Sequence Ontology Workshop

MHC/HLA data in the Immunology Database and Analysis Portal - *ImmPort*

Bioinformatics Integration Support Contract (BISC)
HHSN266200400076C
N01AI40076

12 JUNE 2007

NORTHROP GRUMMAN Information Technology

DEPARTMENT OF HEALTH & HUMAN SERVICES · USA

NATIONAL INSTITUTES OF HEALTH

NIAID

SOUTHWESTERN

Bioinformatics Support for Immunology Community

- Investigators must be able to **extract meaningful information from the vast amounts of data** generated from advanced research technologies
- This requires the **collection, integration and analysis of research data from numerous, diverse sources, and their long-term storage in sustainable databases**
- To facilitate these processes, the scientific community must work to jointly develop essential **minimum information sets** and **shared vocabularies** for experiment records and knowledge description
- In response to these needs, the NIAID/NIH funded the **Bioinformatics Integration Support Contract (BISC)** to provide advanced IT support in the production, analysis, archiving, exchange and integration of genomic, proteomic and related data
- The **users of BISC** include NIAID/DAIT programs that conduct:
  - basic scientific research of immune system development and function
  - genetic determinants of immune disease
  - clinical trials to evaluate the safety, toxicity, and efficacy of immune disease therapies
  - studies of the underlying mechanisms of therapeutic agents
- Immunology Database and Analysis Portal - ***ImmPort*** - www.immport.org

HLA & Immune Disease Program

HLA Region Genetics in Immune-Mediated Diseases

Objective - To support prospective and/or retrospective studies to investigate the role of HLA genetics in susceptibility to or protection from immune-mediated diseases, including autoimmune diseases and primary immunodeficiency diseases, GVHD, and graft rejection or survival in solid organ, tissue and cell transplantation.

Data handling - generate high quality HLA-disease association data for public use that will be submitted to and maintained by dbMHC through the National Center for Biomedical Information (NCBI). Analysis of data results and data submission to dbMHC will be performed through the NIAID Bioinformatics Information Support Contract (BISC)

HLA/KIR Region Genetics in Pediatric Arthritis	**Dr. David Glass** **University of Cincinatti**
HLA Region Genetics and SLE in US Black Women	**Dr. Lynn Rosenberg** **Boston University**
The Role of HLA and KIR in Rheumatoid Arthritis and Crohn's Disease"	**Dr. Henry Erlich** **Children's Hospital Oakland Research Institute**
Hematopoietic Cell Transplantation	**Dr. Effie Petersdorf** **Fred Hutchison Cancer Research Center**
Molecular Genetics of HLA and Disease	**Dr. Stephen Hauser** **University of California, San Francisco**

IMMPORT

Data Interoperability

Minimum data standards

+

Standardized ontologies

+

Extensible data model

=

Data interoperability

Description Framework

ImmPort

Welcome to ImmPort version 1.4.1! The ImmPort system provides advanced information technology support in the production, analysis, archiving, and exchange of scientific data for the diverse community of life science researchers supported by NIAID/DAIT. It serves as a long-term, sustainable archive of data generated by investigators funded through the NIAID/DAIT.

We welcome any comments/suggestions from our users. If you wish to provide feedback on the use of ImmPort, Please feel free to contact us by email at helpdesk@immport.org

Administration

- My Projects
- My Contracts/Grants

Research Data

- Experiment Data Submission Tutorial
- Submission History
- Research Data Management

Search Reference Data

- Genes
- Proteins
- Pathways
- Protein Network
- SNP
- ImmPort Genes

Search Research Data Tutorial

Experiment Attributes

- Simple Search
- Subjects
- Biological Samples
- Experiment Variables
- Reagents
- Protocols
- Experiment Descriptions
- Analytes (Flow Cytometry)

Experiment Results

- Gene Expression
- Flow Cytometry

Tools & Analysis

- tagSNP Selection Tutorial
- Gene Expression Analysis
- Genetic Analysis
- Genome Browser
- Ontology
- Diseases
- Gene Ontology
- Draft GO - Immunology
- Taxonomy

IMMPORT

ImmPort Overview - System Components

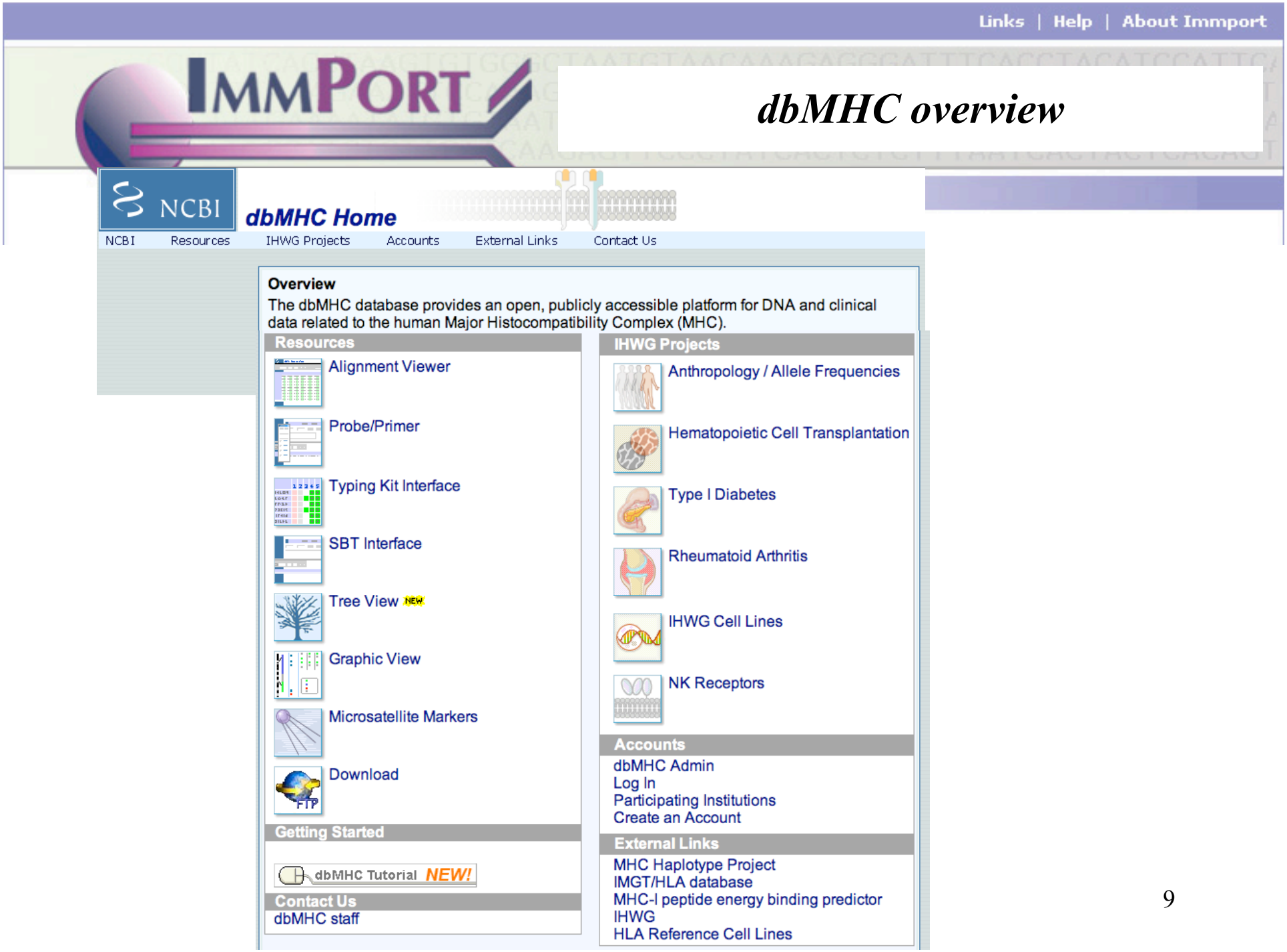
- **Semi-public web-based database and analysis portal**
  - Multi-level access control
  - Data sharing
- **Data**
  - **Reference data**
    - **Types** - Gene structure, protein function, polymorphisms, metabolic, regulatory, signaling and other networks, protein-protein, gene-gene, host-pathogen interactions
    - **Sources** - NCBI, Uniprot, Swissprot, BIND, Reactome
  - **Experiment data**
    - **Metadata** (defining how the experiment was performed) - common features of all experiments
    - **Primary results** - from all experiment measurement techniques
    - **Processed results**
      - Interpreted results
      - Analytical metadata
  - **Clinical trial data**
- **Query tools**
  - To support retrieval of reference and experiment data based on specified criteria
  - Pre-defined QBE
  - Customized semantic queries
- **Ontology**
  - Thesaurus function
  - Organize terms and define relationships
- **Analysis tools**
  - **Genetic analysis**
    - LD analysis
    - TagSNP selection
    - Haplotype reconstruction
    - Genotype-phenotype association
  - **Gene expression analysis**
    - Filtering/normalization
    - Clustering
    - Classification
  - **Cell population analysis**
    - Standard FACS statistics
    - Novel population identification based on high dimensional data clustering
  - **Measurement of immune response (**e.g., ELISA, ELISPOT)
    - Statistical analysis of distributions
  - **Biological network analysis**
    - Quantification of topological parameters
    - Module identification
- **Visualization** tools
  - Genome display, including genes, introns, exons, SNPs, tagSNPs, etc.
  - Genetic analysis results
  - Gene expression results
  - Networks, pathways, and molecular interactions
  - Graphing and charting of statistical results

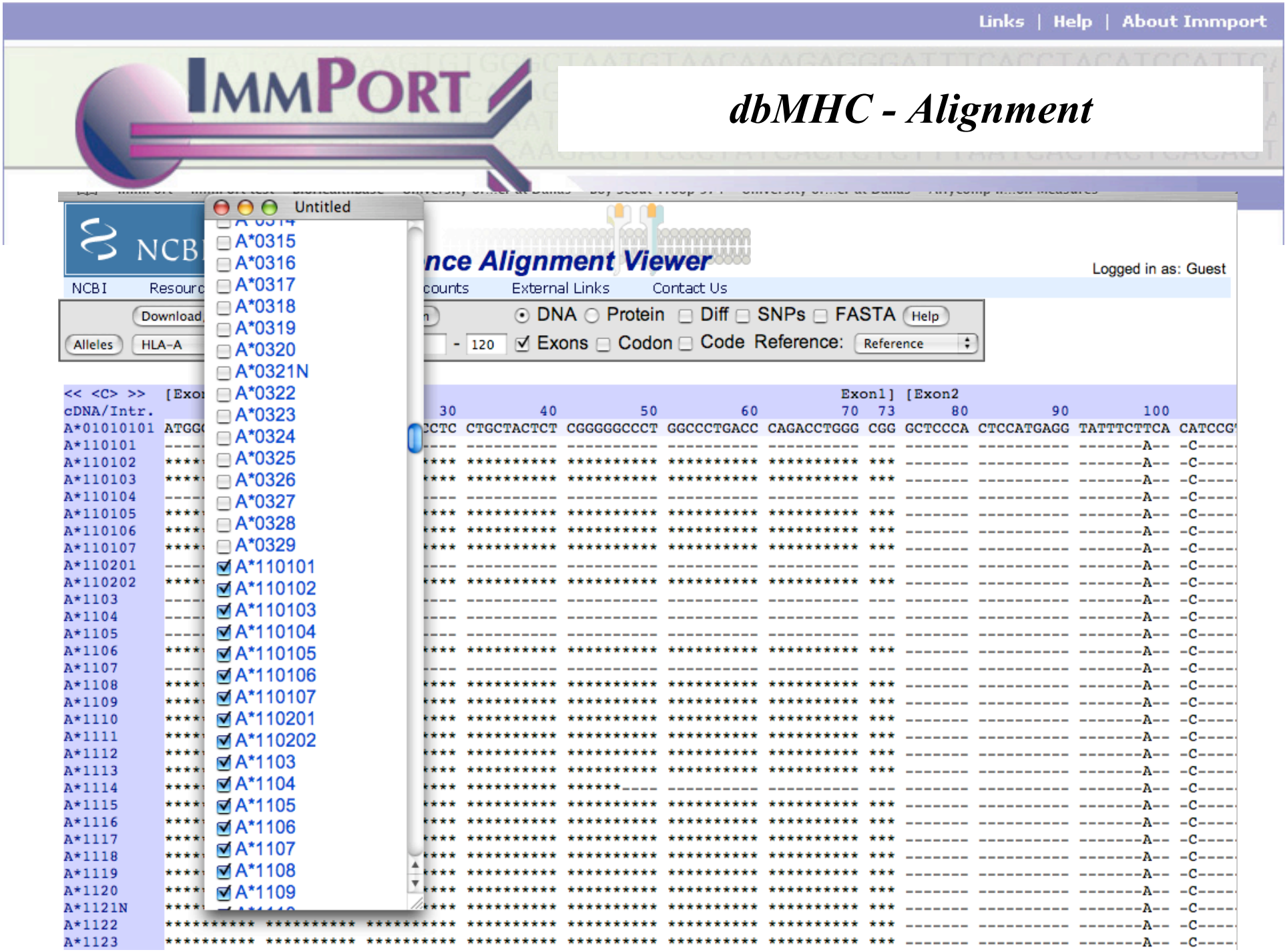
DAIT Interoperability Working Group

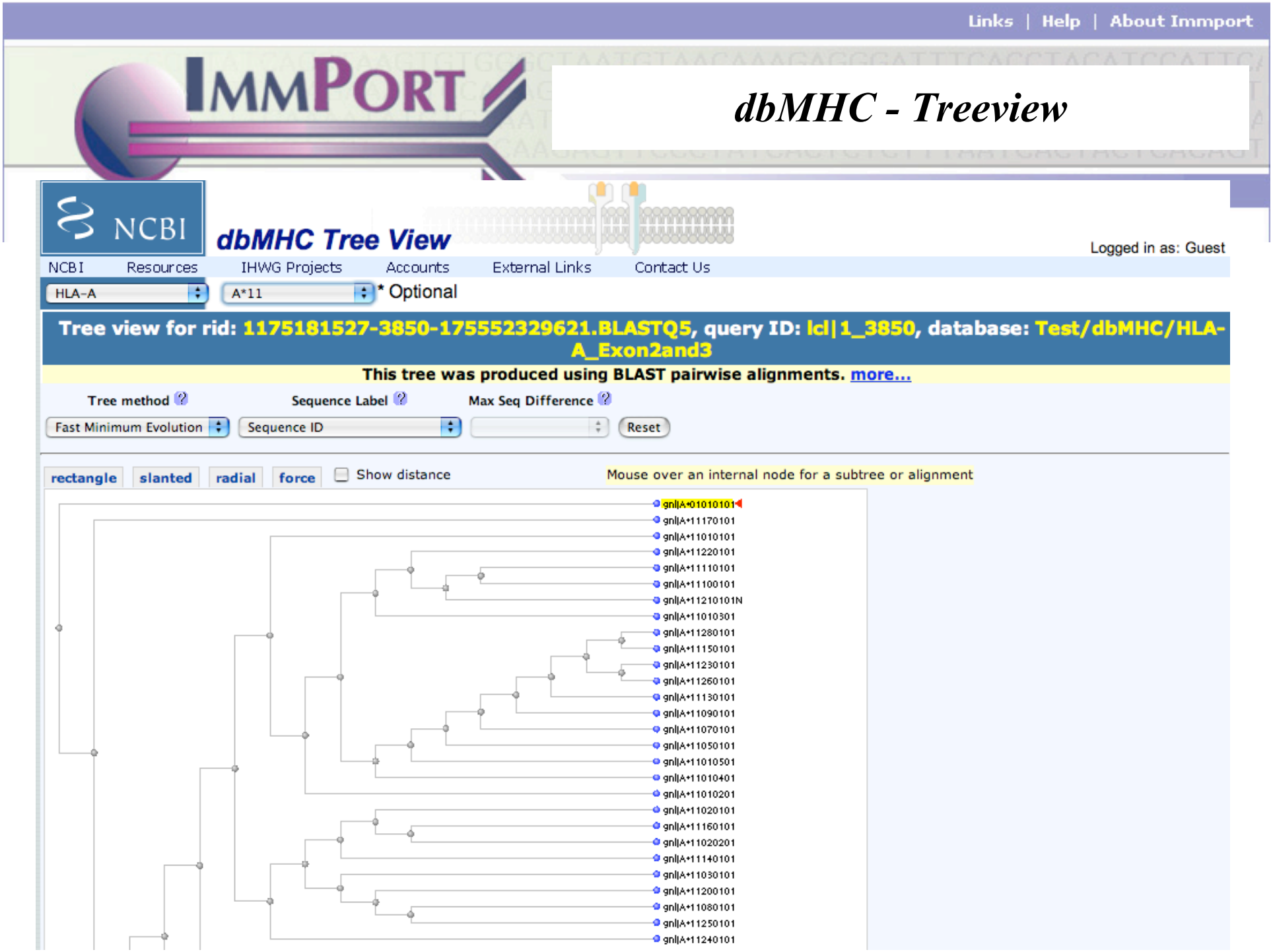
- In February of 2006, DAIT held a Bioinformatics Summit. One of the recommendations of the summit was to establish a committee to develop and support the use of informatics standards to facilitate interoperability among DAIT-funded programs.
- The DAIT Interoperability Steering Committee (DISC) was established in December 2006.
- Two initial projects were established
  - HLA ontology
  - Clinical trial ontology/data model

Use cases

- Is there a significant association between any HLA allele and _____?
- Is there a significant association between any HLA genotype and _____?
- Is there a significant association between any HLA haplotype and _____?
- Is there a significant association between any HLA allele group and _____?


Links | Help | About Immport
IMMPORT
dbMHC overview
NCBI
dbMHC Home
NCBI Resources IHWG Projects Accounts External Links Contact Us
Overview
The dbMHC database provides an open, publicly accessible platform for DNA and clinical data related to the human Major Histocompatibility Complex (MHC).
Resources
Alignment Viewer
Probe/Primer
Typing Kit Interface
SBT Interface
Tree View NEW
Graphic View
Microsatellite Markers
Download
Getting Started
dbMHC Tutorial NEW!
Contact Us
dbMHC staff
IHWG Projects
Anthropology / Allele Frequencies
Hematopoietic Cell Transplantation
Type I Diabetes
Rheumatoid Arthritis
IHWG Cell Lines
NK Receptors
Accounts
dbMHC Admin
Log In
Participating Institutions
Create an Account
External Links
MHC Haplotype Project
IMGT/HLA database
MHC-I peptide energy binding predictor
IHWG
HLA Reference Cell Lines


Links | Help | About Immport
ImmPort
dbMHC - Alignment
Untitled
A*0315
A*0316
A*0317
A*0318
A*0319
A*0320
A*0321N
A*0322
A*0323
A*0324
A*0325
A*0326
A*0327
A*0328
A*0329
A*110101
A*110102
A*110103
A*110104
A*110105
A*110106
A*110107
A*110201
A*110202
A*1103
A*1104
A*1105
A*1106
A*1107
A*1108
A*1109
nce Alignment Viewer
Logged in as: Guest
NCBI
External Links
Contact Us
Download
Alleles
HLA-A
DNA
Protein
Diff
SNPs
FASTA
Help
120
Exons
Codon
Code
Reference:
Reference
Exon1] [Exon2


Links | Help | About Immport
IMMPORT
dbMHC - Treeview
NCBI
dbMHC Tree View
Logged in as: Guest
NCBI Resources IHWG Projects Accounts External Links Contact Us
HLA-A
A*11
* Optional
Tree view for rid: 1175181527-3850-175552329621.BLASTQ5, query ID: lcl|1_3850, database: Test/dbMHC/HLA-A_Exon2and3
This tree was produced using BLAST pairwise alignments. more...
Tree method
Sequence Label
Max Seq Difference
Fast Minimum Evolution
Sequence ID
Reset
rectangle slanted radial force
Show distance
Mouse over an internal node for a subtree or alignment
gnl|A*01010101
gnl|A*11170101
gnl|A*11010101
gnl|A*11220101
gnl|A*11110101
gnl|A*11100101
gnl|A*11210101N
gnl|A*11010301
gnl|A*11280101
gnl|A*11150101
gnl|A*11230101
gnl|A*11260101
gnl|A*11130101
gnl|A*11090101
gnl|A*11070101
gnl|A*11050101
gnl|A*11010501
gnl|A*11010401
gnl|A*11010201
gnl|A*11020101
gnl|A*11160101
gnl|A*11020201
gnl|A*11140101
gnl|A*11030101
gnl|A*11200101
gnl|A*11080101
gnl|A*11250101
gnl|A*11240101

NCBI

dbMHC Microsatellite Markers

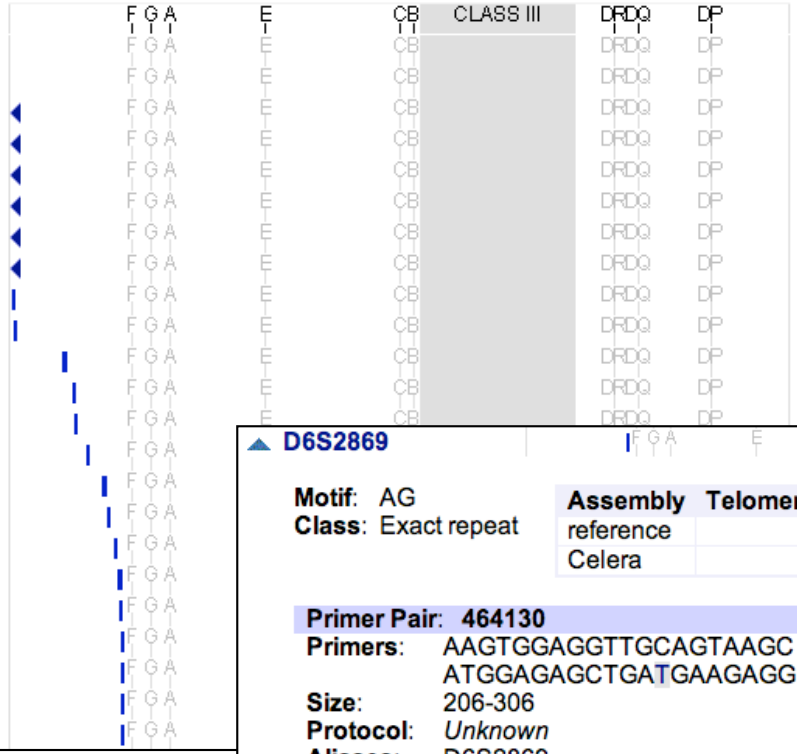
NCBI Resources IHWG Projects Accounts External Links Contact Us

HLA Microsatellite Markers

Search Criteria

Modify Search | New Search | Download Data

View All Details | Hide All Details



D6S2869

Motif: AG
Class: Exact repeat

Assembly	Telomeric Gene (kbp)	Centromeric Gene (kbp)
reference		HLA-F (48)
Celera		HLA-F (45)

Primer Pair: **464130**
Primers: AAGTGGAGGTTGCAGTAAGC
ATGGAGAGCTGATGAAGAGG
Size: 206-306
Protocol: *Unknown*
Aliases: D6S2869
Links: UniSTS e-PCR
SNPs: rs4713230, rs11357168, rs9257941, rs3129194, rs9278233, rs3129195, rs9280637, rs9501680

Assembly	Length	Chr. Location	Contig	Contig Location	Strand	Mismatch	Deletion
reference	248	29750838-29751085	NT_007592.14 sv mv Vega Ensembl Glovar	20501109-20501356	+	0	0
consensus	256	*Unknown*	consensus	940723-940978	+	0	0
Celera	255	29496124-29496378	NT_086688.1 sv mv	2711712-2711966	+	0	0

Type 1 Diabetes

The disease mainly occurs in individuals carrying a genetic predisposition. A number of genes across the genome have been linked to diabetes susceptibility (insulin, CTLA-4, *etc.*) Among these, genes in the HLA (MHC) complex on chromosome 6p account for about 50% of the genetic susceptibility. The HLA class II DR and DQ loci appear as primary determinants of predisposition to and protection from T1D. However, class II genes cannot explain all of the linkage and association between chromosome 6p and T1D. In particular, there is evidence that HLA class I alleles (HLA-A and HLA-B) modulate disease susceptibility and its clinical features. Recent data suggest the possible presence of an additional susceptibility gene telomeric to HLA-F. Other reports have provided evidence for another susceptibility locus, independent of HLA class II genes, located near the TNF locus in the class III region.

Goals

The overall goal of the type 1 diabetes (T1D) component of the 13th IHWG was to gain insight into additional genes within the MHC with effects on disease susceptibility and resistance. The following are the key aims of this effort:

1. To capture existing data from investigators worldwide, to create a resource consisting of available HLA haplotypes and demographic/phenotypic data from patients with T1D, family members, and population matched controls.

2. To make the captured data available to investigators worldwide through dbMHC. Thirty-three laboratories have joined the T1D Component of the 13th IHWG from various parts of the world. These laboratories generously shared data and DNA samples to assemble the 13th IHWG T1D Dataset, which includes samples with significant ethnic diversity from:

   - Over 3,000 patients and 2,000 population-based controls from 20 different data sets (case-control study).
   - Approximately 500 fam

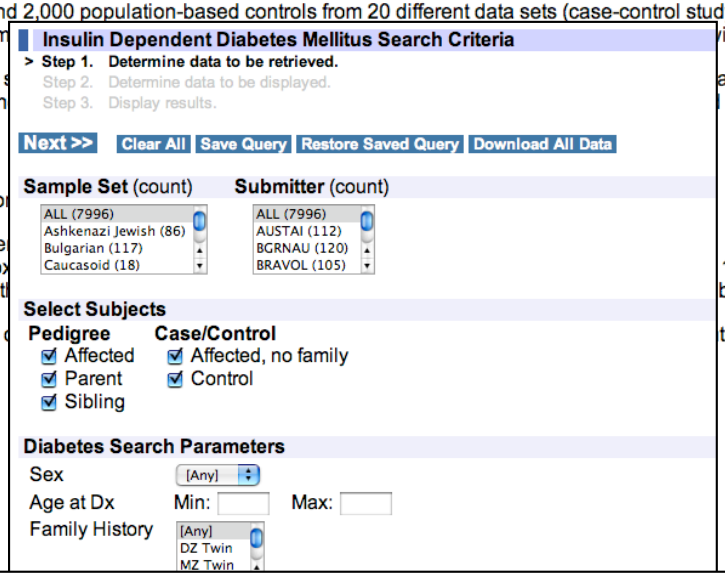
3. Using DNA from these data s
   II loci. The plan is to combin

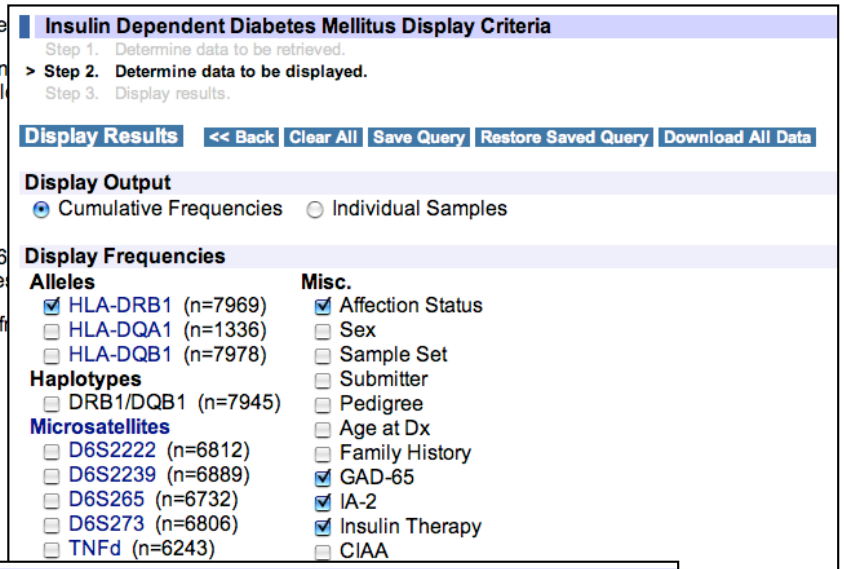
dbMHC Content

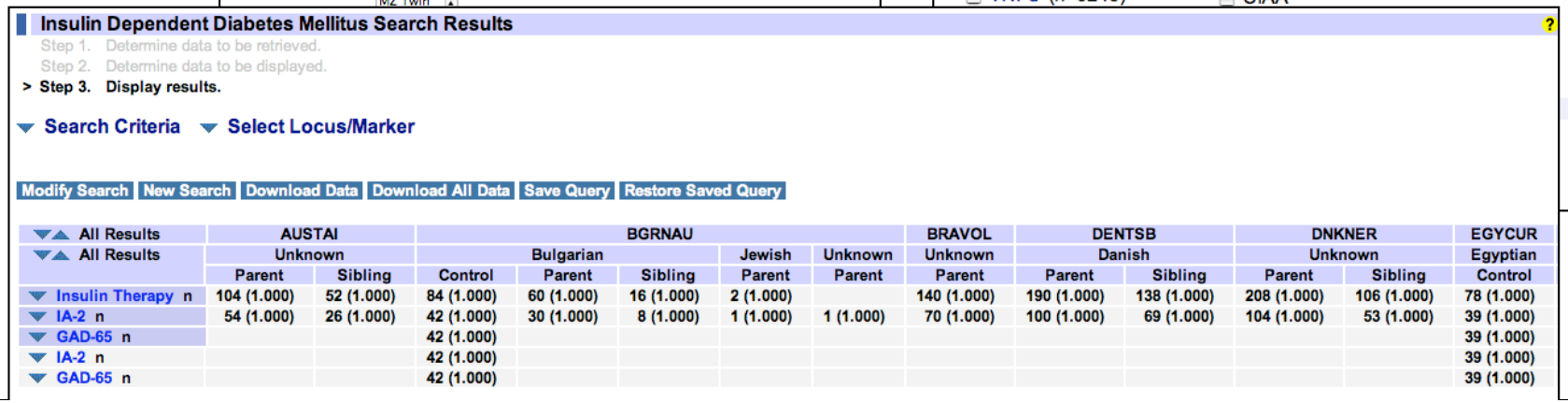
As of October 2005, dbMHC co

- Class II (DRB1/DQB1) ge
- Microsatellite data: approx
- Available data on race, et

NOTE: The database includes



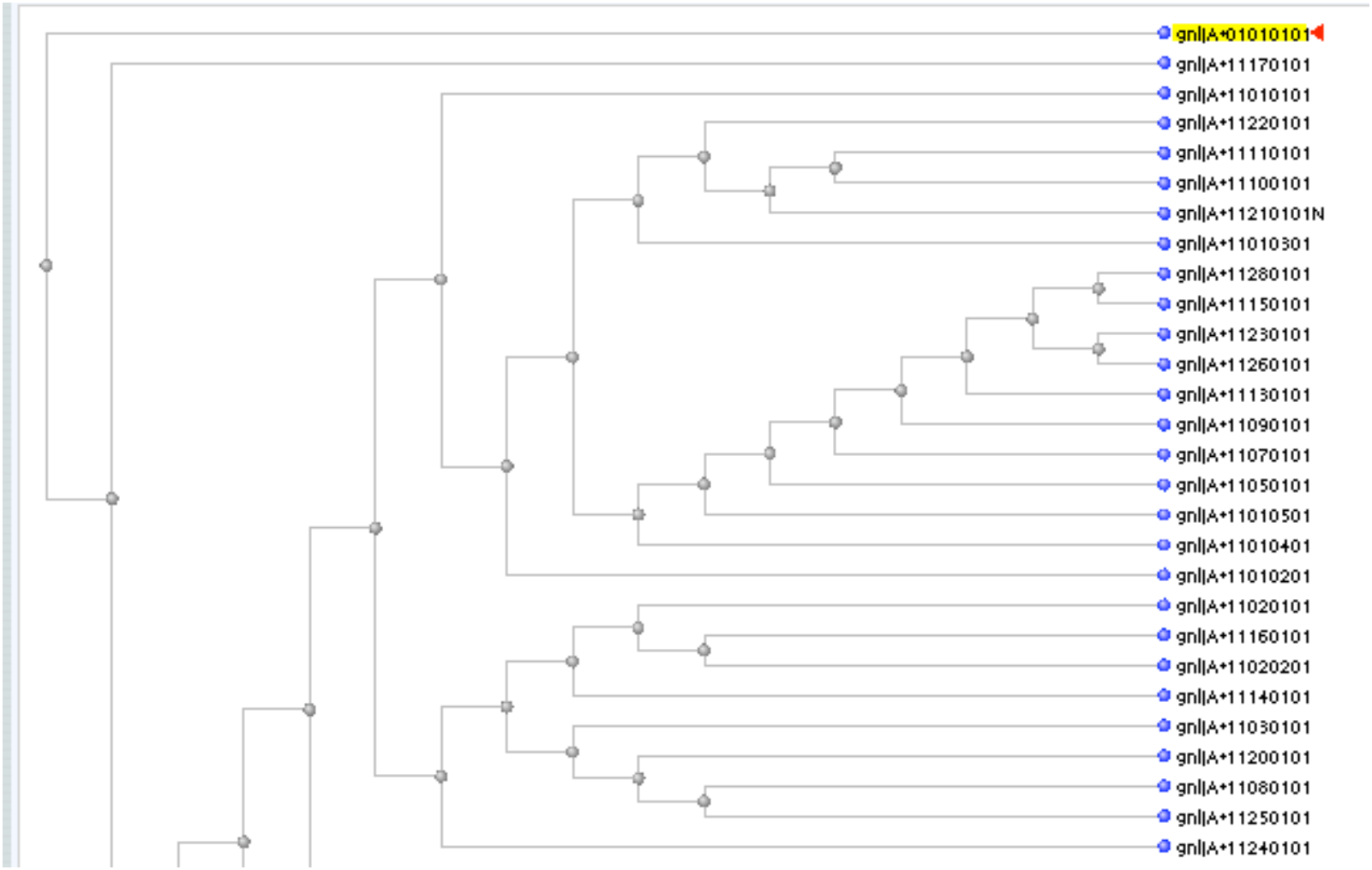




All Results	AUSTAI		BGRNAU					BRAVOL	DENTSB		DNKNER		EGYCUR
All Results	Unknown		Bulgarian			Jewish	Unknown	Unknown	Danish		Unknown		Egyptian
	Parent	Sibling	Control	Parent	Sibling	Parent	Parent	Parent	Parent	Sibling	Parent	Sibling	Control
Insulin Therapy n	104 (1.000)	52 (1.000)	84 (1.000)	60 (1.000)	16 (1.000)	2 (1.000)		140 (1.000)	190 (1.000)	138 (1.000)	208 (1.000)	106 (1.000)	78 (1.000)
IA-2 n	54 (1.000)	26 (1.000)	42 (1.000)	30 (1.000)	8 (1.000)	1 (1.000)	1 (1.000)	70 (1.000)	100 (1.000)	69 (1.000)	104 (1.000)	53 (1.000)	39 (1.000)
GAD-65 n			42 (1.000)										39 (1.000)
IA-2 n			42 (1.000)										39 (1.000)
GAD-65 n			42 (1.000)										39 (1.000)

IMMPORT

HLA allele hierarchy



Toward an HLA Ontology - Why?

- To support data interoperability for data exchange - unambiguous representation
- To support meta-analysis of multiple independent studies
- To support inferential reasoning based on HLA relationships defined in the ontology structure

Toward an HLA Ontology - What?

- Capture the hierarchical relationships between alleles defined by different methodologies and at different levels of resolution
- Define HLA haplotypes
- Develop procedures for adding new alleles to the HLA ontology framework
- Make information encoded in the nomenclature explicit
- Enumerate and distinguish uses of HLA data: presence, restriction, association, ...

IMMPORT

Toward an HLA Ontology - Who?

- HLA nomenclature people, specifically IMGT/HLA, Anthony Nolan Trust, WHO, dbMHC, other IHWG members
- Biomedical ontology people - SO group
- DISC HLA Working Group members - Petersdorf, Karp, Peters, Scheuermann

Toward an HLA Ontology - How?

1. Review history and current status of HLA nomenclature and typing methodologies
2. Review biomedical ontology best practices, e.g. OBO Foundry
3. Review the current Sequence Ontology content and structure
4. Define HLA Ontology design principles and constraints
5. Assemble working groups
6. Define goals, timelines and milestones
7. Develop HLA ontology branches
8. Merge branches into initial HLA ontology draft
9. Vet HLA ontology draft with key stakeholders, including the IHWG/HLA, WHO, dbMHC, DISC HLA WG
10. Revise draft into HLA Allele Ontology v1.0 within the SO
11. Develop procedures for adding new alleles to framework
12. Integrate ontology information into HLA web resources like ImmPort and dbMHC
13. Submit publication

Other nomenclature issues

- Genotype, allele, variant, polymorphism, haplotype
- SNP, microsatellite, VNTR
- Wildtype, normal, reference
- Genetics
  - Homozygous, heterozygous
  - Hardy-Weinberg equilibrium
  - Linkage disequilibrium
- KIR family (killer cell immunoglobulin like receptor)