



IMMUNOLOGY DATABASE
AND ANALYSIS PORTAL

SNP Genotype Results Harmonization

25 September 2008

Teleconference # 1-866-740-2200

Pass code 5630833



IMMUNOLOGY DATABASE AND ANALYSIS PORTAL

BISC

Bioinformatics Integration and Support Contract

Goals:

Enable Storage, Sharing, and Analysis of Research Results

Methods:

- Collect results
- Facilitate Organized Description of Research Subject and Methods
- Transform native result formats into analysis data sets

Example: Generate genetic analysis data sets by aggregating subject descriptions with genotype results



IMMUNOLOGY DATABASE
AND ANALYSIS PORTAL

BISC: A real world example

HLA SNPs Consortium Project

Goal:

Share SNP and HLA genetic data to construct a dense and highly accurate MHC haplotype map.

Requires a very large well-characterized sample set defined with a common set of SNPs



IMMUNOLOGY DATABASE
AND ANALYSIS PORTAL

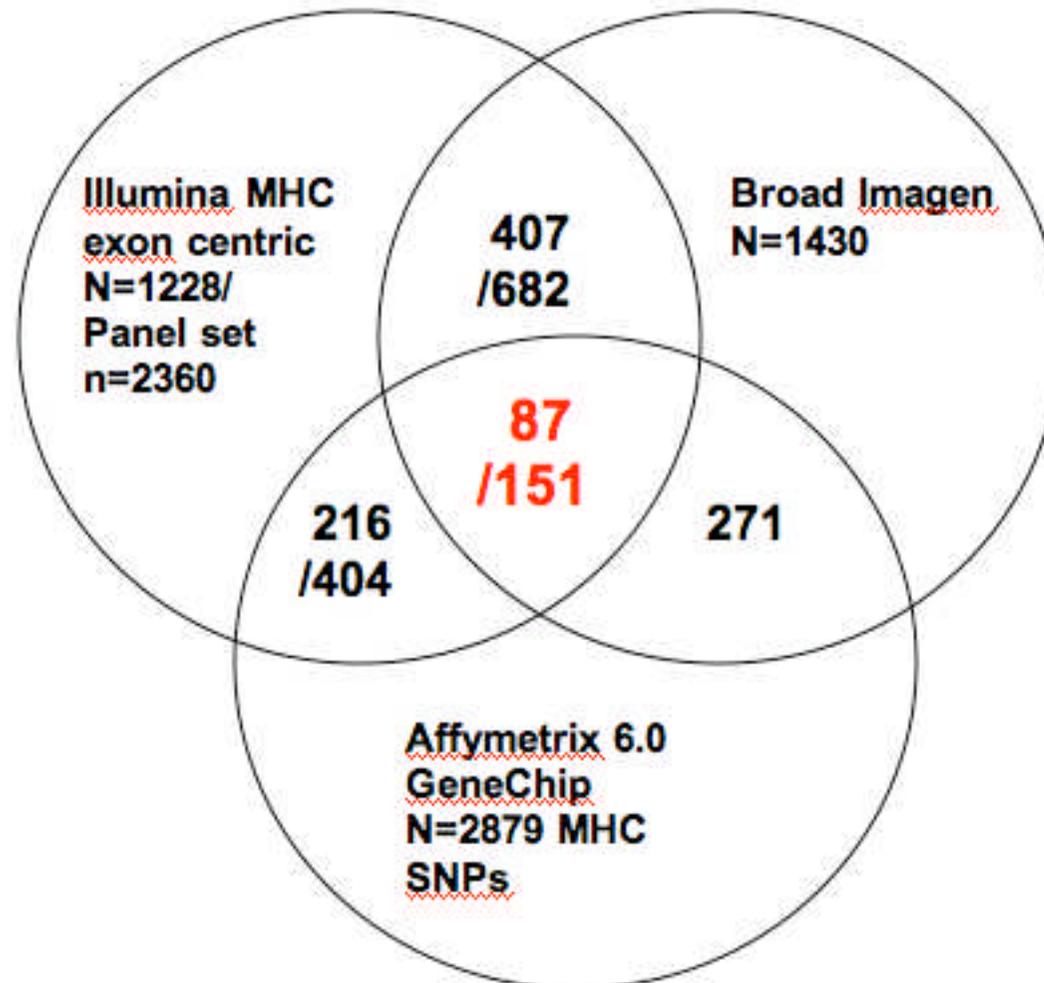
BISC: A real world example

HLA SNPs Consortium Project

Logistics:

- Five Research Teams Collaborating
- Data Repository
- Multiple SNP genotype assay platforms are used:
 - Illumina MHC exon centric
 - Affymetrix 6.0 GeneChip
 - Custom Illumina array
- SNP Allele Reporting
 - Allele Reporting is different for different platforms:
“Top vs Bottom Strand”
- Generate genetic analysis data set (.ped and .info)

HLA SNPs in Common Across Platforms Used By Some Research Teams





“Top” vs. “Bottom” SNP Allele Reporting

dbSNP ...AGCTCTGA[C/A]GTCATAGC...

“Top” ...AGCTCTGA[C/A]GTCATAGC...

SNP assay designed for “Top” strand reports alleles as C or A.

“Bottom” ...TCGAGACT[G/T]CAGTATCG

SNP assay designed for “Bottom ” strand reports alleles as G or T.

Sample ID	SNP ID	Assay Platform	Allele A	Allele B
6773	rs5678	Illumina MHC	A	A
DU367	rs5678	Affymetrix GeneChip 6	C	A
PHA_998	rs5678	HLA_ILL_56	G	T

SNP results collected from multiple assay platforms may look like tetra-allelic results when only bi-allelic results are expected.



IMMUNOLOGY DATABASE
AND ANALYSIS PORTAL

BISC: A real world example

SNPs Genotype Harmonization

- Require ImmPort genotype data include description of assay platform
- Compare the flanking sequence of each SNP on a genotype assay platform with dbSNP flanking sequence
 - Multiple dbSNP versions may need to be evaluated
- Determine whether the genotype assay platform uses “Top” or “Bottom” dbSNP strand
- Recompute SNP genotyping results allele values to use dbSNP “Top” strand value



IMMUNOLOGY DATABASE
AND ANALYSIS PORTAL

BISC: A real world example

“SNPs in Common from Multiple Platforms” Status

- Harmonization of results from three research teams completed
- Genetic analysis data sets were created by BISC team and distributed to research teams
- HLA and SNP haplotype map under development
(BISC HLA QC process was used to confirm consistent HLA results)
- Harmonization code will be integrated into ImmPort 2.4 .ped generation utility



IMMUNOLOGY DATABASE
AND ANALYSIS PORTAL

ImmPort 2.4 .ped Generation Enhancements

- Facilitate search and filtering of large numbers of subjects and SNPs
- Enable user assignment of affection status to large numbers of subjects
- Allow SNP results from multiple platforms to be included in .ped file
- Harmonize SNP results from multiple platforms
- Generate SNP genome-wide unique chromosome coordinates



ImmPort 2.4 .ped file Generation Work Flow

Choose Genetic Analysis from Log in Home page

Send Data

- Templates
- Instructions
- History
- Tools

Find Data

- Public Sources
 - Gene
 - SNP
 - MHC
 - Protein
 - Pathway
- ImmPort Sources
 - Subjects
 - Samples
 - Protocols
 - Results
- My Lists

Analyze Data

- tagSNP
- Gene Expression
- Genetic Analysis
- Correlate Gene Expression and Genotype
- Genome Browser



ImmPort 2.4 Genetic Analysis Home Page

tagSNP

Submit Task
Analysis Results

Gene Expression

User Guide
Submit Task
GDS Results
Single Analysis
Results
Pipeline Analysis
Results

Genetic Analysis

User Guide
Submit Task
Analysis Results
Create Linkage and
marker info file

The ImmPort Genetic Analysis Tools create genetic analysis data sets (.ped and .info format linkage files) and analyzes genotype data using the [Haploview](#) software developed by Mark Daly's laboratory at the Broad Institute.

This software is used to perform

- LD & haplotype block analysis
- haplotype population frequency estimation
- single SNP and haplotype association tests
- permutation testing for association significance

1. Please choose your Genetic Analysis Workflow:

- a. Create a .ped/.info
- b. Use an ImmPort .ped/.info in Haploview
- c. Use your own .ped/.info in Haploview

2. Where do you want to store Your Linkage File(s)?

Research Project(s)

ImmPort Ontology Development

- ImmPort Ontology Development
- Project for Version 2.1.1
- Integrative Genotyping Analysis in the Immunology Database and Analysis Portal
- Public Gene Expression Data
- Regression Test Project for Version 2.1.1
- Regression testing for ImmPort Version 2.1.2
- ImmPort Testscript Version 2.2
- ImmPort Testscript #2 Version 2.2
- Regression testing for ImmPort Version 2.3
- Regression testing for ImmPort Version 2.3 - Beatrice



Assign SUBJECT lists to affection status

ADMINISTRATION

SUBMIT DATA

RESEARCH DATA

SEARCH

ANALYSIS

ONTOLOGY

Creating .ped / .info Linkage Files

Assign Subjects to Affected and Unaffected

Validate Subjects for Linkage File Generation

Choose Genotype Platforms

Choose SNPs to Include

Review Generated Linkage File Results

tagSNP

Submit Task
Analysis Results

Gene Expression

User Guide
Submit Task
GDS Results
Single Analysis Results
Pipeline Analysis Results

Genetic Analysis

User Guide
Submit Task
Analysis Results
Create Linkage and marker info file

Choose one or more SUBJECT lists to assign to an Affection Status

The .ped file supports two affection status values: Affected and Unaffected.

Affection Status

Affected (Case)

	Use in Analysis	View Detail
	List Name	Description
<input type="checkbox"/>	SUBs for case	donor
<input type="checkbox"/>	SUBs for control	patient
<input type="checkbox"/>	Some more Subs	healthy

Do you need to create lists?
Search here for SUBJECTS

Unaffected (Control)

	Use in Analysis	View Detail
	List Name	Description
<input type="checkbox"/>	SUBs for case	donor
<input type="checkbox"/>	SUBs for control	patient
<input type="checkbox"/>	Some more Subs	healthy

Or use the Affection Status Assigned to the Subject

Validate that Subjects you selected have all of the data they need to create a .ped file

What's this?

The subjects will be checked to see that there are genotype results and genotype assay platforms associated with the subjects.

Adv Search link out if you need to make SUBJECT lists

Advanced Search - Selected Return Data Type : Subject

Please Choose Data Type to Find : Results per page :

Advanced Search Attribute Tree

- Subject
 - Species
 - Saved List
- Demography
 - Gender
 - Ethnicity
 - Race
 - Population Name
 - Religion
 - Age at Enrollment

Find the SUBJECTs of interest to you

a. Filter the search results by selecting an item from the lists to the left and entering the search terms.

b. Group Search Results if you wish.

c. Save your SUBJECTs to a list from the Search Result. You will need at least two lists for creating .ped .info file.

Page 1 of 13

Sub Org	Accession	User-Defined	Species	Race/Strain	Gender	Affection Pher	Upload Ticket	Project Title
<input checked="" type="checkbox"/>	SUB00010	Patient1234	Homo sapiens	Asian	male		Patrick_20060601_10	ImmPort Ontology Dev
<input checked="" type="checkbox"/>	SUB00022	Patient1234	Homo sapiens	Asian	male		vyotishah_20060601_2	B cell development an
<input checked="" type="checkbox"/>	SUB00071	Patient1234	Homo sapiens	Asian	male		vyotishah_20060614_5	B cell development an
<input checked="" type="checkbox"/>	SUB00095	Patient1234	Homo sapiens	Asian	male		vyotishah_20060622_5	B cell development an
<input checked="" type="checkbox"/>	SUB00096	Patient1234	Homo sapiens	Asian	male		scheuri_20060623_95	ImmPort Ontology Dev
<input checked="" type="checkbox"/>	SUB00127	Subject_1	9606		female		Patrick_20060623_11	ImmPort Ontology Dev
<input checked="" type="checkbox"/>	SUB00128	Subject_2	9606		male		Patrick_20060623_11	ImmPort Ontology Dev
<input checked="" type="checkbox"/>	SUB00129	Subject_3	9606		female		Patrick_20060623_11	ImmPort Ontology Dev
<input type="checkbox"/>	SUB00130	Subject_4	9606		male		Patrick_20060623_11	ImmPort Ontology Dev

Select a project to save the list *

List Name*



Assign SUBJECT lists to affection status

ADMINISTRATION

SUBMIT DATA

RESEARCH DATA

SEARCH

ANALYSIS

ONTOLOGY

Creating .ped / .info Linkage Files

Assign Subjects to Affected and Unaffected

Validate Subjects for Linkage File Generation

Choose Genotype Platforms

Choose SNPs to Include

Review Generated Linkage File Results

tagSNP

Submit Task
Analysis Results

Gene Expression

User Guide
Submit Task
GDS Results
Single Analysis
Results
Pipeline Analysis
Results

Genetic Analysis

User Guide
Submit Task
Analysis Results
Create Linkage and
marker info file

Choose one or more SUBJECT lists to assign to an Affection Status

The .ped file supports two affection status values: Affected and Unaffected.

Affection Status

Affected (Case)

	Use in Analysis	View Detail
	List Name	Description
<input type="checkbox"/>	SUBs for case	donor
<input type="checkbox"/>	SUBs for control	patient
<input type="checkbox"/>	Some more Subs	healthy

Do you need to create lists?
Search here for SUBJECTS

Unaffected (Control)

	Use in Analysis	View Detail
	List Name	Description
<input type="checkbox"/>	SUBs for case	donor
<input type="checkbox"/>	SUBs for control	patient
<input type="checkbox"/>	Some more Subs	healthy

Or use the Affection Status Assigned to the Subject

Validate that Subjects you selected have all of the data they need to create a .ped file

What's this?

The subjects will be checked to see that there are genotype results and genotype assay platforms associated with the subjects.

Validation results of SUBJECT for determining if they meet .ped file criteria

ped file generation criteria
Subjects must be linked to a single genotype assay platform and to genotype results.

Creating .ped / .info

- Find Subjects
- Assign Subjects to Affected and Unaffected
- Validate Subjects for Linkage File Generation
- Choose Genotype Platforms

124 Subjects meet all .ped file criteria

Save Subjects to a List

45 Subjects do not meet all .ped file criteria

Save Subjects to a List

There are 3 Genotype Platforms linked to your Subjects. Please select the platform(s) you wish to use:

- Affy 800k
- LII885
- ILL MHC exon centric

Subject ID	Description	Genotype Platform	Number of with Genot Results
<input type="checkbox"/> D13445	donor	Affy 800k	789,445
<input type="checkbox"/> P5883	patient	Affy 800k	790,344
<input type="checkbox"/> BIS345	healthy	LII885	78,999
<input type="checkbox"/> Y677	case	ILL MHC exon centric	7,986

Page 1 of 1	View Details	Save Items	Save All
Subject ID	Description	Genotype Platform	Number of SNP Genotype Resu
<input type="checkbox"/> D13445	donor	Aff 800k Aff250k Sty	789,445
<input type="checkbox"/> P5883	patient	NA	790,344
<input type="checkbox"/> BIS345	healthy	LII885	NA

If the user selects only one platform

Do you want to use all of the SNP on the platform or choose from a list?

Use All SNPs on the Platform

- 45 SNPs I like
IL-2, pathway SNPs
SNP, SNAP, SNUP

Do you need to create SNP lists?
Search here for SNPs

Generate .ped and .info files

If the user selects more than one platform

Do you want to use all of the SNP that are in common between the platforms or choose from a list?

Use SNPs in Common Between the Platforms

- 45 SNPs I like
IL-2, pathway SNPs
SNP, SNAP, SNUP

Do you need to create SNP lists?
Search here for SNPs

Generate .ped and .info files

Advanced Search utility if you need to make a SNP list

Advanced Search - Selected Return Data Type : Subject

Please Choose Data Type to Find : Results per page :

- Advanced Search Attribute Tree
- Genes
 - Symbol
 - Entrez ID
 - Proteins
 - Symbol
 - UniProt ID
 - SNPs
 - rs ID
 - Gene Symbol
 - MHC Allele
 - Locus Symbol
 - Allele Name
 - Pathways
 - Pathway Name
 - Pathway ID

Find SNPs of interest to you

a. Filter the search results by selecting an item from the lists to the left and entering the search terms.

b. Group Search Results if you wish.

Page 1 of 13 | View Details | Save Items | Save All

Sub Org Accession	Gene Symbol
<input checked="" type="checkbox"/> rs123	IL-12
<input checked="" type="checkbox"/> rs9988444	IL-12
<input checked="" type="checkbox"/> rs27499	IL21
<input checked="" type="checkbox"/> rs19955	IL-112
<input checked="" type="checkbox"/> rs888440	GAPC
<input checked="" type="checkbox"/> rs9911994	CBBE
<input checked="" type="checkbox"/> rs993393	UDS
<input checked="" type="checkbox"/> rs09973	PWW
<input type="checkbox"/> rs80021	UNNA

c. Save your SNPs to a list from the Search Result and/or the Grouping.

Select a project to save the list *

List Name*



IMMUNOLOGY DATABASE
AND ANALYSIS PORTAL

Questions and Comments?

Thank you for your attention.

helpdesk@import.org