

## Image treatment and classification for genetic polymorphism analysis by DNA microarray

M. GUERISOLI - M. GIACOMINI - C. RUGGIERO

DIST, University of Genova

The Hla complex, that is the human system of histocompatibility, is composed by intrinsic proteins that are present on all the cells and that behave like antigens. When these proteins come to contact with the immune system of one different individual, they are recognized like strangers and provoke the immune answer.

Its genes are situated on the short arm of chromosome 6, and they have been subdivided in two main classes: I, that comprises genes of type A, B and C and II, that comprises genes of type DP, DQ, DR.

The antigen of class I are present on all the nucleate cells of the organism. This distribution of the antigens of class I stress their importance as acknowledgment markers between that the organism recognizes like pertaining to it (self) and that that is stranger (non-self). It can be assumed therefore that the HLA is at the basis of the procedure of immunity acknowledgment. They are fundamental for the study of the problem of the refusal in the transplants. If the transplanted woven does not turn out to be histocompatibile, that is if the cells that compose it do not have the same HLA antigen of receiving patient, the transplant is rejected.

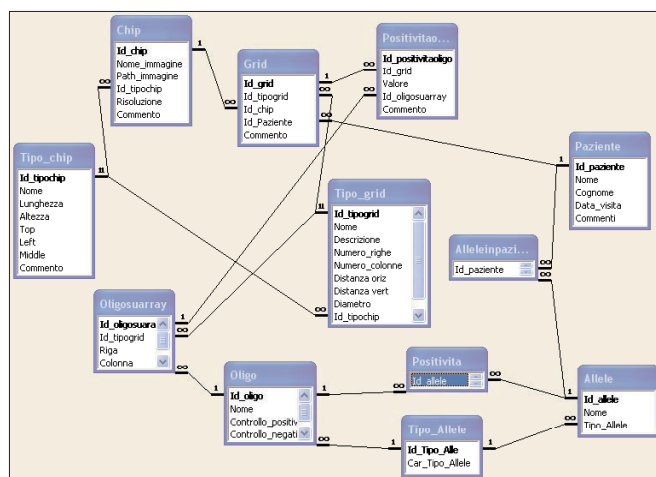
Our laboratory is developing a new project for automatic elaboration of images, obtained from experiments of HLA polimorfism on transplant attending patients; based on the well known technique of the microarray. The main scope is to obtain a software for high resolution classification of HLA loci both of class I and of class II.

The software mainly allows:

- to visualize saved images;
- to elaborate them;
- to identify fundamental data on the type of the allele that characterizes the patient;
- to record all data on a database;
- to visualize every moment all data present in the

database.

The database, structured according to the relational model, contains all data relative to the geometry of microarray slabs and to the identification of patients whose genetic material has been examined.



The entity-relation diagram, shown in figure, describes the inner structure of the database: all these tables can be group into three main groups. The first group contains the tables in which the archetype of the test is saved; that is the structure of the chip and the grid with all measures and distances that are necessary to the definition of an algorithm for the automatic identification of spots on the image.

This first group comprises:

- the 'Tipo\_chip' table, in which all data relative to the used slab are present; the slab turns out subdivided in two parts, one containing the grid in which the spot are inserted and another with its copy. The criterion of validation of the examination involves that the grid and its copy contains the same number of coated spot and in the same position. In the table the values length and height

of the chip, the values top and left of grid and the distance that separates the grid from its copy are also inserted.

- The 'Tipo\_grid' table, contains all data relative to the grid; the fields contain the number of the lines and the number of the columns that characterize the grid, the horizontal and vertical distance that separates each single spot and, finally, the value of their diameter. Moreover, in one of the fields a connection to the Tipo\_chip table is inserted in order to put in relation the used grid with the chip on which it is put.

The second group of tables comprises:

- the 'Oligo' table, that contains a list of all the oligonucleotides considered in this study;

- the 'Allele' table, with one list of the names of all alleles in study and a reference to the type of allele, A or B;

- the 'Positività' table, in which the two previous tables are put in relation; in this table the connections between the alleles and the oligonucleotides that compose to them are put in evidence.

- the 'Oligosuarray' table, with the values related to the theoretical position of each spot of the grid and the connection to the really type of grid used in this case.

All the values of the tables of the first and second group have been inserted in the DB before writing the MATLAB program, these data are fundamental for the correct elaboration of the images.

The third group of tables contain data relating to the patient and is subdivided into two parts, one relative to data identification and the other the obtained results.

This group includes:

- the 'Paziente' table, with the first name and the last name of the patient, the health identification code and the date in which the visit has been carried out;

- the 'Grid' table, connected to 'Patient', that indicate the specific grid used for the analysis of the patient;

- the 'Chip' table, connected to 'Grid', with the name and of the image of the considered grid and the path where it has been stored;

- the 'Positività oligo' table, with the list of the names and the intensities of all the oligo recognized to be excited during the specific analysis;

- the 'Allele in paziente' table, with the name of the allele or of the alleles that have been observed on

the image on the base of the excited oligonucleotides.

Once the db has been created, we implemented the Matlab program, subdivided in five interfaces.

The 'Scelta' interface is the first one presented to the user at the beginning of the program; it allows the user to choose between the analysis of one new image or the presentation of data relating to patients and images already inserted in the Db.

In order to elaborate an image the user must select the first choice. The interface 'Elabora immagine' is then shown; it allows the opening of the image and the selection of data related to the chip and to the grid that have generated the image. In the second place, by selection of the 'Segmenta immagine' key, two rectangles will appear on the main image, to indicate the zone in which the positive control should be present on each type of grid. The user can therefore verify if the experiment is successful; in fact, the positive controls, present you in various positions according the type of allele to which the image refers, must always be excited. If this did not happen the experiment would be considered not successful and the image would be useless.

Later with the 'Rilevamento spot' key some parameters of the positive control will be calculated; specifically: the contour and the intensity value of the spot. This phase is the more sensitive part of the elaboration process; in fact, it has been attempted to create a program able to find the presence of the spot, starting uniquely from the geometric shape and the intensity of the positive control spots that, in some cases can have small differences with the background.

For this purpose several functions of the image processing toolbox of Matlab have been used:

- 'Imcrop', it selects the part of the image contained in the rectangle previously enlighten;

- 'Imbowthat', it maximizes the contrast between spot and the background;

- 'Graythresh', it calculates the value of intensity of the image using the algorithm by Otsu;

- 'Bwmorph', it removes most of the noise, obtaining one image of the spot, white, on black background;

- 'Bwperim' e 'Immultiply', they define the contour of the spot and bring back it on the image.

All this data are then shown to the user on the main image. If positive controls are found, their real positions will determine the position of all

other oligoes. Then the same algorithm will be applied in these positions to define the presence of other excited plots.

Once all excited spots are shown on the main image, the program opens a new interface, 'Oligo\_accesi', in which the names of all excited oligos are shown, and the name of the allele or the alleles, that can be justified by the presence of these excited oligoes. In fact, according to the quality of the image, it is possible that the answer of the program is not uniquely determined. Moreover, in this interface, the user must insert the first name, the last name, the health identification code and the date in which the visit has been carried out. Selecting the key 'Salva', this data are saved permanently on the db so available for further db interrogation.

This interrogation can be performed by using the

key 'Apri' of 'Ricerca' of the 'Scelta' interface. The 'Paziente' interface is opened, in which a list of all the patients already saved on the db is shown. The user, by selection of one of these patients, obtains the opening of the 'Paziente\_scelto' interface, in which many data are shown: patient identification data, the list of the excited oligo and the allele/s that can be justified by these oligoes. Another key is also present on this interface, 'Vedi dati' that shows the real microarray image with excited oligoes singled out by white squares.

The project is in a preliminary phase, but we have already been obtained preliminary data, in particular as regard as the ability to find the spot, distinguishing them from the errors. A further appraisal will be necessary later, when the program will come used in the clinical routine.