CytoGenetic Assistant a New Cytogenetic Application Dedicaced to Telomeres Analysis.

Jean-Patrick Pommier

June 2000

Laboratoire de Biologie et de Radiobiologie
CEA, Fontenay-aux-Roses, France.

Draft for Image Processing Europe

The telomeres as molecular hourglasses

The number of divisions that most of normal human cells is limited *in-vitro*. This limit is due to the extremities of the linear structure of the chromosomes found in all the eucaryotic cells, called the telomeres. The telomeres are nucleo-proteic complexes located at the very end of the eucaryotic chromosomes. In each cell, the DNA sequence of each of the 92 human telomeres is composed of hundreds to thousands repetition of a six nucleotidic base pairs TTAGGG.These motifs are recognized by specific proteins that ensure the cellular stability of the chromosomes. Due to the mechanisms of DNA replication, the replication of the telomeres is incomplete leading to a progressive loss of TTAGGG DNA motifs in absence of a specialized enzyme : the telomerase. In human tissues, the telomerase is only expressed in a subset of normal cells at variable rates which do not always compensate the telomere shortening, the telomerase is constituvely expressed in most cancer cells. The limit of proliferatif potential of the normal cells might weight on the human longevity but also imposes constraints to the *in-vitro* cellular cultures. Thus the telomere length is an indicated tool to predict the proliferative capacities of cells, which had direct applications in cell therapies technology.

The length of the telomeres can be estimated by a molecular biology technique : the *Southern* blot. The *Southern* blot allow a direct estimation of the length of DNA molecules essentially constituted by telomeric DNA motifs. However, this technique has its drawbacks, apart from the heaviness of its carrying out, the *Southern* blot mixes the 92 telomeres from one cell and the telomeres from the million of cells requirred to perform an assay. The telomeric signal obtained by this technique, is highly heterogeneous as well as in terms of intercellular heterogeneity and in terms of intracellular heterogeneity.

Quantification of one telomeric signal in one cell by fluorescence microscopy

The fluorescence microscopy allows to quantify the telomeric signals of a single telomere of one chromosome of a single cell. The quantization of the telomeric signals on chromosomes by fluorescence microscopy leans on multispectral imaging, a minimum of two channels are required. The signal obtained after binding synthetic telomeric probes labelled with a fluorochrome such as the cyanine 3, is observed in the red channel and recorded with a CCD sensor. The chromosomes are colored with a specific DNA dye, such the DAPI, and observed in the blue channel of spectrum. The more a telomere has TTAGGG motifs, the more synthetic probes will bind to it and the more the signal will be bright, thus depending on the shape of the

chromosome and on the length of its telomeres up to four telomeric signals can be detected. The aim of a software dedicaced to the quantization of the telomeric signals is not only to detect and quantify the signals but also to link the signals to the chromosomes in order to construct organized datas which will allows to compare the telomeric signals of one chromosome to the others in one cell or to compare the length of a given telomere beyond different cells.

Waited functions for a cytogenetic application dedicaced to quantization of the fluoresent in situ hybridization (FISH) telomeric signals on metaphasic chromosomes.

The quantization of the telomeric signals is a sequential process. It begins with the acquisition of the images step, followed by images preprocessing, object segmentation, data processing steps.

The images acquisition is the critical step, from it, will depend the validity of the quantization approach. This step should allow to check the quality of the image and possibly to give the possibility to perform corrections on the acquirred images, this step also should allows to record the images in order to retrieve them easily.

Generaly an image preprocessing step is require to correct the imperfection of the image acquisition system. The corrections concern typically the correction of the uneven illumination of the microscopic field and the colors compensation when color spread occurs between the differents color channels. Uneven illumination of the images of the chromosomes as well as the images of the FISH signals, is corrected with a flatfield image acquirred with a slide constituted by a homogeneously fluorescent material. Color spreading is due to the overlapping of the excitation/emission spectrum between the fluorochromes. Depending on the characteristics of the excitation/dichrïc/emission filters, a substantial amount of the light emitted by one fluorochrome (for example DAPI) can be observed in the channel of another fluorochrome (for example FITC). It might be also necessary to check the registration of the images belonging to the different channels. After these corrections, it is still necessary to perform background substraction to ease the segmentation steps.

The segmentation steps are necessary in order to extract objects from the images. In the channel detecting the chromosomes, different kind of objects can potentially be detected from nucleus, clusters of overlapping chromosomes and single chromosomes. Until truly performant algorithms capable of resolving automaticaly the problem of the touching or overlapping chromosomes, interactive tools are required to performe this task even if automatic segmentation

algorithms are usefull for a first segmentation which had to be refined. The segmentation of the images of the FISH signals, in the case of the detection of the human telomeres, yields mainly to two kinds of FISH signals. The former case of signal corresponds to a single telomeric spot belonging to one chromatid of the chromosome, and the latter case corresponds to a pair of unresolved spots of two close chromatids. A second level in the segmentation steps consists in *image understanding* process which uses the extracted objects from the segmentations steps. The *image understanding* task consists in assembling the independant particules that the telomeric spots are, into pair of spots belonging to a particular chromosome, at a given cytogenetic location.

The end of the telomere analysis is close to be achieved when a database containing the intensities of the telomeric FISH signal can be constructed. This database can be constructed in different ways. A two columned matrix can be constructed when the chromosomes are not karyotyped, the two columns will correspond to the intensity of telomeric spots of the short and of the long arms of the chromosomes, and the number of line of the matrix will be equal to the number of chromosomes of the cell. From this first matrix of measurements, additional informations can be calculated such that the ratio of the intensity of the telomeric spots between the short and the long arms of the chromosomes, the ratio of the intensity between the spots of each chromatids of a given chromosomal arm and statistical parameters describing the shape of the distribution of the intensity of the telomeric signals of the cell. When the chromosomes of the cells are karyotyped, an other matrix of measurements can be constructed. This matrix will have a number of chromosomes equal to the number of pair of chromosomes in the cell, that is 46 for female human diploïd cells and 48 for male cells, with a number of lines equal to the number of homolog chromosomes per chromosomes, that is 2 for normal diploïd cells, thus for n analysed cells, this matrix will comprise $2n$ lines. This matrix allows to represent each chromosome of a given pair, as a two dimensional vector and to search for difference of telomere length between the homolog chromosomes. If additional FISH signals are quantified, this matrix can be extented to represent each chromosome as vector of higher dimension. It is desirable to expressed the intensity of the telomeric FISH signal in order to make them independent of the images acquisition conditions. The unit in which the intensity of the telomeric FISH signals are expressed, can be converted using an external calibration, this external calibration will be construct with telomeric molecules with a known amount of TTAGGG motifs, in this case the

converted units will be expressed in DNA length unit (kilobases). The external calibration can simply be the flatfield image itself, and the converted unit will be expressed as a percentage of the mean fluorescence value of the flatfield image. The major disadvantage of the choice of the external calibration, is that cell to cell variability due to the experimental conditions cannot be taken into account. Under the assumption that the rate of hybridization of two different probes varies the same way from one cell to another one even if their absolute rate of hybridization is different, a more promising approach is the use of another probe, such the alphoïd repeat sequences, as internal calibration. An external or an internal calibration require in both cases to include the expure time in the converted unit. When no external or internal calibration is available, it is still possible to render the unit of telomeric signals independant of the images acquisiton conditions, using an internal standardization calculated for each telomere of a cell from the mean and the standard deviation of the whole telomeric signals of the cell.

The end of the analysis of the telomere length distribution is achieved when the database containing the telomeric signal values can be submited to a data mining. Data mining can be performed when tools are available to compare the telomeres of different chromosomes of the same cell or belonging to different cells. For that purpose a feed back between graphical representations, such that uni or bidimensional histogramms, the database and objects of the images, is very useful for example to determine which chromosomes have statistically the shortest or the longest telomeres and gaver them in a gallery of chromosomes. An other part of the postprocessing step, consists in making color images, specific tools are often necessary to produce images where the shade and the contrast of the color have been mofified to make visible some details which appears well in a monochrome image but remains poorly visible in an unprocessed color image. Depending on the number of spectral component of the image, simple RGB color tools can be sufficient, but when four or more fluorochromes or when sequential hybridization experiments are performed, more specific color tools become necessary.

***CGA* (CytoGenetic Assistant), a dedicaced software for the analysis of FISH signals quantization**

The *CGA* software, developped by ADCIS for the Laboratoire de Radiobiologie et d'Oncologie (CEA, France), uses the low level and the intermediate symbolic representation image processing toolbox of Aphelion 3.0 ™ developped by ADCIS (http://www.adcis.net). CGA is designed to minimize the operator's manipulations throughtout the images acquisition

phase to the construction of a database containing the telomeric intensity values. CGA is organized into different modules dedicaced for the users and files management, the images acquisition, the images segmentation, the chromosomes classification, the extraction of the telomeric intensity values, the RGB color processing. All the functionalities required to analyse the telomeric signals on metaphasic chromosomes are implemented, data mining of the length of the telomeres is performed on conventional statistical analysis software after exportation of the database. CGA runs under Windows NT on PC platforms.