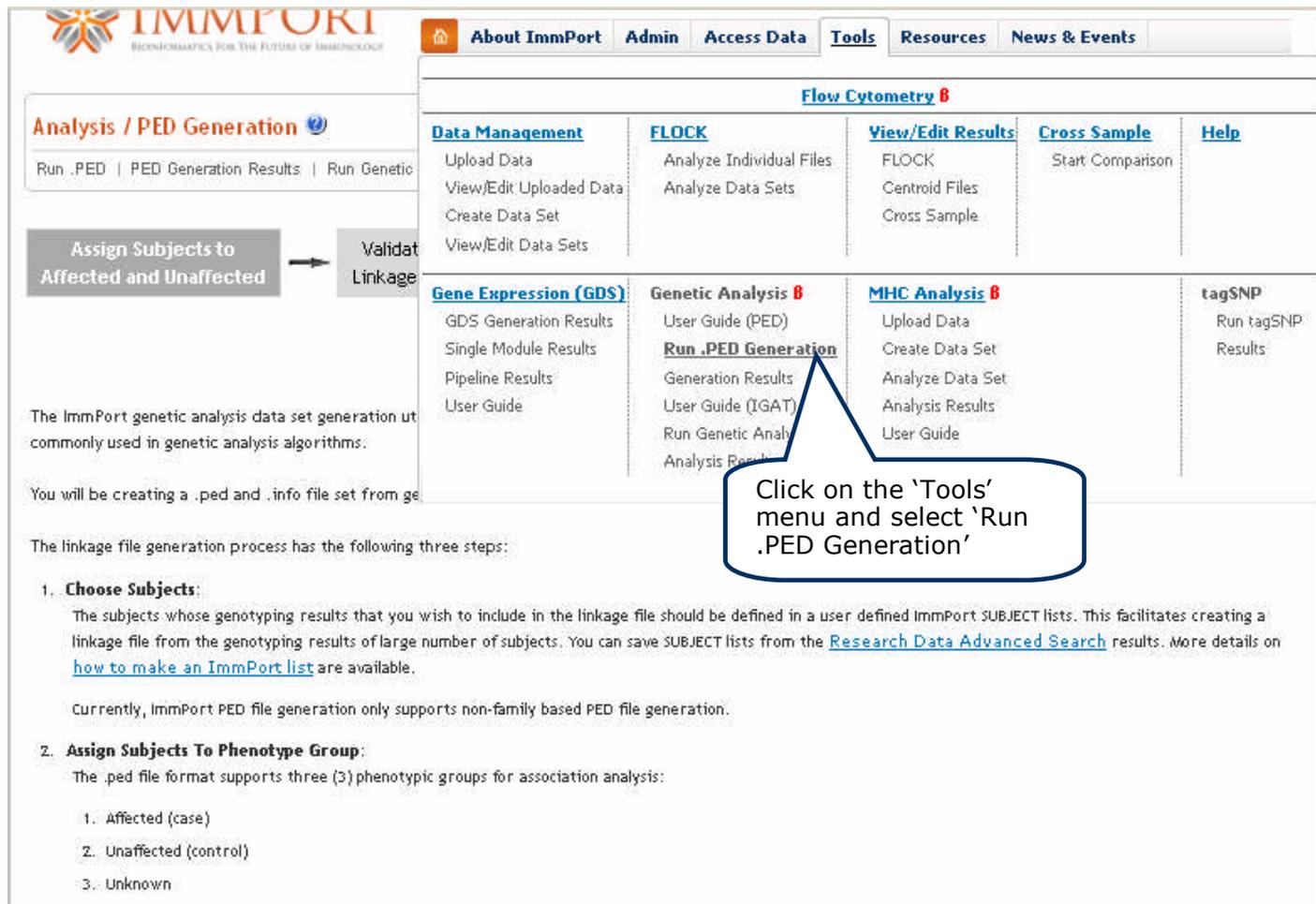


Introduction

The .ped file is the shorthand, commonly used name for the **LINKAGE 'Pre-makeped'** file format. In ImmPort, .ped and .info files are used with the Haploview genetic analysis suite and are inputs to other genetic analysis tools. The Haploview suite enforces certain data simplification strategies to enable the analysis to proceed. This requires that a .ped/.info file generation tool validate input requests to ensure all criteria are met. The current changes in ImmPort is to enhance the .ped file generation utility by leveraging lists use (subjects, genetic markers) to make the process more intuitive and scalable to large data sets.



The screenshot shows the ImmPort web interface. The top navigation bar includes 'About ImmPort', 'Admin', 'Access Data', 'Tools', 'Resources', and 'News & Events'. The 'Tools' menu is expanded, showing several categories: 'Flow Cytometry', 'Data Management', 'Gene Expression (GDS)', 'Genetic Analysis', 'MHC Analysis', and 'tagSNP'. Under 'Genetic Analysis', the 'Run .PED Generation' option is highlighted with a blue callout box that says 'Click on the 'Tools' menu and select 'Run .PED Generation''. The main content area on the left shows a workflow: 'Assign Subjects to Affected and Unaffected' followed by 'Validate Linkage'. Below this, there is a section titled 'Analysis / PED Generation' with a sub-header 'Run .PED | PED Generation Results | Run Genetic'. The text below explains that the ImmPort genetic analysis data set generation utility is commonly used in genetic analysis algorithms and that the linkage file generation process has three steps: 1. Choose Subjects, 2. Assign Subjects To Phenotype Group, and 3. Unknown.

.ped and .info File Generation

When the user clicks on Ped Generation link from the Analysis module shown in Slide 2, the Ped Generation main page is as shown below:



PED Generation 

Beta Release

Assign Subjects to Affected and Unaffected | Validate Subjects for Linkage File Generation | Choose SNPs to include | Review Generated Linkage File

Selected Projects | **Change Projects**

The ImmPort genetic analysis data set generation utility allows you to create the .ped (pedigree) and .info (information) files for the 'pre MAKEPED' [linkage file format](#) commonly used in genetic analysis algorithms.

You will be creating a .ped and .info file set from genotype data stored in ImmPort. The linkage file generation process has the following three steps:

- 1. Choose Subjects:**
The subjects whose genotyping results that you wish to include in the linkage file should be defined in a user defined ImmPort SUBJECT lists. This facilitates creating a linkage file from the genotyping results of large number of subjects. You can save SUBJECT lists from the [Research Data Advanced Search](#) results. More details on [how to make an ImmPort list](#) are available.

Currently, ImmPort PED file generation only supports non-family based PED file generation.
- 2. Assign Subjects To Phenotype Group:**
The .ped file format supports three (3) phenotypic groups for association analysis:
Affected (case)
Unaffected (control) and
Unknown

You may assign a SUBJECT list to a phenotype group by selecting a radio button in the appropriate column. One or more SUBJECT lists may be assigned to a phenotype group. At least one SUBJECT list must be assigned to at least one phenotype group to proceed. It is not necessary for a list to be assigned to each phenotype group.
- 3. Validate Subjects:**
When you have assigned all of the lists you wish to use, click the 'Validate' button. This will evaluate whether any of the subjects in the lists is assigned to a phenotype group, then the PED generation tool will proceed. If not, you will be notified. [Click here for more details on the validation process.](#)

Limitations on PED Generation:
For PED generation, a user can determine the number of subjects that can be processed. We have successfully tested up to 50 millions data-points. ImmPort currently supports generating non-family based PED files.

Newly submitted data will be available for PED generation the next day after submission.

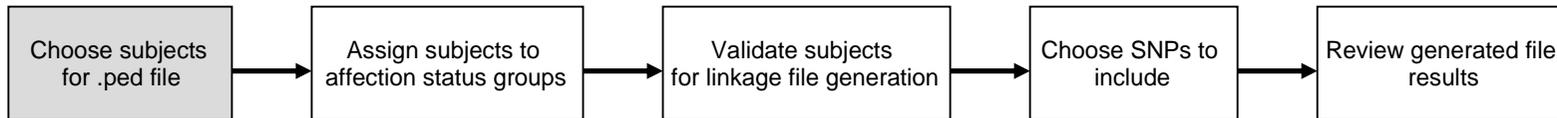
Button Tips:
During the PED generation process, the subject list choices remains the same unless you change the SUBJECT list selection on this page or 'Clear' all SUBJECT lists. The Clear option removes all SUBJECT list choices so that you can start the process over again. To deselect a SUBJECT list from a category, just click on that select again. You can change the category into which a SUBJECT list selected by merely clicking on the phenotype group you wish the list to reside in.

Subject List Name	Description	Case	Control	Unknown
20090710 subject SUB1455	1 item - SUB1455	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20090720 AS subject CP25	total - 81	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20090720 SFVT Sub & Exp	item - 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



.ped and .info File Generation

Choose subjects for .ped file



Research Advanced Search

Using Advanced Search:

1. Choose the type of data you want to find (e.g. subjects).
2. Choose a folder to refine your search. It can be a subject or a project.
3. Click on the data feature(s) you want to search.
4. Enter search terms for one or more features.

Advanced Search - Selected Return Data Type : Subject

Tip: You can select a range of rows by holding down the 'Shift' key.
Selected items: SUB1, SUB10, SUB11, SUB12, SUB13, SUB14

Select the arrow to the attribute tree for editing your search criteria

Select a project to save a list and give it a name and description

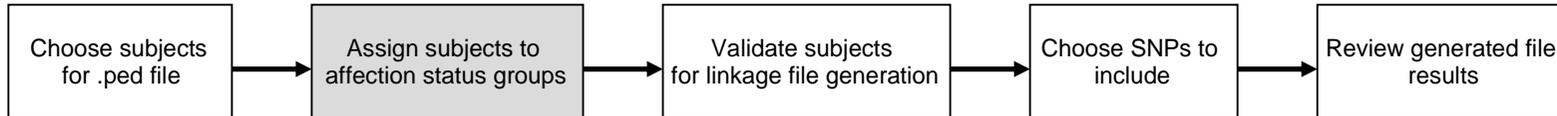
Save selected items

Select any or all the boxes for detailed or saving to a list

Sub Org Accession	User-Defined ID	Species	Race/Strain	Gender	Description	Project Title
<input checked="" type="checkbox"/> SUB1	ELISPOT Subject			Female	"Acute systemic lupus erythematosus with ne	TESTING: Bioinformatics Integrati
<input checked="" type="checkbox"/> SUB10	Subject_Exp13_38		African Amer	Male	This subject is the Father of Subject_Exp13_;	TESTING: Bioinformatics Integrati
<input checked="" type="checkbox"/> SUB11	Subject_Exp13_4073	Homo sapiens	Native Hawaiians or P	Female	This subject is the Mother of Subject_Exp13_;	TESTING: Bioinformatics Integrati
<input checked="" type="checkbox"/> SUB12	Subject_Exp13_3546	Homo sapiens	Black or African Amer	Female	This subject is the Offspring of the parents -{	TESTING: Bioinformatics Integrati
<input checked="" type="checkbox"/> SUB13	Subject_Exp13_227	Homo sapiens	Native Hawaiians or P	Male	This subject is the Father of Subject_Exp13_;	TESTING: Bioinformatics Integrati
<input checked="" type="checkbox"/> SUB14			Black or African Amer	Female	This subject is the Mother of Subject_Exp13_;	TESTING: Bioinformatics Integrati
<input type="checkbox"/> SUB15			Black or African Amer	Male	This subject is the Offspring of the parents -{	TESTING: Bioinformatics Integrati
<input type="checkbox"/> SUB1576			African American	Male	"Allogeneic transplant uses donated cells from	TESTING: Bioinformatics Integrati
<input type="checkbox"/> SUB1577	HLA Subject C2	Homo sapiens	Native Hawaii or Pac	Male	"Allogeneic transplant uses donated cells from	TESTING: Bioinformatics Integrati
<input type="checkbox"/> SUB1578	HLA Subject C3	Homo sapiens	Black	Female	"Allogeneic transplant uses donated cells from	TESTING: Bioinformatics Integrati



Assigning subjects to affection status groups



The .ped file column 'Affection Status' provides a means to stratify subjects (SUBs) into one of three groups: Affected, Unaffected, and Unknown for genetic association tests.

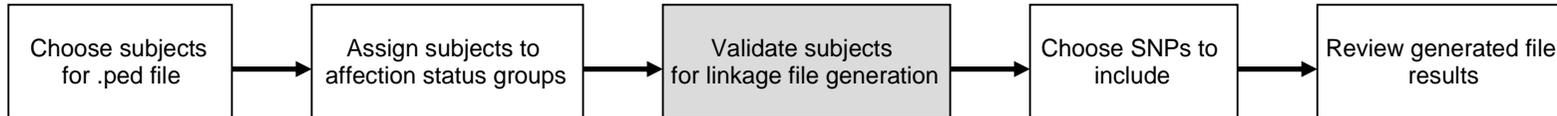
The user must have the means to decide which criteria they wish to use for assigning SUBs to the Affection Status values. The ImmPort subject metadata includes provisions for describing affection status, but this only partially satisfies the need to allow the user to choose the criteria by which to group SUBs.

A more robust strategy is to allow the user to assign SUB lists to the Affection Status values. As noted above, the Advanced Search and Grouping functions allow the users to define SUB lists. The system will need to allow the user to choose from their SUB lists and assign to each of the Affection Status values. This is intended to allow the user to apply the affection status value stored in the SUB metadata or apply any other metadata criteria to define affected, vs. unaffected SUBs.

No database metadata content needs to be changed. The Affection Status assignments are only applied to the .ped file content. This feature may lead to deprecation of the Affection Status attribute in the metadata schema as the user can dynamically assign affection status to a SUB upon .ped file creation.



Validation of records for .ped file generation criteria



PED Generation

Beta Release

If the list does not exist, click Advanced Research Data Search to create a list – Slide 3

The ImmPort genetic analysis data set generation utility allows you to create the .ped and .info (information) files for the 'pre MAKEPED' [linkage file format](#) commonly used in genetic analysis algorithms.

You will be creating a .ped and .info file set from genotype data at ImmPort. The linkage file generation process has the following three steps:

1. Choose Subjects:

The subjects whose genotyping results that you wish to include in the linkage file should be defined in a user defined ImmPort SUBJECT lists. This facilitates creating a linkage file from the genotyping results of large number of subjects. You can save SUBJECT lists from the [Research Data Advanced Search](#) results. More details on [how to make an ImmPort list](#) are available.

Currently, ImmPort PED file generation only supports non-family based data.

Click 'validate' to validate subjects for .ped generation

2. Assign Subjects To Phenotype Group:

The .ped file format supports three (3) phenotypic groups for assignment: Affected (case), Unaffected (control) and Unknown.

You may assign a SUBJECT list to a phenotype group by selecting a radio button in the appropriate column. One or more SUBJECT lists may be assigned to a phenotype group.

At least one SUBJECT list must be assigned to at least one phenotype group to proceed. It is not necessary for a list to be assigned to each phenotype group.

3. Validate Subjects:

When you have assigned all of the lists you wish to use to their appropriate group, please click the 'Validate' button. This will evaluate whether any of the subjects in the lists is assigned to more than one phenotype group. If no subjects are assigned to more than one phenotype group, then the PED generation tool evaluates whether subjects in each group meet the [PED file generation criteria](#) and which do not.

Limitations on PED Generation:

For PED generation, a user can determine the number of subjects and SNPs to generate a PED file. For the current system, we have successfully tested up to 50 millions data-points. Data points are defined by the formula = #subjects x #SNPs. ImmPort currently supports generating non-family based PED files.

Newly submitted data will be available for PED generation on the next day after submission.

Button Tips:

During the PED generation process, the set of list choices remains the same until you click the 'Clear' button. The 'Clear' option removes all SUBJECT list choices so you can start over. To deselect a SUBJECT list from a category, just click on that select again. You can change your selection on this page or over again. To deselect a SUBJECT list selected by merely clicking on the phenotype group you wish the list to reside in.

Click 'clear' to start the validation process over

Page 1 of 2 | Validate | Clear | Displaying 1 - 25 of 46.

Subject List Name	Description	Case	Control	Unknown
20090710_subject_SUB1455	1 item - SUB1455	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20090720_AS_subject_CP25	total - 81	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20090720_SFVT_Sub_&Exp	item - 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Display of validation results

The user will have an option view the SUB records that meet all .ped generation criteria and those that do not. This would involve creating a summary result table with columns that are unique to .ped file generation. The SUB summary result currently shows attributes intrinsic to SUBs. For consistency, the user should be able to see details of the SUBs that met all criteria.



Subject Validation Results for Linkage File

Subject Data Issues Tutorial

Note: When using **'Show Details'** to view the subject details for a subject category, the result table may not be able to display subject data if there are too many subjects to display. Select the **Save All** option on the table to save the subjects as a list. You may use the list in Advanced Search to reduce the number of subject records.

Affected / Case

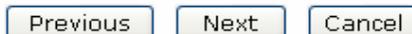
3 Subjects meet all .ped file criteria. [Show Details](#)
 101 Subjects do not meet all .ped file criteria. [Show Details](#)

Unaffected / Control

0 Subjects meet all .ped file criteria.
 0 Subjects do not meet all .ped file criteria.

Unknown

0 Subjects meet all .ped file criteria.
 0 Subjects do not meet all .ped file criteria.



Click to show details of the validated results. Click 'Next' to go to the SNP selection page



Detailed view of validation results for the case and control

Affected/Case Subjects that meet all .ped file criteria

Extra subject data is available using the optional display columns. This data include: species, race/strain, gender, affection phenotype, affection status, project title, and submit date. The data is sorted by ImmPort subject accession.

User can 'view details', 'save items', or 'save all' items selected

Page 1 of 1							View Details Save Items Save All		Displaying 1 - 6 of 6	
	Sub Org Acce	Sub User Id	Subject Desc	GenoType Platform	No. of SNPs with Genotype Results	Exp Sample User ID				
<input type="checkbox"/>	SUB215828	Subject2	Father of affe	Demonstration Illumina panel 1	12	BeadStudio_FinalReport_12se				
<input type="checkbox"/>	SUB215831	Subject5	Father of affe	Demonstration Illumina panel 1	12	BeadStudio_FinalReport_12se				
<input type="checkbox"/>	SUB215832	Subject6	Mother of affe	Demonstration Illumina panel 1	12	BeadStudio_FinalReport_12se				

Unaffected/Control Subjects that meet all .ped file criteria

Extra subject data is available using the optional display columns. This data include: species, race/strain, gender, affection phenotype, affection status, project title, and submit date. The data is sorted by ImmPort subject accession.

Page 1 of 1							View Details Save Items Save All		Displaying 1 - 6 of 6	
	Sub Org Acce	Sub User Id	Subject Desc	GenoType Platform	No. of SNPs with Genotype Results	Exp Sample User ID				
<input type="checkbox"/>	SUB215827	Subject1	"Affected child	Demonstration Illumina panel 1	12	BeadStudio_FinalReport_12se				
<input type="checkbox"/>	SUB215829	Subject3	Mother of affe	Demonstration Illumina panel 1	12	BeadStudio_FinalReport_12se				
<input type="checkbox"/>	SUB215830	Subject4	"Affected child	Demonstration Illumina panel 1	12	BeadStudio_FinalReport_12se				



.ped and .info File Generation

Validation results do not meet the all .ped generation criteria

Assign Subjects to Affected and Unaffected | **Validate Subjects for Linkage File Generation** | Choose SNPs to include | Review Generated Linkage File Results

Affected / Case

0 Subjects meet all .ped file criteria.
 8 Subjects do not meet all .ped file criteria. [Show Details](#)

Unaffected / Control

0 Subjects meet all .ped file criteria.
 0 Subjects do not meet all .ped file criteria.

Unknown

0 Subjects
 0 Subjects

Affected/Case Subjects that do not meet .ped file criteria

Subject Data Issues Tutorial

[Previous](#)

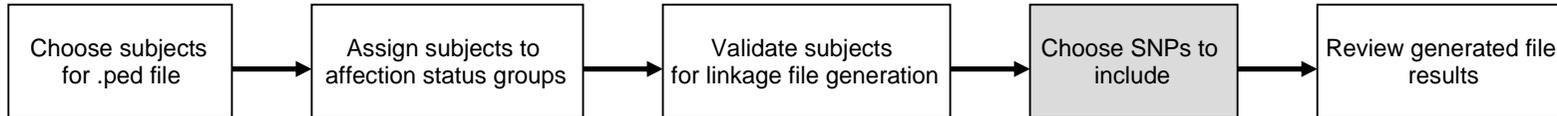
Extra subject data is available using the optional display columns. This data include: species, race/strain, gender, affection phenotype, affection status, project title, and submit date. The data is sorted by ImmPort subject accession.

Note the GenoTyping and Subject Related Errors and click on the Tutorial for a description on these errors

Sub Org Accession	Sub User Id	Subject Desc	GenoTyping Errors	Subject Related Errors
<input type="checkbox"/> SUB03032	11		Unknown Geno-Typing Reagent/Resu	Missing Experiment Samples
<input type="checkbox"/> SUB03036	4		Unknown Geno-Typing Reagent/Resu	Missing Biological Samples
<input type="checkbox"/> SUB122969	Subject_HLAEp2_1952	This subject is the	Unknown Geno-Typing Reagent/Resu	Missing Experiment Samples
<input type="checkbox"/> SUB128182	Subject_HLAEp3_2668	This subject is the	Unknown Geno-Typing Reagent/Resu	Missing Experiment Samples
<input type="checkbox"/> SUB152366	Study_Pat-451	CNTLSING: Contr	Unknown Geno-Typing Reagent/Resu	Missing Experiment Samples

.ped and .info File Generation

Selected subjects that meet .ped generation criteria are linked to ESs



The subjects you selected and that meet all .ped file generation criteria are linked to the result sets(experiment samples) listed below. The genotype assay platforms linked to the result sets are also listed. Please select one or more result sets to include in the .ped file.

Exp UserID (Affy)/ Exp Sample UserID (Illumina)	Genotype Platform	Number of SNPs with Genotype Results *
Illumina_FHCRC_MNDP_final_repo	Illumina MHC-Exon-Centric Panel , Illumina MHC-Exon-Centric Panel	1199

Confirmation of genotype platform, subjects, and scope of SNPs

The values in the column "Number of SNPs with Genotype Results" are computed using the total record counts for the experiment samples in an experiment. As a result, the numbers may not match the dbSNP rs accession counts provided below.
The number of dbSNP rs accessions that are found in common between all of the result sets is : 1201
The total number of dbSNP rs accessions that are found in all the result sets is : 1201

Indicate dataset Name and Description and click 'Generate .ped and .info files'

Fields marked with an asterisk * are required.

Generate the .ped file from: *

SNPs in common between all selected results

All SNPs present in the result sets

If you want to use an Import SNP list to further select the SNPs to include in the .ped file, please choose an Import SNP list to select the SNPs. If you need to create the desired SNP list, please go to [Advanced Reference Search](#). On return from the advance reference query page you will need to refresh this page so that the SNP list will appear in the SNP lists menu below:

NO SNP LIST

OR

If you want to use your own SNP list to further select the SNPs to include in the .ped file, please select your file:

Select a research project below to save the PED results to:

Packages3

Dataset Name: *

Dataset Description: *

Analysis Type: *

Non-family based

The family based linkage file will include the pedigree information available for each subject

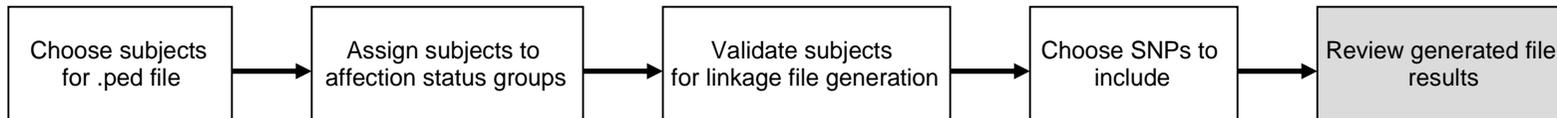
The page at <https://www.import.net> says:

You are about to generate a ped file. This may take a few minutes.



.ped and .info File Generation

Review generated linkage file results



PED Generation

Beta Release



Your PED Generator request was submitted successfully. This generation determined that you requested 1201 snps and 3 subjects using a single genotype platform reagent ESR00038. Your analysis task ID is 4236. Click the "Show PED Converter Results" button to proceed to the PED generator History page. Once the task you requested is completed, you will be able to download your:

- '.ped' (linkage) file
- '.info' (marker) file
- missing markers file (Markers that did not satisfy all criteria for .ped file generation. Please see user guide for further details.).

Previous

Show Ped Converter Results

Click the ped converter results to obtain the .ped and .info results

Once the user has obtained linkage file results, the next step is to download the associated files from the IGAT Analysis results page.

OR

The user can also Submit Task under Genetic Analysis module



Use the linkage files generated in the Genetic Analysis module

IGAT Analysis History

This page is a list of the tasks you have submitted to the ImmPort Genetic Analysis Toolset. The history list can show up to 50 tasks. If the number exceeds 50, the oldest tasks will be dropped off the list. Click on a task ID to view more information about the task or to see the results when the task has completed execution.

127 items found, displaying 1 to 20. [First/Prev] [1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [7](#) [Next/Last]

Task Id	Start Date	End Date	Task Type	Data Set Accession/Name	Dataset Source	Algorithm	Status
4236	05/01/2009 09:09:11	05/01/2009 09:09:19	Create_PED_File	HDS619 / 2.5 test	import_created	n/a	Completed
4231	04/30/2009 17:29:33	04/30/2009 17:31:48	Run_Haploview	HDS544 / TS020509.1	import_created	Haploview	Completed
4227	04/30/2009 13:58:33	04/30/2009 14:06:06	Run_Haploview	HDS544 / TS020509.1	import_created	Haploview	Completed

Click on the Task Id to view results of completed tasks. This link takes user to the ImmPort Genetic Analysis Tool (IGAT)