaps (BIGDAWG): An integrated case-control analysis pipeline.

Permalink

<https://escholarship.org/uc/item/6j12j2vt>

Journal

Human immunology, 77(3)

ISSN

0198-8859

Authors

Pappas, Derek J
Marin, Wesley
Hollenbach, Jill A
[et al.](#)

Publication Date

2016-03-01

DOI

10.1016/j.humimm.2015.12.006

Peer reviewed



HHS Public Access

Author manuscript

Hum Immunol. Author manuscript; available in PMC 2017 March 01.

Published in final edited form as:

Hum Immunol. 2016 March ; 77(3): 283–287. doi:10.1016/j.humimm.2015.12.006.

Bridging Immunogenomic Data Analysis Workflow Gaps (bigdawg): An Integrated Case-Control Analysis Pipeline

Derek J. Pappas¹, Wesley Marin^{2,3}, Jill A. Hollenbach^{#2}, and Steven J. Mack^{#1}

¹Center for Genetics, Children's Hospital & Research Center Oakland, Oakland, CA

²Department of Neurology, University of California, San Francisco, CA

³Molecular and Cell Biology, University of California, Berkeley, CA

These authors contributed equally to this work.

Abstract

Bridging ImmunoGenomic Data-Analysis Workflow Gaps (BIGDAWG) is an integrated data-analysis pipeline designed for the standardized analysis of highly-polymorphic genetic data, specifically for the HLA and KIR genetic systems. Most modern genetic analysis programs are designed for the analysis of single nucleotide polymorphisms, but the highly polymorphic nature of HLA and KIR data require specialized methods of data analysis. BIGDAWG performs case-control data analyses of highly polymorphic genotype data characteristic of the HLA and KIR loci. BIGDAWG performs tests for Hardy-Weinberg equilibrium, calculates allele frequencies and bins low-frequency alleles for $k \times 2$ and 2×2 chi-squared tests, and calculates odds ratios, confidence intervals and p-values for each allele. When multi-locus genotype data are available, BIGDAWG estimates user-specified haplotypes and performs the same binning and statistical calculations for each haplotype. For the HLA loci, BIGDAWG performs the same analyses at the individual amino-acid level. Finally, BIGDAWG generates figures and tables for each of these comparisons. BIGDAWG obviates the error-prone reformatting needed to traffic data between multiple programs, and streamlines and standardizes the data-analysis process for case-control studies of highly polymorphic data. BIGDAWG has been implemented as the *bigdawg* R package and as a free web application at bigdawg.immunogenomics.org.

Keywords

BIGDAWG; HLA KIR data analysis; R package; web app; Hardy-Weinberg testing; case-control analysis; amino-acid analysis; haplotype analysis

Corresponding Author, Derek Pappas, Ph.D., Children's Hospital Oakland Research Institute, 5700 Martin Luther King Jr Way, Oakland, California 94609.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

The extensive polymorphism, linkage disequilibrium and genotyping ambiguity commonly associated with the HLA and KIR loci (described here collectively as *immunogenomic* loci) pose challenges for the consistent analyses of these data [1]. Modern genetic analysis programs are designed for use with bi-allelic single nucleotide polymorphisms (SNPs) or SNP haplotypes generated in genome-wide association studies (GWAS), but cannot be applied to highly polymorphic immunogenomic data. New tools are needed to leverage modern computational resources for the analysis of immunogenomic data, and to integrate the analysis of immunogenomic loci with genomic SNP/GWAS data. The few Ad-hoc tools designed to handle immunogenomic data, such as PyPop [2] and Arelquin [3] are limited by operating systems, outdated with spurious maintenance cycles, and often times require cumbersome data formatting.

A typical immunogenomic data analysis workflow involves the trafficking of data between several programs; this usually involves reformatting of these data for each program, a process that is time intensive, error prone and limits reproducibility. Quite often, this data-trafficking involves the use of Microsoft Excel, which is particularly poor choice for immunogenomic data-management [1]. In addition, the management of data in a typical workflow is often idiosyncratic to the analyst, which further limits reproducibility across studies. The automated manipulation of immunogenomic data in a single analysis workflow will reduce errors and allow true analytical reproducibility.

We have developed Bridging ImmunoGenomic Data-Analysis Workflow Gaps (BIGDAWG), an automated software pipeline that performs a suite of common case-control analyses of multi-locus highly polymorphic genetic data [4-6]. Unlike SNP/GWAS case-control analysis tools, BIGDAWG is tailored for use with immunogenomic data. In addition, BIGDAWG can be applied to any highly polymorphic genetic data, including SNPs and SNP haplotypes. BIGDAWG is implemented as an R package (named, *bigdawk*) and as a web application running at bigdawk.immunogenomics.org.

2. Methods

2.1. Implementation

BIGDAWG has been developed in the framework of the R statistical environment (<http://www.r-project.org>). The *bigdawk* R package provides documentation of all BIGDAWG functions, and includes a vignette detailing package use along with a sample dataset. The *bigdawk* vignette is included here as Supplementary Material. BIGDAWG's functionality depends on the *epicalc* [7] and *haplo.stats* [8] R packages, along with the R base package *parallel*. The R *XML* package [9] is required for updating the protein alignment object to adhere to the most current IMGT/HLA Database [10] (version 3.20.0 released 2015-04-17 as of this writing). The *bigdawk* R package (version 1.1) is covered under the GNU general public license version 3 or higher and has been made available through the Comprehensive R Archive Network (CRAN) repository.

The BIGDAWG web application (BWA) is a *shiny* [11] implementation of the *bigdawg* R package. As such, BWA requires only a modern web browser and Internet access to function, and does not require the R environment to be installed on a user's system. BWA input data and analytical parameters (described in Section 2.5) are specified in the user's web-browser, and the results files can be downloaded from the browser.

2.2. Functions

BIGDAWG accepts unambiguous genotype data for case-control groups as input, and calculates allele frequencies for chi-square (χ^2) testing, along with odds ratios, confidence intervals and p-values for each allele (for a processing flowchart see Supplementary Material Figure 1). BIGDAWG combines rare alleles into a common class ("binning"; see section 2.3) which are included for testing, performs overall locus-level ($k \times 2$) tests of significance, followed by a series of allele-level (2×2) tests of significance for each locus. In addition, the control group is tested for deviations from expected Hardy-Weinberg equilibrium proportions (HWEP) at the allele level. When multi-locus genotype data are available, BIGDAWG estimates user-specified haplotypes and performs the same binning and statistical calculations for each haplotype [$k \times 2$ tests at the multi-locus level (e.g. HLA-A~HLA-B or HLA-DRB1~HLA-DQA1~HLA-DQB1) followed by 2×2 tests at the haplotype level]. For HLA data, BIGDAWG integrates protein sequence alignments from the IMGT/HLA database to run case-control association tests on individual amino-acid positions within exon 2 and exon 3 (class I) or exon 2 (class II) ($k \times 2$ tests for each polymorphic amino-acid position, followed by 2×2 tests for each amino-acid residue). For these amino acid analyses, HLA allele names must conform to the colon-delimited HLA allele name nomenclature as defined by the WHO Nomenclature Committee for Factors of the HLA System in April 2010 [12].

2.3. Statistics

All HWEP and phenotype association (haplotype, locus and amino acid) analyses are currently based on a traditional χ^2 test. For HWEP deviation testing, BIGDAWG combines rare genotypes into a single common class (binning) for analysis and performs a goodness-of-fit test. The degrees of freedom (dof) are calculated as $dof = g - (a - 1)$, where g is the number of unique non-binned genotypes and a is the number of unique non-binned alleles.

For testing phenotype associations, BIGDAWG runs a test-of-independence, automatically tabulating the $k \times 2$ contingency tables, where k is the number of unique haplotypes, alleles or amino acids. For either testing scenario, rare cells (with expected counts less than five) are combined into a common class (binned) prior to computing the χ^2 statistic, except in cases of the test-of-independence where all cells of a given $k \times 2$ contingency table are ≥ 1 and fewer than 20% of the cells have expected counts less than five. BIGDAWG's haplotype estimation function requires the R *haplo.stats* package, whereas calculation of the individual haplotype/allele/residue confidence intervals, odds ratios, and p-values requires the R *epicalc* package.

2.4. Input and Output Data Structures

BIGDAWG input files are tab delimited text files with columns for subject IDs, phenotype association analysis (labeled 1 or 0), and column pairs of unambiguous, unphased alleles for each locus. Allele names can be of any format (e.g., 1, 2, 3, a, A, b, B, s, S, t, T, p, P, q, Q, etc. can be supplied as allele names). For HLA data, allele names (with or without a locus prefix) can include from a single field up to the full length name for a given allele (e.g., “01”, “01:01”. “01:01:01” and “01:01:01:01” are all recognized as valid alleles). BIGDAWG treats the absence of a locus (e.g., resulting from structural variation) as an allele of that locus, and recognizes “00:00” as a convention for identifying absent loci. This can be especially relevant for HLA loci such as *HLA-DRB3*, *HLA-DBR4*, *HLA-DRB5* as well as members of the *KIR* gene family, where locus absence may be informative and associated with the pertinent phenotype.

After input data have been read, BIGDAWG provides a short summary of the relevant architecture of the supplied data (e.g., the number of unique alleles and the number of instances of missing data at each locus), and runs a set of data consistency checks to ensure the most compatible data set for analysis (e.g., identifying large-scale discrepancies between the number of HLA allele-name fields in case and control groups). An example of this summary is shown in Figure 1. The *bigdawg* vignette, included in the Supplementary Material, provides more detailed description of input file requirements.

Summaries of each analysis are displayed on the R console/terminal window (Figure 2), or web-browser pane. However, all analytical results are recorded as tab delimited text files, which include more detailed descriptions of each analysis. In addition, each BIGDAWG analysis generates a “run parameters” file identifying the options used in that run, allowing each analysis to be reproduced. Descriptions of each BIGDAWG result file are included in the Supplementary Material as part of the *bigdawg* vignette.

2.5. Parameters

BIGDAWG offers considerable flexibility in the selection of parameters for running an analysis. Users can specify individual levels of analysis (for Hardy-Weinberg (“HWE”) or for case-control at the haplotype (“H”), locus (“L”) or amino-acid (“A”) levels) or combinations of these tests (data permitting) using the Run.Tests parameter in the *bigdawg* R package, or using checkboxes for BWA. Users can specify a threshold for the per-subject missing data allowance (the Missing parameter); missing data can dramatically impact haplotype estimation performance as the frequency of missing data increases. For the haplotype analysis, user can identify which loci or combinations of loci using the Loci.Set parameter, and can specify analysis of all pairwise combinations of available loci using the All.Pairwise parameter. Functions specific to HLA allow finer formatting of the data, including trimming to a desired level of resolution using the Trim parameter. Finally, BIGDAWG takes advantage of R base functions for parallel computing to increase processing performance during the amino acid analysis (operating system dependent). For more information on setting parameters and their defaults, please refer to the *bigdawg* vignette included in the Supplementary Material.

2.6. Built-in Example Dataset

A synthetic example dataset (“HLA_data”) consisting HLA genotype data for 2000 subjects (998 cases, 1002 controls) has been included in the *bigdawg* R package. This multi-locus dataset has been designed to illustrate specific BIGDAWG functions and features, and does not represent actual HLA data for a real study. In particular, this dataset includes loci that do and do not require binning for locus-level analyses, that do and do not display significant phenotype associations, and that do and do not display deviations from HWEP for the control group. In addition, the multi-locus data can be used to perform different sets of haplotype analyses, and amino-acid level analyses can be performed. Finally, this dataset includes examples of missing allele data and absent loci.

3. Results

3.1. Running *bigdawg*

In this section, we demonstrate running *bigdawg* on the built-in example data set (described in section 2.6). The example set can be accessed by setting the ‘Data’ parameter to the value ‘HLA_data’ (case sensitive). The first two lines of the following code snippet specify all possible parameters that a user can change. The subsequent two lines will load *bigdawg* from the R library (step 1) and run the full analysis with all defaults using the built-in dataset (step 2).

```
># All possible user parameters:
># bigdawg(Data, HLA=TRUE, Run.Tests, Loci.Set, All.Pairwise=FALSE,
Trim=FALSE, Res=2,
EVS.rm=FALSE, Missing=0, Cores.Lim=1L, Results.Dir, Output=TRUE)
>library(bigdawg) # step 1 load bigdawg
>bigdawg('HLA_data') # step 2 run bigdawg, all defaults, sample data set
```

For any dataset, BIGDAWG will initially provide summary statistics of the dataset (Figure 1) including the number and name of loci available, the number of alleles per locus, and the number of cases and controls. Moreover, with the default setting of ‘HLA=TRUE’ (the data is HLA genotyping data), BIGDAWG will also determine the maximum allele-name length of the alleles for cases and controls, and will alert the user if an allele-name length imbalance exists between them. Following the summary statistics, BIGDAWG will test the controls for HWEP at each available locus.

BIGDAWG's default setting is to test all available loci. Therefore all association tests will run on all loci, including the haplotype, locus, and amino acid tests. The console/terminal output (Figure 2) summarizes the results of the individual tests. For a more detailed description of the association for each haplotype, allele, and residues, BIGDAWG writes text-formatted output files that can be reviewed later.

The following lines demonstrate the different options to fine tune an analysis.

```
# Run the haplotype analysis on all loci including all pairwise comparisons
>bigdawg(Data="HLA_data", Run.Tests="H", All.Pairwise=T)
# Run the Hardy-Weinberg and Locus analysis with non-HLA data, ignoring
missing data
>bigdawg(Data="HLA_data", HLA=F, Run.Tests=c("HWE","L"), Missing="ignore")
# Run the amino acid analysis, trimming HLA data to 2-Field resolution
>bigdawg(Data="HLA_data", Run.Tests="A", Trim=T, Res=2)
```

BIGDAWG's HWEP, locus, and amino acid analyses evaluate each locus independently; multiple locus subsets are only evaluated as part of the analysis of haplotypes. To analyze multiple haplotypes, locus subsets should be specified only for the haplotype analysis, in order to avoid performing redundant locus and amino acid analyses.

```
# Run the haplotype test on a list of locus subsets
>bigdawg(Data="HLA_data", Run.Test="H", Loci.Set=list(c("A","DRB1",
"DQB1"),c("DRB1","DQB1","DPA1")))
```

4. Discussion

BIGDAWG is a standardized pipeline for the case-control analysis of immunogenomic data. Available as the *bigdawg* R package, and as BWA, the [BIGDAWG web application](#), BIGDAWG has been designed for the analysis of highly-polymorphic HLA data, but can be applied to any genotype data, including genotype data derived from disparate genetic systems (e.g., HLA, KIR and SNPs) or from a variety of sources. BIGDAWG performs case-control analyses at the haplotype, locus and amino-acid levels, and also performs tests for deviation from HWEP in control groups. Most importantly, BIGDAWG automates the analytic accommodations required for the analysis of highly polymorphic data, a capacity that is not available in SNP-focused data-analysis software.

BIGDAWG's functions are modular; each analysis is self-contained, and can be run separately, or sequentially, from a single command. Similarly, each analytic result is reported separately, and can serve as the input for other analyses. BIGDAWG's modular nature allows functionality to be added, edited, or removed in future releases as needed. When used in the R environment, *bigdawg* can be called to analyze hundreds of datasets using simple loops or, more efficiently, through the use of the *apply()* family of R base functions. Finally, BIGDAWG documents the settings used for each analysis, allowing any analysis to be reproduced.

4.1 Ongoing BIGDAWG Development

The currently available version of BIGDAWG is version 1.0. However, BIGDAWG development is active and ongoing, with the goal of adding support for new features and data-types in future releases.

For example, BIGDAWG 1.0 employs a χ^2 statistic for HWE testing. The utility of the χ^2 test for highly polymorphic datasets is limited due to the propensity for “sparse cells” in the table of all possible genotypes. Future releases of BIGDAWG will include the more accurate Monte-Carlo-based exact-test approximation developed by Guo and Thompson (1992) for HWE testing [13].

In addition, functions calculating the linkage disequilibrium (LD) measures, D' [14] and W_n [15], as well as the conditional asymmetric LD (ALD) measure [16], will be included in future BIGDAWG releases to complement extant haplotype analyses. In particular, the application of ALD will foster the more granular dissection of disease association in cases where allele polymorphism is asymmetric between loci in haplotypes.

Finally, future BIGDAWG versions will support generalized linear models, the validation of KIR allele names derived from the IPD-KIR Database [17], and amino-acid level analysis of KIR polymorphisms.

4.2. Conclusions

The goal in developing BIGDAWG was to eliminate the tedious, time-consuming and error-prone reformatting of datasets required for the use of multiple data-analysis programs. BIGDAWG starts with a table of genotypes for case and control groups, requires only that a user decide which analyses should be done, and generates a series of result tables that can be incorporated into a presentation or manuscript with minimal effort. Use of BIGDAWG will represent a significant increase in productivity for any research effort, freeing investigators to focus on discovery rather than data-management.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The work described here was performed with the support of National Institutes of Health (NIH) grants R01GM19030 (JAH, SJM, DP) awarded by the National Institute of General Medical Sciences (NIGMS) and U01AI067068 (JAH and SJM) awarded by the National Institute of Allergy and Infectious Diseases (NIAID). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, NIGMS, NIAID or United States Government. We thank Hannah Ollila, Richard M. Single, and Julia Udell for beta-testing early versions of BIGDAWG. We also thank Paul J. Norman, Patricia Francis-Lyon, Anthony G. Pena and Robert M. Horton for helpful discussions.

Abbreviations

ALD	Asymmetric Linkage Disequilibrium
BIGDAWG	Bridging ImmunoGenomic Data-Analysis Workflow Gaps
BWA	BIGDAWG web application
DOF	Degrees Of Freedom
GWAS	Genome Wide Association Study

HLA	Human Leukocyte Antigen
HWEP	Hardy-Weinberg Equilibrium Proportions
IMGT	ImMunoGeneTics
IPD	Immuno Polymorphism Database
KIR	Killer-cell Immunoglobulin-like Receptor
LD	Linkage Disequilibrium
SNP	Single Nucleotide Polymorphism

References

- Hollenbach JA, Mack SJ, Gourraud PA, Single RM, Maiers M, Middleton D, Thomson G, Marsh SG, Varney MD. A community standard for immunogenomic data reporting and analysis: proposal for a Strengthening the REporting of Immunogenomic Studies statement. *Tissue Antigens*. 2011; 78(5):333. [PubMed: 21988720]
- Lancaster AK, Single RM, Solberg OD, Nelson MP, Thomson G. PyPop update--a software pipeline for large-scale multilocus population genomics. *Tissue Antigens*. 2007; 69(Suppl 1):192. [PubMed: 17445199]
- Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010; 10(3):564. [PubMed: 21565059]
- Mack SJ, Gourraud PA, Single RM, Thomson G, Hollenbach JA. Analytical methods for immunogenetic population data. *Methods Mol Biol*. 2012; 882:215. [PubMed: 22665237]
- Gourraud PA, Hollenbach JA, Barnetche T, Single RM, Mack SJ. Standard methods for the management of immunogenetic data. *Methods Mol Biol*. 2012; 882:197. [PubMed: 22665236]
- Hollenbach JA, Mack SJ, Thomson G, Gourraud PA. Analytical methods for disease association studies with immunogenetic data. *Methods Mol Biol*. 2012; 882:245. [PubMed: 22665238]
- Chongsuvivatwong V. epicalc: Epidemiological calculator. 2012
- Sinnwell J, Schaid J. haplo.stats: Statistical Analysis of Haplotypes with Traits and Covariates when Linkage Phase is Ambiguous. 2013
- Lang DT. XML: Tools for parsing and generating XML within R and S-Plus. 2013
- Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res*. 2015; 43(Database issue):D423. [PubMed: 25414341]
- Chang W, Cheng J, Allaire J, Xie Y, McPherso J. shiny: Web Application Framework for R. 2015
- Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, Fernandez-Vina M, Geraghty DE, Holdsworth R, Hurlley CK, Lau M, Lee KW, Mach B, Maiers M, Mayr WR, Muller CR, Parham P, Petersdorf EW, Sasazuki T, Strominger JL, Svejgaard A, Terasaki PI, Tiercy JM, Trowsdale J. Nomenclature for factors of the HLA system, 2010. *Tissue Antigens*. 2010; 75(4): 291. [PubMed: 20356336]
- Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*. 1992; 48(2):361. [PubMed: 1637966]
- Hedrick PW. Gametic disequilibrium measures: proceed with caution. *Genetics*. 1987; 117(2):331. [PubMed: 3666445]
- Cramer, H. *Mathematical Models of Statistics*. Princeton University Press; Princeton, NJ: 1946.
- Thomson G, Single RM. Conditional asymmetric linkage disequilibrium (ALD): extending the biallelic r^2 measure. *Genetics*. 2014; 198(1):321. [PubMed: 25023400]
- Robinson J, Halliwell JA, McWilliam H, Lopez R, Marsh SG. IPD--the Immuno Polymorphism Database. *Nucleic Acids Res*. 2013; 41(Database issue):D1234. [PubMed: 23180793]

>>>> CASE - CONTROL SUMMARY STATISTICS

```

Sample Summary
Sample Size (n): 2000
...Number of Controls/Cases: 1002/998
Allele Count (2n): 4000
Total loci in file: 6
Unique loci: A, DRB1, DQB1, DRB3, DRB4, DRB5
Unique alleles per locus: 22, 20, 20, 8, 6, 4
...Unique in Controls/Cases: 22/22, 17/18, 17/19, 7/8, 6/4, 3/4
Missing alleles per locus: 0, 30, 0, 30, 30, 30
...Missing in Controls/Cases: 0/0, 22/8, 0/0, 22/8, 22/8, 22/8

Observed Allele Resolution
Max Resolution Controls: 4-Field
Max Resolution Cases: 4-Field
Defined Resolution: Not Defined

```

```

>>>> STARTING HARDY-WEINBERG ANALYSIS...
HWE performed at set resolution on controls.

```

```

> HARDY-WEINBERG ANALYSIS COMPLETED
Locus X.square df p.value sig
A 6.8714 54 1 NS
DRB1 0.4065 38 1 NS
DQB1 191.9252 65 1.8319e-14 *
DRB3 0.4081 10 1 NS
DRB4 0.0206 3 0.99922 NS
DRB5 0.0121 2 0.99399 NS

```

Figure 1. Summary Statistics and Hardy-Weinberg Equilibrium Analysis

Sig (significance) column. * indicates a significant p-value. These p-values have not been corrected for multiple comparisons.

```

>>>> STARTING HAPLOTYPE ANALYSIS...
-----
Estimating Haplotypes ...

> HAPLOTYPE ANALYSIS COMPLETED: All.Loci
X.square df    p.value sig
665.7496 84 < 2.22e-16  *

>>>> STARTING LOCUS LEVEL ANALYSIS...
-----
> LOCUS LEVEL ANALYSIS COMPLETED
Locus X.square df    p.value sig
  A 82.4359 21 3.1495e-09  *
DRB1 125.01 14 < 2.22e-16  *
...other loci omitted for brevity

>>>> STARTING AMINO ACID LEVEL ANALYSIS...
-----
> AMINO ACID ANALYSIS COMPLETED
Significant Amino Acid Position(s):
Locus      Position X.square df    p.value sig
  A Position.43  5.7117  1  0.016852  *
  A Position.62  10.04  4  0.039759  *
  A Position.74  5.5698  1  0.018273  *
  A Position.76  6.2869  2  0.043135  *
  A Position.97 17.1988  2 0.00018422  *
  A Position.107 5.5698  1  0.018273  *
  A Position.149 8.9534  1  0.0027695  *
  A Position.156 12.0245  3  0.0072995  *
DRB1 Position.11  9.75  4  0.044857  *
DRB1 Position.33  5.1192  1  0.023662  *
DRB1 Position.71 32.7722  3 3.5975e-07  *
DRB1 Position.86 26.9263  1 2.1136e-07  *
...other loci omitted for brevity

```

Haplotype

Locus

Amino Acid

Figure 2. Summarized Association Testing Results

Sig (significance) column. * indicates a significant p-value. These p-values have not been corrected for multiple comparisons. The Amino Acid Analysis results have been shorted for publication.