SKDM HLA tool:

SKDM tool helps to investigate HLA associations, and automate their analysis in the context of a case–control design through complex computation.

The User-Interface:

A Java UI helps provide speed and portability to the analysis. The UI provides space for copying and pasting HLA typing for CASE and another space for Control population.

The SKDM HLA Tool can test for HLA allele differences between two populations and perform amino acid analysis by retrieving amino acid sequences. Once primary associations are identified, the program examines zygosity and tests for strongest association, interaction and linkage disequilibrium among amino acid epitopes of the same HLA molecule or between HLA isotypes. A summary of the analysis is output in plain language.

1. Getting Started

This is the upgraded version of the application. The older version is here: http://sourceforge.net/projects/skdm/

1.1 Availability

The upgraded version of SKDM is available at https://github.com/chopdgd/CHOP_SKDM.
You only need the main executable "SKDM_2.1.jar" under the "dist" folder to run this program. In order to execute the program, simply double click on the icon or on the filename, SKDM_2.1.jar. If you cannot execute this file for any reason, you may need to install the latest version of Java on your machine. To do that, go to the following website and download a free version of Java:
http://www.java.com/en/download/

1.2 Quick Start

On the top input panel, type (as in the table below)

```
A
0101
```

'A' is the HLA locus and '0101' is the allele. Then click on "R U N …"
You should get a new tab on the output panel, labeled "Results" with the following information:

```
SKDM HLA Tools by S.Kanterakis-D.Monos (c)2007
Timestamp: Wed Dec 05 14:13:39 EST 2007


CASES
=====

HLA-A summary
Allele  Pop Freq              Allele Freq
0101    1  100.00%     2  100.00%
1       unique alleles total.
1       samples total.

Alignment of HLA-A alleles
Allele      .  10|    .  20|    .  30|    .  40|    .  50|    .  60|    …
0101   GSHSMRYFFTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQKMEPRAPWIEQEGPEYWDQE …


Analysis completed in 0.547 seconds.
*EOF****************************************************
```

This is the simplest way to run the program. In the following pages you will learn how to do a case-control analysis and interpret the program's output.

1.3 Machine Requirements

SKDM was programmed for Java version 6 or later. It was tested on Windows XP, Mac OS X and Linux 6.5, although it should run on any operating system. Computation time ranges from a few seconds to several minutes, depending on sample size and the number of HLA loci examined. A Pentium 4 machine at 2GHz (or equivalent) with 2GB of memory is recommended. A test run on 104 cases and 138 control individuals, typed at HLA-A, B, Cw, DRB1 and DQB1, on such a system was completed in 2 minutes.

2. Input

2.1 File format

SKDM can analyze sample sets that have been molecularly typed -in high or low resolution- for the HLA class I and/or class II regions. The program accepts a set of HLA alleles for any of the A, B, C, DRB1, DQB1, DQA1, DPB1 and DPA1 HLA and MICA, MICB polymorphic loci, for a list of individuals. Typical input to the program will have the following form.

```
My Population
Id      HLA-A          HLA-B          HLA-C          HLA-DRB1
03      0201    3101   5101    4002                   1407    1501
30      0201           5101    3503   12      15      1401    1601
32      0201           4402    3503   05      12      0407    1104
33      0201    3201   5101    3502   04      15      1104    1101
36      0201    2501   3901    1801   07      12      0405    0405
38      0201    3201   5101    3503   15      15      1601    1101
```

This input can be interpreted as: a list of 6 individuals from “My Population”, with HLA typing for the A, B, C and DRB1 loci. Individual “03” is missing the HLA-C typing, while individuals “30” and “32” are homozygous for HLA-A. Alternatively, an HLA allele may be added twice to denote homozygosity, as in the case of individual “36” for the DRB1 locus.

HLA typing can be in high resolution (4 digits, as in 0201) or low resolution (as for HLA-C in the above example). When low resolution typing is used, the program takes the consensus string of all possible high-resolution typings, marking polymorphic amino acid positions as unknown.

The dataset title (“My Population”) and the Id column are optional but recommended. In a case-control study, the dataset title is used to differentiate between the two populations. If no title is found, the default labels “cases” and “controls” are used. The dataset files should be tab-delimited. The best way to input data into the SKDM is by copying and pasting from a spreadsheet application (such as Excel) directly into the SKDM input panels.

2.2 Summary statistics and case-control analysis

For population summary statistics, one dataset is sufficient. SKDM will produce an allele summary and retrieve amino acid alignments.

For a case-control run, you need to input two datasets. The top input panel will usually correspond to the case dataset while the bottom panel to the controls. If both datasets are present, the program will first summarize the two datasets and then run a comparison analysis.

3. Output

SKDM’s output will generally have the following structure (Results tab):

- The first piece of output the user will see is a header with a timestamp.

```
SKDM HLA Tools by S.Kanterakis-D.Monos (c) 2007
Timestamp: Thu Dec 06 17:14:57 EST 2007
```

- Right below the header is a comparison title (for case-control runs)

```
My CASES vs My CONTROLS
```

• The output below is an allele summary for each HLA locus in each population, where the population frequency (Pop Freq) indicates the number of subjects (and the corresponding percentage), positive for the HLA A*0201 allele, whereas the allele frequency (Allele Freq) refers to the number of HLA- A*0201 alleles present in this population. The total number of samples and distinct alleles is given at the end.

```
My CASES
========

HLA-A summary
Allele  Pop Freq       Allele Freq
0201    45  43.27%     55  26.44%
…
…
23      unique alleles total.
104     samples total.
```

• The output below is the difference (Delta) in frequency between case and control alleles for a particular locus where the difference in population frequency for each allele is given. A corresponding odds ratio (OR) and a corrected p-value are also supplied. P-values are corrected by the number of distinct alleles present in cases and controls (30 in this case). This list is ordered by "Delta" to aid in visualizing the most differentially distributed alleles.

```
Delta between My CASES and My CONTROLS for locus HLA-A
Allele  Delta   p^corr  OR
0101    8.31%   1       1.58
…
…
30      alleles total.
```

• The output below is the amino acid alignment for each of the alleles of a particular HLA locus, ordered by "delta" to aid in visualizing the most differentially distributed alleles.

```
Alignment of HLA-A alleles
Delta   Allele       .  10|    .  20|    .  30|    .  40|    .  50|
8.31    0101     GSHSMRYFFTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQKMEPRAPWIEQ…
…
```

• The output below shows a list of statistically significant residues as a table denoting the alleles (Alls) where a residues is present, it's position (Pos) in the alignment, the single letter alias of the amino acid (AA), whether it is associated (Assoc) with cases (+) or controls (–), a p-value, a p-value corrected (p^corr) by the number of AA interrogated and an associated odds-ratio (OR).

HLA-DPB1 Residues

> p-value correction is 42 (= number of AA interrogated)

Alls	Pos	AA	Assoc	p-val	p^corr	OR
0101 1401 0601 1001 1301 1701 0301	8	V	+	0.01338	0.56207	5.57
0101 1301 0401 … …	36	A	-	0.02304	0.9678	0.19

• The output below shows a list of HLA pocket residues whose distribution is significantly different between the two groups with a similar format as in the residue list, except here AAs are ordered by pocket (with pocket positions indicated in brackets), and the correction is equal to the number of pocket AAs. Elements whose p-value is less than the significance threshold (here 95% or 0.05), are marked with asterisks (***).

HLA-DPB1 Pocket Residues

> p-value correction is 24 (= pocket AAs interrogated)

Pocket 1 [87,84]

Pos	AA	Assoc	p-val	p^corr	OR
84	D	+	0.0156	0.3732	5.13
87	V	+	0.0156	0.3732	5.13

Pocket 4 [13,69,76,68,72,24]

Pos	AA	Assoc	p-val	p^corr	OR
69	E	+	2.2E-4	0.0053	12.89 ***

… …

• The output below is an assessment of zygosity where for each previously identified AA, it is indicated whether a homozygote or heterozygote condition differentiates susceptibility to disease. Three tests are performed here and the p-value correction is 3.

Zygosity analysis for My CASES and My CONTROLS

> p-value correction is 3 (= zygosity tests)

Description [p-val, OR]

DPB1_E-69; homozygotes associated [0.04348, 14.14], heterozygotes associated [0.00131, 11.97],no difference in zygosity.

DPB1_DEAV-84,85,86,87; homozygotes NOT individually associated, heterozygotes associated [0.04665, 4.78],no difference in zygosity.

…

• A variety of tests for statistically significant associations are shown below, where tests for independence, difference in association, combined action, interaction and linkage disequilibrium (LD) are used to determine the strongest association amongst the list of the associated AAs. For each of the 8 tests, p-values and associated OR are given.

```
Interaction analysis for My CASES and My CONTROLS
=================================================
> p-value correction is 5 (= tests 1-5, for strongest association), 8
(for less critical tests, 6-8)
Description [test; p-val, OR]

DPB1_V36 NOT independent of DPB1_E69, DPB1_E69 NOT independent of
DPB1_V36, DO NOT interact, associations DO NOT differ, have combined
action, in LD (CASE), in LD (CTRL).
[1; 1, 1.59]   [2; 0.26417, 3] [3; 1, 2.19]  [4; 1, 4.13]   [5; 1, 0.73]
[6; 0.00209, 6.56] *** [7; 0.02144, 9] *** [8; 0.0053, 17] ***

DPB1_E69 NOT independent of DPB1_DEAV84,85,86,87, DPB1_DEAV84,85,86,87
NOT independent of DPB1_E69, DO NOT interact, associations DO NOT
differ, DO NOT have combined action, NOT in LD (CASE), NOT in LD (CTRL).
[1; 1, 1.59]   [2; 0.01215, 4.9] *** [3; 1, 1.55]   [4; 0.15164, 4.77]
[5; 1, 1.03]   [6; 0.07559, 7.57]    [7; 1, 1.06]   [8; 1, 3.29]
…
```

• A footer indicating the completion of analysis

```
Analysis completed in 1.812 seconds.
*EOF*****************************************************
```

A detailed list of all two by two contingency tables used to complete the above tests is given in the “2x2” tab.

Additionally, you may notice 4 files being created in the directory where the program was run. These files are not mandatory to run the program.

- ctrl.txt, contains the last control population which was input
- pat.txt, contains the last case population
- log.txt, lists all output produced by the last analysis
- settings.txt, saves the settings of the last run.

4. Description of Tests.

In this section, the details of the tests for association and associations in pocket positions, zygosity and interaction will be given. For each comparison, the Odd's Ratio (OR) is calculated by Haldane's modification of Woolf's method: OR = [(a + 1⁄2)(d + 1⁄2)]/[(b + 1⁄2)(c + 1⁄2)], and the significance of it's derivation from unity is estimated by Fisher's exact test.

4.1 Tests for association of polymorphic amino acids and pockets

	Number of	
Factor	Cases	Controls
+	x1	y1
–	x2	y2

	Entries of 2 x 2 Table				
Comparison	a	b	c	d	Test
Factor presence vs absence	x1	y1	x2	y2	Factor associated?
Factor presence vs absence	x1	y1	x2	y2	Pocket factor associated?

The pocket position that the program recognize are the following: HLA-A (ref 1 and 2)

HLA-B (same as HLA-A)

HLA-Cw (same as HLA-A)

HLA-DRB1 (refs 3, 4 and 5)

Pocket 1: 85, 89, 86

Pocket 4: 13, 71, 78, 70, 74, 26

Pocket 6: 9, 11,30

Pocket 7: 28, 61, 71, 47, 67

Pocket 9: 9, 60, 57, 37, 38

HLA-DQB1 (same as HLA-DRB1)

HLA-DPB1 (ref 6)

Pocket 1: 87, 84

Pocket 4: 13, 69, 76, 68, 72, 24

Pocket 6: 9, 11, 28

Pocket 7: 26, 59, 69, 45, 65

Pocket 9: 9, 58, 55, 35, 36

HLA-DQA1 (same as DRB1)

Pocket 1: 34, 46, 56, 35, 55, 57, 27

Pocket 4: 11, 65, 14

Pocket 6: 14, 68, 66, 69

Pocket 7: 68, 72

Pocket 9: 75, 76, 72, 79

HLA-DPA1 (same as DRB1)

Pocket 1: 31, 43, 53, 32, 52, 54, 24

Pocket 4: 9, 62, 11

Pocket 6: 11, 65, 62, 66

Pocket 7: 65, 69

Pocket 9: 72, 73, 69, 76

4.2 Test for zygosity

Factor	Number of Cases	Number of Controls
Homozygous	x2	y2
Heterozygous	x1	y1
Absent	x0	y0

Comparison	Entries of 2 x 2 Table a	b	c	d	Test
Homozygote factor vs absent	x2	x0	y2	y0	Homozygosity associated?
Heterozygote factor vs absent	x1	x0	y1	y0	Heterozygosity associated?
Homozygote factor vs heterozygote factor	x2	x1	y2	y1	Zygosity associated?

4.3 Tests for residue interactions

(ref 7)

Factor A	Factor B	Number of Cases	Number of Controls
+	+	x1	y1
+	–	x2	y2
–	+	x3	y3
–	–	x4	y4

Comparison	Entries of 2 x 2 Table a	b	c	d	Test [Number]	
+ + vs – +	x1	x3	y1	y3	[1] A associated in B-positives?	Is A associated
+ – vs – –	x2	x4	y2	y4	[2] A associated in B-negatives?	independently of B?
+ + vs + –	x1	x2	y1	y2	[3] B associated in A-positives?	Is B associated
– + vs – –	x3	x4	y3	y4	[4] B associated in A-negatives?	independently of A?
					[1] + [3]	A and B interact?
+ – vs – +	x2	x3	y2	y3	[5] Difference between A and B associations?	
+ + vs – –	x1	x4	y1	y4	[6] Combined A·B association	
Association between A and B in cases	x1	x2	x3	x4	[7] Linkage disequilibrium in cases?	
Association between A and B in controls	y1	y2	y3	y4	[8] Linkage disequilibrium in controls?	

5. Contact information

Please contact Dimitri Monos [monosd (at) email.chop.edu] for comments, suggestions and clarifications.

6. References

- Garrett TPJ, **S**aper MA, Bjorkman PJ, Strominger JL, Wiley DC. Specificity pockets for the side chains of peptide antigens in HLA-Aw68, Nature 1989; 342: 692-696.

- Saper MA, Bjorkman PJ, Wiley DC. Refined structure of the human histocom- patibility Antigen HLA-A2 at 2.6 Å resolution, Journal of Molecular Biology 1991; 219: 277-319.

- Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, et al. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 1993; 364: 33-9.

- Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, et al. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature 1994;

368: 215-21.

- Androulakis IP, Nayak NN, Ierapetritou MG, Monos DS, Floudas CA. A predictive method for the evaluation of peptide binding in pocket 1 of HLA- DRB1 via global minimization of energy interactions.Proteins 1997; 29:87-102.

- Chicz, RM, Graziano DF, Trucco M, Strominger JL, Gorga JC. HLA-DP2: self peptide sequences and binding properties.Journal of Immunology 1997; 159; 4935-4942.

- Svejgaard A, Ryder LP. HLA and disease associations: detecting the strongest association. Tissue Antigens 1994; 43: 18-27.