

IDAWG 16th IHIW PROJECT: Immunogenomic Data Management Methods

Jill A. Hollenbach and Steven J. Mack

Children's Hospital Oakland Research Institute, Oakland, CA, USA

The Immunogenomics Data Analysis Working Group (IDAWG) is an international collaboration of histocompatibility and immunogenetics investigators who share the goal of making immunogenomic data (HLA, KIR, etc.) more useful across studies and between different groups. In pursuit of this goal, the IDAWG has launched an International Histocompatibility and Immunogenetics Workshop (IHIW) project. The goal of this project is to develop data-management tools and documentation standards that are tailored to work with the HLA and KIR data-management practices in use by the immunogenetics community.

The project will proceed in two phases, with the first phase now underway and continuing through the 16th Workshop meeting. For more information, visit <http://www.immunogenomics.org/workshop.html>

PHASE ONE

Survey of current HLA and KIR data-management practices available at <http://www.surveymonkey.com/s/IDAWG>

We are surveying the immunogenetics and immunogenomics community, to determine current practices in:

- HLA and KIR data-management and transmission

* 3. 3/3 Which HLA genes are studied in your laboratory? (check all that apply)

HLA-A
 HLA-C
 HLA-B
 HLA-DRB1
 HLA-DRB3/4/5
 HLA-DQA1
 HLA-DQB1
 HLA-DPA1
 HLA-DPB1
 Other HLA gene (please describe below)

* 4. 4/4 Using the information listed below, which version(s) of the HLA nomenclature do you use for each locus? (Check all that apply)

Version of HLA Nomenclature available:
 VERSION 1 (IMGT/HLA database release 1.0-1.16)
 VERSION 2 (IMGT/HLA database release 2.0-2.28)
 VERSION 3 (IMGT/HLA database release 3.0 or higher)

* 13. 13/7 Does your lab generate its data on its own, receive data from another lab, or both? (select one)

My laboratory generates data
 My laboratory receives data
 My laboratory generates and receives data

* 14. 14/15 What methods/programs are used to store and manage data in your lab? You will be asked to describe in the next question. (select all that apply)

Commercial Database System (e.g. MySQL, MS Access, FileMaker, FoxPro)
 Custom Database System (e.g. SCORE)
 Microsoft Excel
 Other Spreadsheet Program
 Word Processing Program (e.g. MS Word)
 Other electronic method
 Non-electronic method (e.g. paper cards)

- HLA ambiguity management and resolution

* 28. 28/29 Are the results of your Hardy-Weinberg analyses part of a data quality check? (check one)

No
 Yes

29. 29/30 What software does your laboratory use to carry out Hardy-Weinberg analyses? (select all that apply)

Arlequin
 GenePop
 R package(s)
 PyPop
 Other software

- Primary data-analysis

* 28. 28/29 Are the results of your Hardy-Weinberg analyses part of a data quality check? (check one)

No
 Yes

29. 29/30 What software does your laboratory use to carry out Hardy-Weinberg analyses? (select all that apply)

Arlequin
 GenePop
 R package(s)
 PyPop
 Other software

The survey is available at <http://www.surveymonkey.com/s/IDAWG>, and should take approximately 15 minutes to complete. The survey can be submitted anonymously, and although survey participants are invited to take part in the larger IDAWG Workshop project, information specific to individual laboratories will be kept confidential; individual laboratories will not be identified without providing consent. When survey participants have indicated their interest, the survey will be followed-up with synthetic datasets for the purpose of demonstrating HLA ambiguity resolution and primary data-analysis practices.

* 25. 25/24 Thank you for agreeing to demonstrate your laboratory's ambiguity reduction approach on test datasets. Please provide a contact name and email address. test datasets following the survey."

Name:
 Email Address:

The IDAWG 16th IHIW Survey of Immunogenetic Data Management and Analysis Methods has been developed and made available with the generous support of the American Society for Histocompatibility and Immunogenetics (ASHI), and we thank ASHI for their support of this project.

Please direct any questions or comments about the survey to

survey@immunogenomics.org

Your contact information will be used only for 16th Workshop-related news, and will not be shared.

Thank you for participating in the Immunogenomic Data-Analysis Working Group (IDAWG) survey of immunogenetic data-management and analysis practices. Information provided will be kept confidential.

PHASE TWO

Impact of data management practices on downstream analysis

The second part of this project will begin after the first few surveys are completed, and will use the information provided by the community to determine the effects of the various practices in use on common applications for these data, including:

- Registry Searches,
- Disease-Association Studies, and
- Population Studies.

By the 16th Workshop meeting, we hope to have compiled enough data from Phase One of the project to present some meaningful recommendations based on the outcomes of current practices from Phase Two. The results of the survey will be presented at the Workshop meeting and subsequently published.

The project will continue after the 16th Workshop meeting, with the aim of developing:

- Ambiguity Resolution Documentation Formats
- Single-task Data-Management Tools
- Community Data-Sharing Standards
- Novel Data-Analysis Methods/Applications, and
- Reporting Guidelines for Immunogenomic Studies.

While additional project details will be forthcoming, we welcome the input and participation in these projects from the histocompatibility and immunogenetics community.

01000100011000010111010001100001
IMMUNOGENOMICS DATA
ANALYSIS WORKING GROUP
 GCTAATGCACTGTACAGTATATCA