

01000100011000010111010001100001

IDAWG

GCTAATGCACTGTACAGTATATCA

IMMUNOGENOMICS DATA

ANALYSIS WORKING GROUP

Martin Maiers, NMDP, USA

APHIA 2010-11-18

igdawg.org

01000100011000010111010001100001
IDAWG
GCTAATGCACTGTACAGTATATCA
IMMUNOGENOMICS DATA
ANALYSIS WORKING GROUP

HOME

OVERVIEW

RATIONALE

NEWS

CALENDAR

SOFTWARE

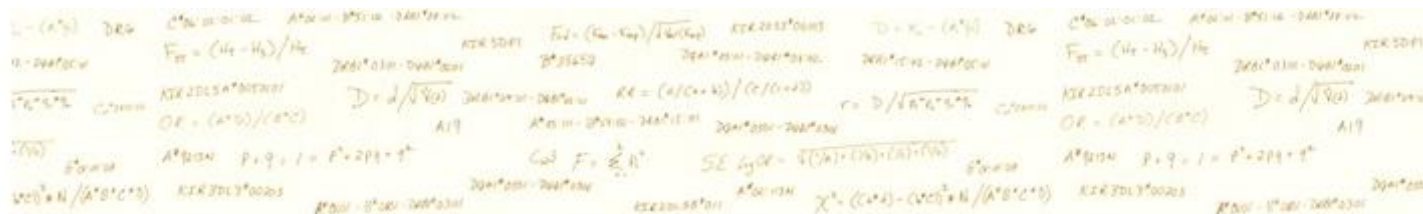
PUBLICATIONS

PARTICIPANTS

16TH IHIWS

CONTACT

LINKS



01000100011000010111010001100001
IMMUNOGENOMICS DATA
ANALYSIS WORKING GROUP
GCTAATGCACTGTACAGTATATCA

The immunogenomics data analysis working group (IDAWG) is a collaboration of twenty-three investigators from ten nations with the goal of facilitating the sharing of immunogenomic data (HLA, KIR, etc.) and fostering the consistent analysis and interpretation of those data by the immunogenomics community and the larger genomics communities.

The working group has been formed in advance of the [16th International HLA and Immunogenetics Workshop and Conference \(IHIWC\)](#) with the intent to present its recommendations on topics of data-management and data-analysis at the 16th IHIWC in 2012, followed by the publication of a reference manual.

For information about IDAWG projects for the 16th Workshop, [visit our IDAWG IHIWS page](#).

For information about the ANTT and UNCL software tools, [visit our software page](#).

Messages

- HLA is ambiguous
- HLA is not portable
- HLA is opaque
- Need Data Standards

Motivation

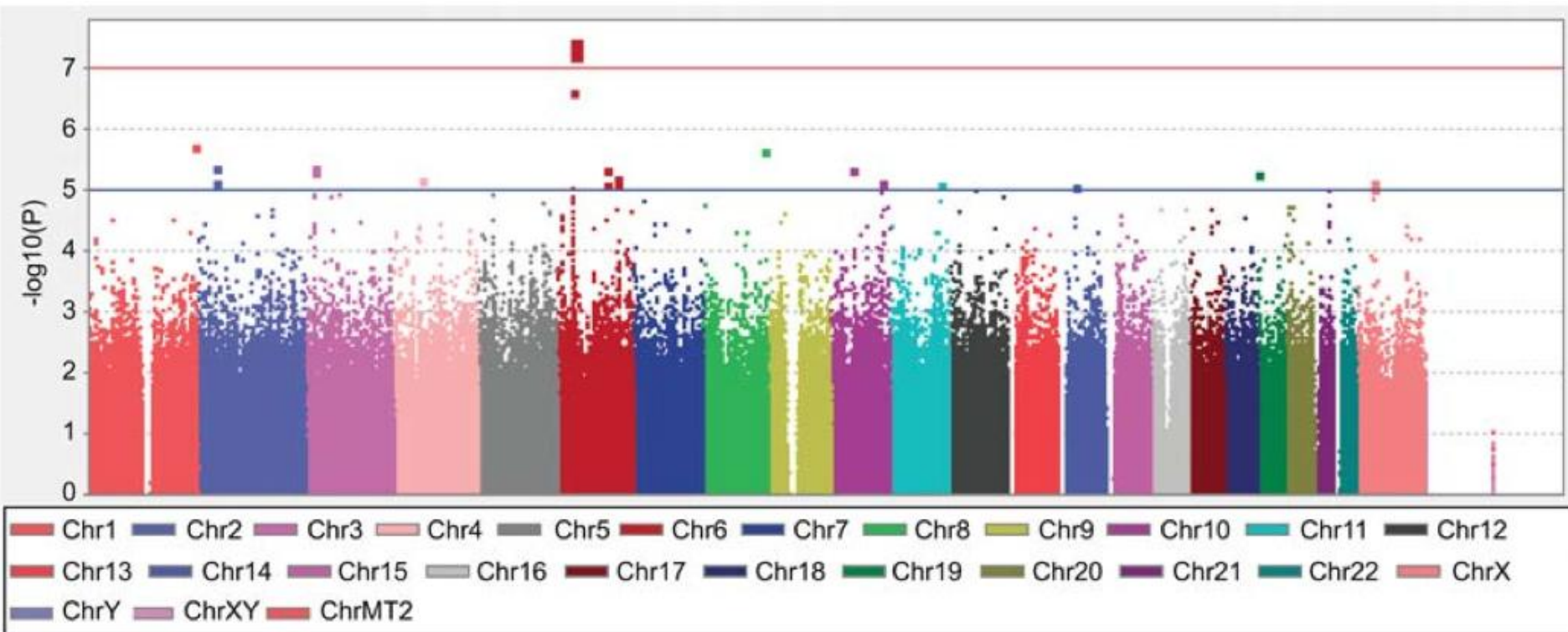
- Imagine a world without GenBank
- Imagine HLA without IMGT/HLA
- Look at BMDW:
 - 14,625,649 donors and 447,184 CBUs
 - All with some ambiguity
 - Ambiguity is weakly specified



- 405 platform presentations
- 2744 poster presentations
- 208 exhibiting companies and over
- 6,000 scientific registrants

ASHG

- Typical GWAS result



This horrendous alphanumeric mélange

“I must admit here and now to a serious personal weakness. I have a complete mental block when confronted with the bewildering variety of tissue types. Some of my best friends are cellular immunologists who live, work and breathe tissue types, and the Institute where I work is packed with them. Yet something switches off in my brain when they start describing the various types. All of them begin with the three letters HLA. Then numbers and letters are tacked on the end: HLA-DRB1, HLA-DPB2, HLA-B27 and so forth. Time and again I go to seminars which kick off with a slide showing a table of this horrendous alphanumeric mélange. For years I concentrated, thinking it would sink in eventually if I tried hard. After all, I have to teach this stuff in my genetics classes. But to no avail. I reluctantly conclude that I am genetically incapable of understanding tissue types beyond knowing that there are an awful lot of them. Which, fortunately, is all you need to know as well.”

Brian Sykes – “The seven daughters of Eve” p91

Consistency in Genetic Studies

- Human Genome Epidemiology Network (HuGENet)
 - 1998 CDC/Office of Public Health Genomics
 - to advance the synthesis and interpretation of population data on human genetic variation in disease association.
- Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) www.strobe-statement.org
- Strengthening the Reporting of Genetic Association studies (STREGA) statements



STROBE → STREGA

- Community-based reporting guidelines
- Seven of the authors are, or have been, editors of journals.
 - *PLoS Medicine*
 - *International Journal of Epidemiology*
 - *American Journal of Human Genetics*
 - *Paediatric and Perinatal Epidemiology*
 - *European Journal of Epidemiology*
 - *Lancet*
 - *Genetics in Medicine*
- Analyses of published studies suggest that implementation of **data reporting standards** for genetic association studies could greatly enhance the utility of these studies to the larger community through increased transparency

What about HLA/Immunogenetics?

- Many of the data-reporting issues described in these statements are pertinent to immunogenetic studies.
- However, these statements pertain primarily to large cohorts and single nucleotide polymorphism (SNP) based studies.
- The high level of polymorphism associated with the HLA and KIR loci requires specific consideration for the development of reporting standards and recommendations that go **beyond** those defined in the STROBE and STREGA statements.

Goal

- Outline the principle challenges to consistency in immunogenomic studies, in support of the development of a STREIS (STrengthening the REporting of Immunogenomic Studies) statement.

Constituencies

- Anthropology
- Disease Association
- Transplantation
 - Stem Cell Registries
 - Clinical Laboratories
- Science
 - Reproducibility
 - Immortality

HLA Ambiguities

- While an individual has only two HLA alleles per locus, many HLA genotyping results are *ambiguous* in that they include more than two possible alleles for a given sample at a given locus.

HLA ambiguities

At least 4 flavors

- Allelic ambiguity
- Genotypic ambiguity
- Undefined exons
- Historical context (expanding allele list)

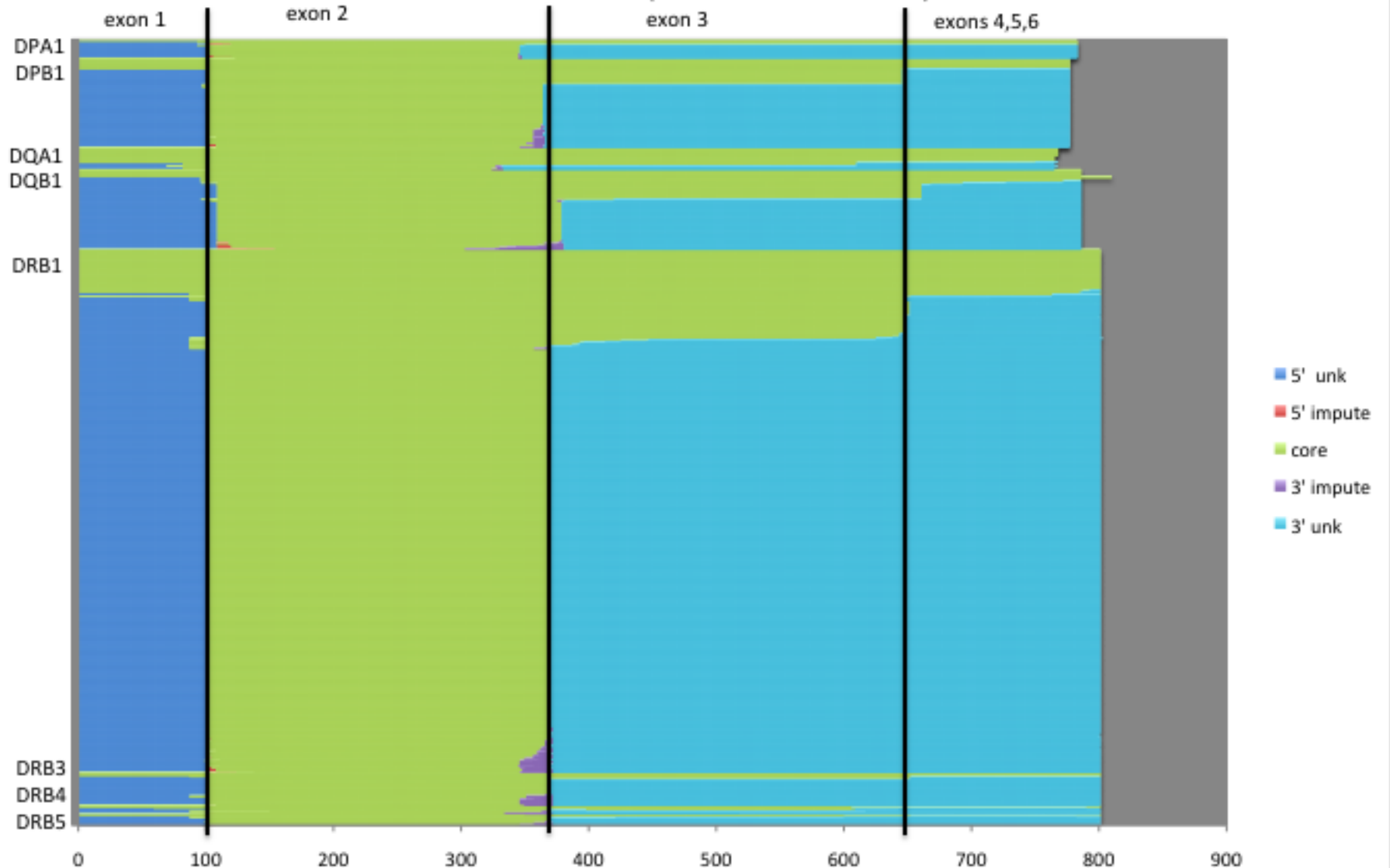
Allele Ambiguity

- *Allele ambiguity* results when the polymorphisms that distinguish alleles fall outside of the regions assessed by the genotyping system.
- A*02:07/A*02:15N/A*02:265

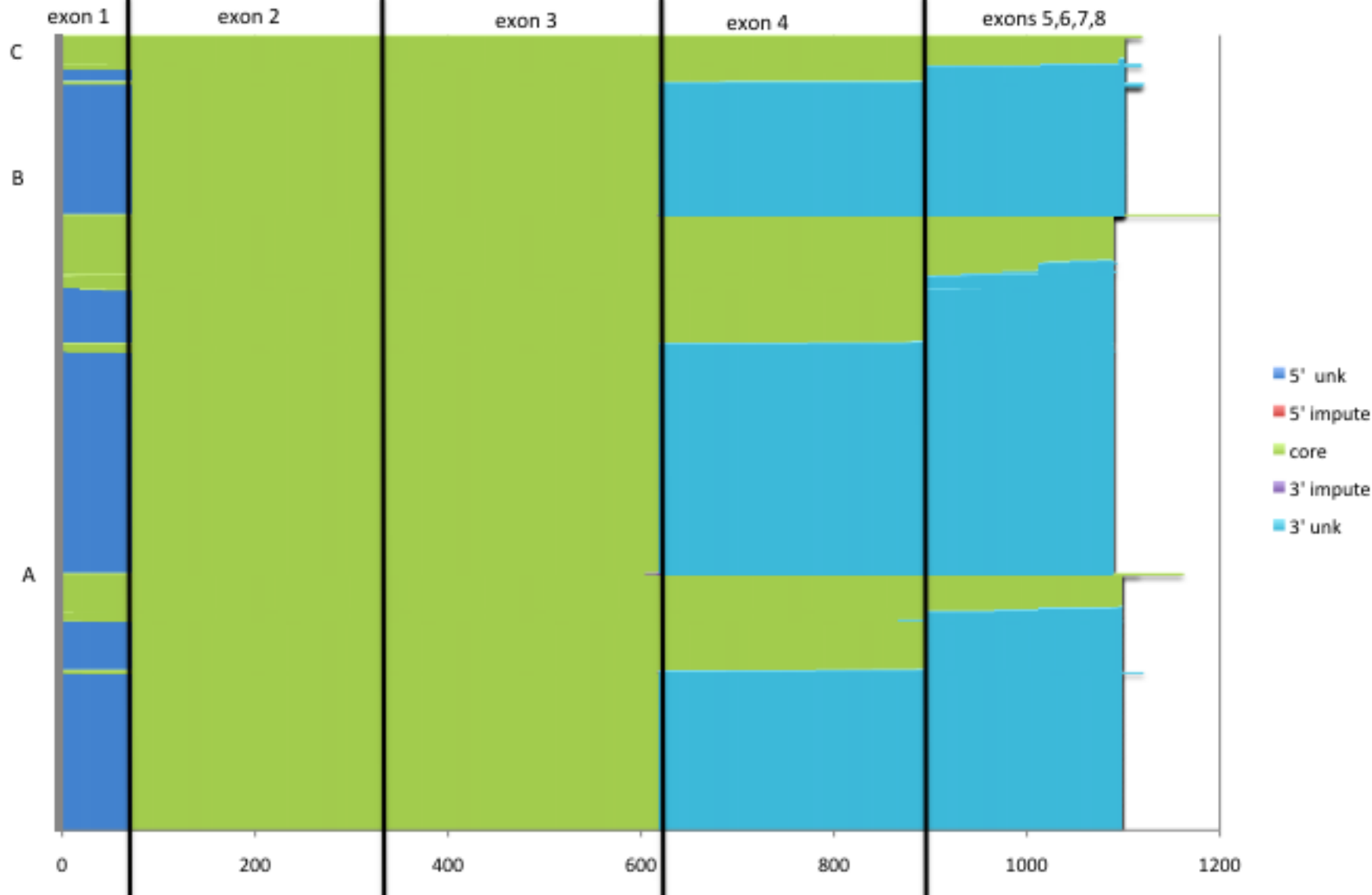
Genotype Ambiguity

- *Genotype ambiguity* results from an inability to establish chromosomal phase between identified polymorphisms.
- DRB1*04:01:01+DRB1*13:01:01,
DRB1*04:13+DRB1*13:02:01,
DRB1*04:14+DRB1*14:21,
DRB1*04:35+DRB1*13:40, and
DRB1*04:38+DRB1*13:20
have identical heterozygous DRB1 exon 2

HLA Class II (HLADB 3.2.0 N = 1248)



HLA Class I (HLADB 3.2.0 N = 4268)



Historical Context

- Expanding Allele list
- Increases all three forms of ambiguity over time
- Typing strategies adapt to new alleles
 - 2002 DRB1 exon3 for DRB1*1201/1206
(now DRB1*12:01/12:06/12:10/12:17)
 - 2008 DRB1 exon3 for DRB1*14:01/14:54

Historical Context

- Expanding Allele list
- Increases all three forms of ambiguity over time
- Typing strategies adapt to new alleles
 - 2002 DRB1 exon3 for DRB1*1201/1206
(now DRB1*12:01/12:06/12:10/12:17)
 - 2008 DRB1 exon3 for DRB1*14:01/14:54
 - 2012 DRB1 exon3 for DRB1*15:01/(future)

**Bioinformatics Integration Support Contract (BISC), Phase II
PROPOSAL FOR HLA DATA VALIDATION**



Version 1.0

Period Of Performance: September 30, 2004—September 29, 2010

Developed Under Contract Number: HHSN266200400076C

ADB Contract Number: N01-A1-40041

Delivered: June, 2008

Project Sponsor:

National Institutes of Health (NIH)
National Institute of Allergy and Infectious Diseases (NIAID)
Division of Allergy, Immunology, and Transplantation (DAIT)

Prepared by:

Steven J. Mack
Richard M. Single
Henry A. Erlich
Glenys Thomson

Silver Standard

- data validation via a standard reporting of HLA data to include:
 - Documentation of the **typing method** used. For each locus, this includes:
 - Manufacturer
 - Version
 - Molecular nature of the data (e.g., SBT, PCR-SSOP, SSP).
 - Exons interrogated
 - A list of all detectable alleles (IMGT/HLADB version)
 - A sublist of all alleles that can be detected unambiguously by the system
 - A sublist of all allele sets detected ambiguously by the system
 - A file listing **all allele and genotype ambiguities** per individual per locus
 - Documentation of the **rules** used to assign the most likely allele “calls”
 - A file listing the two most likely allele “calls”
- **The gold standard** for HLA genotyping data-collection would be to record the sequence, primer, probes and reactivity patterns for every individual per locus.
 - Such systems has been developed at the NCBI & NMDP

IDAWG 16th IHIWS PROJECT: Immunogenomic Data Management Methods

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.

Survey current HLA and KIR data-management practices to determine:

- Current HLA and KIR data-capture and data transmission standards,
- Current HLA ambiguity resolution practices, and
- Current primary data-analysis methods.

With the aim of developing:

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

- we welcome the input and participation in these projects from the histocompatibility and immunogenetics community.

Institution	Participant
Children's Hospital Oakland Research Institute, Oakland, United States of America	Jill A. Hollenbach Steven J. Mack
Roche Molecular Systems, Pleasanton, United States of America	Henry Erlich
National Institutes of Health, Bethesda, United States of America	Michael Feolo
University of Texas, Houston, United States of America	Marcelo Fernandez-Vina
University of California, San Francisco, San Francisco, United States of America	Pierre-Antoine Gourraud
Medizinische Universität Graz, Graz, Austria	Wolfgang Helmberg
All India Institute of Medical Sciences, New Delhi, India	Uma Kanga
National Blood Centre, Thai Red Cross Society, Bangkok, Thailand	Pawinee Kupatawintu Oytip Nathalang
University of Arizona, Tuscon, United States of America	Alex Lancaster
National Marrow Donor Program, Minneapolis, United States of America	Martin Maiers
Royal Free Hospital, London, United Kingdom (Great Britain)	Hazael Maldonado-Torres Steven G.E. Marsh
Universidade de Sao Paulo, Sao Paolo, Brazil	Diogo Meyer
Royal Liverpool University Hospital, Liverpool, United Kingdom (Great Britain)	Derek Middleton
Zentrales Knochenmarkspender-Register fuer die Bundesrepublik Deutschland GmbH, Ulm, Germany	CarlHeinz R. Müller
Seoul National University Hospital, Seoul, Korea, Republic of	Myoung Hee Park
University of Vermont, Burlington, United States of America	Richard M. Single
The Australian Red Cross Blood Service, Melbourne, Australia	Mike Varney , Brian Tait
University of California, Berkeley, Berkeley, United States of America	Glenys Thomson
King's College London, St Thomas' Hospital, London, United States of America	Ana Maria Valdes

idawg.org

Welcome to Idaho Domestic Animal Welfare Group or IDAWG.

We are under construction. Please visit our [PetFinders](#) site. Thanks!

You will automatically be redirected to this site.



Hurricane Katrina survivor, Rosa, taken into foster care by IDAWG.

HARMONISATION OF DEFINITIONS OF HISTOCOMPATIBILITY TYPING TERMS

The purpose of the **Harmonisation of Histocompatibility Typing Terms Statement** is to recommend and to initiate discussion regarding terminology, definitions of resolution levels, format for reporting HLA assignments and match reporting for transplantation.

Statement issued Oct 13th 2010.

The Deadline for receipt of comments via EFI is December 31st 2010.

Harmonisation Group 2010

Marcelo Fernandez-Vina (ASHI)

Brad Eisenbrey (CAP)

Kay Poulton (EFI)

Cyndi Taves (ASHI)

Gottfried Fischer (EFI)

Eduardo Nunes (AABB)

Carolyn Hurley (NMDP, WMDA)

Dawn Wagenknecht (ASHI)

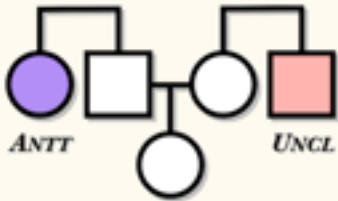
Helen Heslop (ASBMT)

Kara Wacker (FACT)

Harriet Noreen (ASHI, NMDP)

Nicoletta Sacchi (WMDA)

ANTT & UNCL



Allele Name Translation

We have developed two applications that will translate non-colon-delimited HLA allele names (e.g., A*01010101) to their colon-delimited counterparts (e.g., A*01:01:01:01).

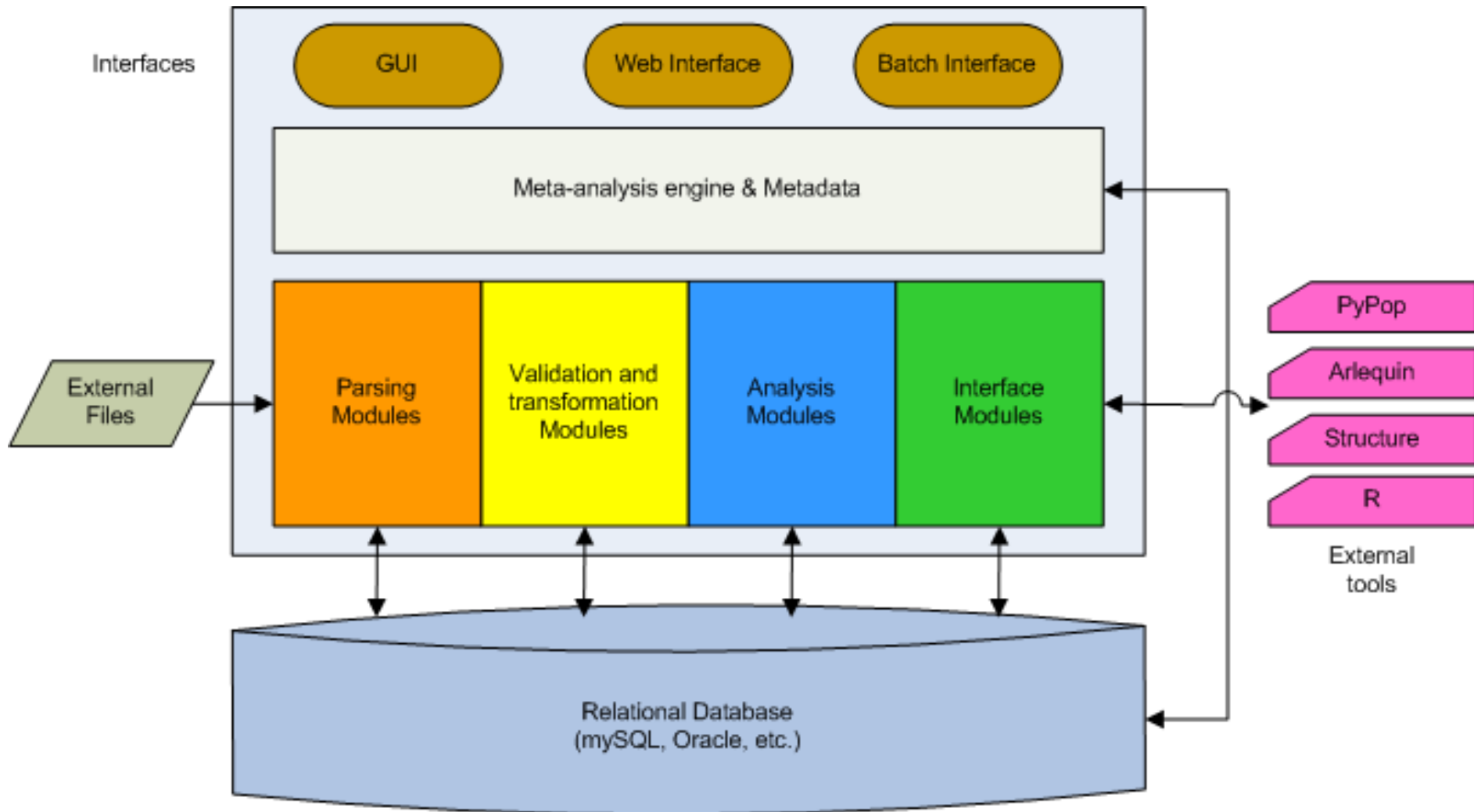
The **Allele Name Translation Tool (ANTT)** is a downloadable application that translates the allele names in entire datasets, and can be customized to translate allele names between any pair of naming conventions (including user-defined naming conventions).



Update NomenCLature (UNCL) is a web-based implementation of an R-script that translates the allele names in entire datasets. UNCL can be used by anyone with an internet connection and a modern web-browser.



Standard Data Management Tools



Fundamental HLA Model

- The foundation of the standard is a **formal model** of HLA
- UML (Unified Modeling Language)
 - Software engineering tool
 - Curated terminology (concepts)
 - Relationships
 - Generate
 - Data formats (XML)
 - Database schemas
 - Standard interfaces