doi:10.1017/S1431927610057065
Microsc. Microanal. 16 (Suppl 2), 2010

Automatic Detection of Combed DNA

Shlomo Hareli,[1] Efrat Ozeri-Galai,[2] Batsheva Kerem[2] and Aryeh Weiss[3,4]
[1]Department of Applied Mathematics, Hebrew University, Jerusalem, Israel
[2]Department of Genetics, Hebrew University, Jerusalem, Israel
[3]School of Engineering, Bar Ilan University, Ramt-Gan, Israel
[4]Bio-imaging Unit, Institute of Life Sciences, Hebrew University, Jerusalem, Israel

Molecular combing is a process whereby many molecules of DNA (hundreds of kb) are identically stretched and aligned on a silanized glass surface, in a size and sequence independent manner. The addition of nucleotide analogs (for example, CldU and IdU) to cell cultures, allows detection of the replication pattern at different sites in the genome. Fluorescence in-situ hybridization (FISH) of DNA probes on combed DNA allows direct mapping of specific loci, with respect to their replication patterns, with a resolution of 1 to 4kb. Because the density of the fibers is high, rapid scanning and recording of hundreds or thousands of molecules are possible [1].

Images of combed DNA may include large numbers of DNA strands, but these images also include significant noise, and the strands often appear as relatively short, broken line segments, which makes automated identification and subsequent quantitative analysis of DNA replication patterns difficult. Here, we present automatic detection of strands formed by labeled nucleotide analogs on combed DNA. This is a necessary step towards the ultimate goal of identifying DNA probes and correlating such probes with replication patterns near the probe.

Slides of combed DNA with fluorescently labeled CldU and IdU were imaged with an Olympus IX81 inverted fluorescence microscope, using a 60x/NA=1.35 oil objective, and filter sets appropriate to the dyes which were used to label the nucleotide analogs. The images were acquired using an EMCCD camera (Andor IXON-885). The microscope and camera were controlled by Andor IQ software. Although the EMCCD acquires 14 bit images, these were converted to 8 bit images for subsequent processing. All image processing was done using ImageJ[2], as will be described below.

First, a 3x3 median filter was applied to the images, followed by a rolling ball filter of radius 3. Since the images contain at least a few parallel lines, a Fast Fourier transform (FFT) was used to find the angle of rotation of the lines, and the input image was then rotated so that in all subsequent processing, the DNA strands were horizontal. Once the lines are known to be horizontal, a morphological closing operation with a horizontal structuring element is used to generate lines from disconnected dots. This process may also generate short line segments. Such segments are excluded unless they are close enough to a valid line to be considered as part of that line. In that case, the existing line is extended to the segment. The ImageJ object analyzer is used to impose conditions for rejecting invalid lines. The program outputs position and length information to a file, so that those data can be used for further processing, or for comparing to "ground truth", that is, to manually detected lines in the same field.

A tool which allows manual marking of two classes of lines on an image of combed DNA was developed. This tool outputs the manually marked lines in the same format as the automatic detection tool, so that the two outputs can be easily compared.

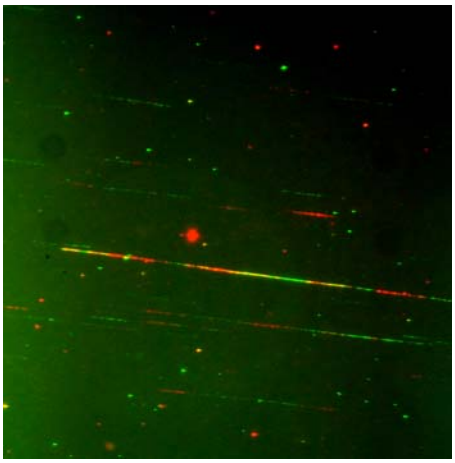
Figure 1a shows a typical input image, with two nucleic acid analogs labeled in red and green. Figure 1b shows the rotated image, following median filtering and background subtraction. Figure 1c shows the result of the segmentation. Note the rejection of most isolated dots, while groups of dots that

form lines are detected as such. The images in Figure 1 are actually cropped subsets of the original image, which was too large to include in this abstract.

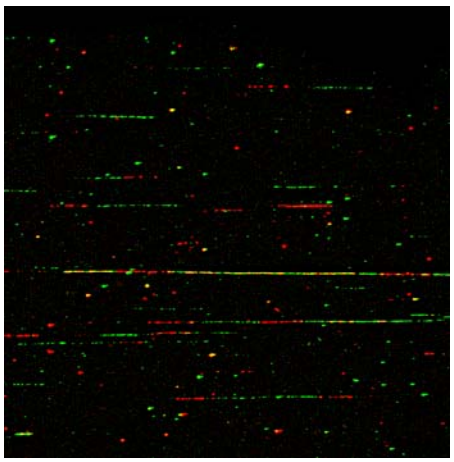
The algorithm is currently designed for a high detection rate, at the expense of some false negatives. When we introduce known correlations between the two color channels, we will be able to disqualify many of the false positives. In some experiments, there is a FISH marker encoded in a third color channel. Replication sites marked by the FISH label are very hard to find due to the noise caused by nonspecific adherence of the FISH labels to the silanized surface. By restricting the search for the FISH markers to locations near valid lines of DNA, we expect to greatly enhance the detection probability of valid FISH labels.[3]

References:

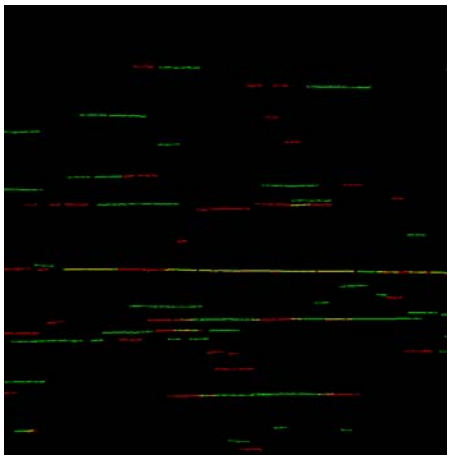
[1] J. Herrick. and A. Bensimon, *Biochimie* 81, 859-71 (1999).
[2] W.S. Rasband, ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/, 1997-2010.
[3] The support of the Israel Science Foundation (grant no. 467/07) is gratefully acknowledged.



(a)



(b)



(c)

Figure 1: (a) Orignal two channel combed-DNA image; (b) Rotated image, following median filtering and background subtraction; (c) Segmented image, showing detected lines. Yellow indicates that a red and green line are overlapping. This was possible in this experiment because teh first probe was not washed out before the second wes introduced.

https://doi.org/10.1017/S1431927610057065 Published online by Cambridge University Press