IDAWG: The Immunogenomics Data-Analysis Working Group

Jill A. Hollenbach, Henry A. Erlich, Michael Feolo, Marcelo Fernandez-Vina, Wolfgang Helmberg, Uma Kanga, Pawinee Kupatawintu, Alex Lancaster, Martin Maiers, Hazaél Maldonado-Torres, Steven G.E. Marsh, Diogo Meyer, Derek Middleton, Carleheinz R. Müller, Oytip Nathalang, Myoung Hee Park, Richard M. Single, Brian Tait, Glenys Thomson, Ana Maria Valdes, Michael Varney and Steven J. Mack

ABSTRACT
The goal of the immunogenomics data analysis working group is to foster consistent analytical interpretation of immunogenetic data by the immunogenomics and larger genomics communities. Comprised of investigators from five continents, the working group aims to develop a set of community standards intended to facilitate the sharing of these data (HLA, KIR, etc.) and analyses, as well as develop novel methods and tools for immunogenomics data management and analysis.

The goal of the working group is to develop methods, standards, tools and recommendations intended to:

1) record, store and transmit immunogenomic data without obscuring the limitations of the typing method used, allowing easy identification of allele equivalency under successive nomenclatures, make data both human-readable (e.g., flat-text file) and machine-readable (e.g., XML file), conform to extant nomenclature rules, all without the use of proprietary platforms;

2) document ambiguity reduction (AR) methods used, permit reproducible AR, and permit equivalency under different AR methods;

3) foster portability between extant analysis tools and methods for maximum access to investigators (e.g., web-based tools);

4) encourage consistent data formats in future analytical methods, promoting widespread accessibility and application;

5) foster methodological consistency in the analysis of low frequency alleles and heterogeneous data, haplotype estimation, Hardy-Weinberg testing of highly polymorphic data, the application of measures of and adjustment for linkage disequilibrium, tests for selection and measures of population differentiation, the calculation of odds ratios, relative risks, etc., corrections for multiple testing, mitigation of false positive readings; and

6) develop novel methods of data analysis for highly polymorphic loci in disease association and population studies (e.g., peptide and nucleotide-level analyses, multidimensional scaling analyses, and neural network analyses).

We envision a collaborative effort by investigators particularly interested in issues of immunogenomics data management and analysis, with the goal of presenting our recommendations on these topics at the 16th BHWC follow-up by the publication of a reference manual.

Immunogenomics Data-Analysis Working Group Participants
Jill A. Hollenbach, co-chair
Center for Genetics, Children’s Hospital Oakland Research Institute, Oakland, CA, USA
Henry Erlich
Department of Human Genetics, Roche molecular Systems, Pleasanton, CA, USA
Michael Feolo
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA
Marcelo Fernandez-Vina
M.D. Anderson Cancer Center, University of Texas, Houston, TX, USA
Wolfgang Helmberg
Universitätshôd für Blutgruppentechnologie and Transfusionsmedizin, Medizinische Universität Graz, Graz, Austria
Uma Kanga
Department of Transplant Immunology and Immunogenetics, All India Institute of Medical Sciences, Amritsar, New Delhi, India
Pawinee Kupatawintu
Thai National Stem Cell Donor Registry, National Blood Centre, Thai Red Cross Society, Bangkok, Thailand
Alex Lancaster
Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA
Martin Maiers
National Marrow Donor Program, Minneapolis, MN, USA
Hazaél Maldonado-Torres
Anthony Nolan Research Institute, Royal Free Hospital, London, UK
Steven G.E. Marsh
Anthony Nolan Research Institute, Royal Free Hospital, London, UK
Diogo Meyer
Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de Sao Paulo, Sao Paulo, Brazil
Derek Middleton
ASUJ London
Transplant Immunology, Laboratory, Royal Liverpool University Hospital, Liverpool, UK
Carsteno R. Müller
Zentralbl-Knochenmarkspender-Register für die Bundesrepublik Deutschland GmbH, Ulm, Germany
Oytul Nathalang
Thai National Stem Cell Donor Registry, National Blood Centre, Thai Red Cross Society, Bangkok, Thailand
Myoung Hee Park
Department of Laboratory Medicine, Seoul National University Hospital, Seoul, Korea
Richard M. Single
Department of Medical Biostatistics, University of Vermont, Burlington, USA
Brian Tait
The Australian Red Cross Blood Service, Melbourne, Victoria, Australia
Glenys Thomson
Department of Integrative Biology, University of California, Berkeley, CA, USA
Ana María Valdes
Twin Research Unit, King’s College London, St Thomas’ Hospital, London, UK
Mike Varney
The Australian Red Cross Blood Service, Melbourne, Victoria, Australia

Solutions for Immunogenomic Data Management and Analysis
Given these challenges to consistent data-analysis, the Immunogenomics Data-Analysis Working Group proposes to develop data equivalency standards intended to foster consistency in the use of extant and future analytical methods, and to develop novel statistical and computational methodologies for the analysis of highly polymorphic loci.

In addition, we will determine the impact of various standards and methods for data management on downstream data-analyses, comparing them to extant immunogenomic data analysis systems, and producing recommendations for consistency in the analysis of highly polymorphic datasets.

Finally, we will promote the widespread accessibility and application of novel data equivalency and analytical tools by making them available to the community using web-based and multi-platform approaches.