Automatic segmentation of centromeres, foci and delineation of chromosomes

Eva Hoffmann[1]
http://www.sussex.ac.uk/lifesci/hoffmannlab/
C. C. Reyes-Aldasoro [2]
reyes@city.ac.uk
http://staff.city.ac.uk/~sbbk034/

[1] Genome Damage and Stability Centre, University of Sussex, Brighton UK
[2] Biomedical Engineering Research Group, City University London, London UK

Abstract

The observation of chromosomes has been crucial for our understanding of their structure, function, organization, and evolution of genes and genomes [1] as well as morphological changes during mitotic and meiotic divisions [2]. In this work, we present an automatic algorithm for the segmentation of centromeres and foci of DNA processing proteins, as well as the delineation of convoluted chromosomes. The algorithm is fully automatic and does not require tuning of parameters. Statistical measurements of numbers, areas distance and lengths are provided by the algorithm. The work is preliminary as this algorithm has not been tested on a large database nor used to differentiate between populations, however, it is considered that given it is fully automatic and fast it should be a useful tool for the analysis of chromosomes.

1 Introduction

Chromosomes carry the genetic material and the genes are shuffled in the germline of sexually reproducing organisms before being passed on to their offspring in eggs and sperm. This occurs during a process known as recombination, which differs in males and females [3]. In females, recombination occurs during foetal development and the rates are twice as high compared to males. One factor implicated in the sex-specific differences as well as the tremendous variation between individuals of the same sex and even within individuals is how chromosome structure is set up: the DNA is arranged as chromatin loops on an axis, and the chromosome axes are approximately twice as long in females compared to males [4]. Measured *per* physical distance, the population-averaged number of recombination events is the same in the two sexes [5]. In addition, the accumulation of proteins into cytologically-detectable foci as biomarkers has been used as phenotypic measurements in a wide range of biological applications like biomarker for genotoxic insult [6].

Understanding how chromosome structure and recombination are correlated is important, since low recombination rates are associated with increased risk of chromosome transmission errors in the egg [7] and decreased reproductive success in women [8]. To assess chromosome structure and recombination simultaneously in oocytes, this work describes an automatic algorithm for the segmentation of centromeres, foci and delineation of chromosomes.

It has been recognised that some manual tasks like interpretation of [6] and bands [9], can be a technically difficult task even for highly skilled personnel. Even an experienced user can struggle to segment foci and delineate the chromosomes manually, besides that it is tedious, time consuming and error prone [10]. Automated methodologies for the detection of foci have been recently being proposed [6], however they are not integrated with the analysis of chromosomes and centromeres. Conventional tracing algorithms [11] do not consider the large amount of crossings and overlaps as those presented with chromosomes. In this work we describe an automatic algorithm that can combine the detection of centromeres and other foci and combine those results with the delineation of chromosomes.

2 Materials and Methods

2.1 Materials

Oocytes were lysed in a hypotonic solution and the chromosomes fixed on glass slides before being stained with antibodies against proteins that mark recombination events (MLH1), the chromosome axis (SYCP3), and centromeres (CREST). The 2D spread chromosomes were imaged using a fluorescence microscope. The images were acquired with an Olympus fluorescence microscope running Applied Spectral Imaging software and 100x magnification.

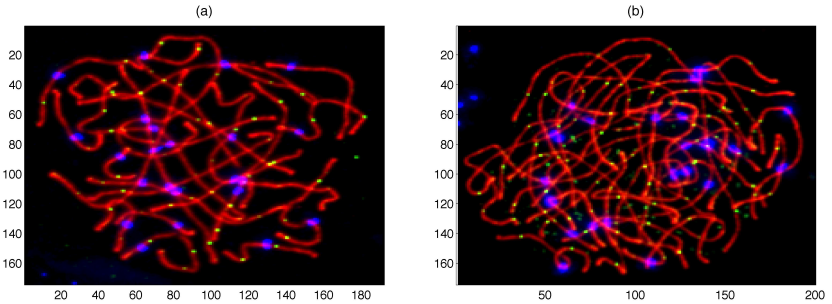

Figure 1: Two representative examples of the data. Chromosomes have been stained to appear red, the centromeres appear blue and the foci appear green. Notice the uneven intensity and complexity of the chromosomes. Notice also that some centromeres and foci do not belong to the chromosomes (middle right in (a), upper left corner of (b)), and also some centromeres that are very close to each other, there should be 23 in total in each.

2.2 Algorithm description

The first striking characteristic of the images is the convoluted nature of the chromosomes (Fig. 1), which is more clearly appreciated in the red channel of the images (Fig. 2a,d). Whilst it is of interest to delineate and identify each chromosome separately, and then analyse the distribution of foci on them, this was considered as a very complicated task. Therefore the algorithm proposed in this work is the following:

1. The RGB colour channels of the original image were separated (Fig. 2). The red channel was further processed in the following way: a foreground that consisted of all

pixels>0 was labelled, if more than one region was detected, the largest region, i.e. the chromosomes, was retained and all other regions, were considered background.

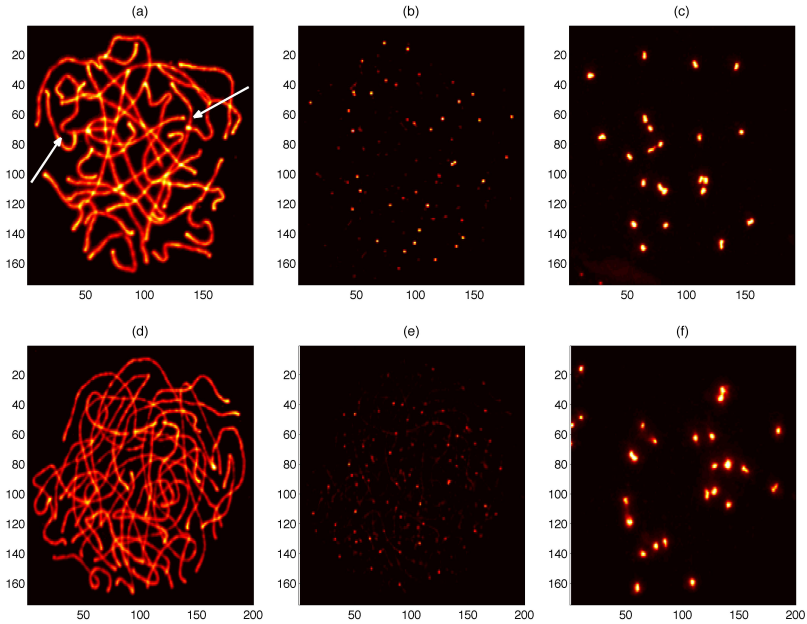

Figure 2: Data separated into (a, d) Red (b, e) Green and (c, f) Blue channels. The hotspots in (a, d) are due to overlapping chromosomes as well as leaks from the frequency bands of the centromeres (c, f). Notice the low intensities in some regions as well as proximity of between chromosomes (arrows)

2. All the lines that form the chromosomes in the red channel were traced (Fig. 3). The tracing was non-trivial as there were areas of low intensity and close proximity that can be artifactually joined. In addition, a simple watershed transform [12] would not trace the "tails" of the chromosomes that extend beyond crossing points as these would drain into a single basin. Therefore, the background region obtained in step 1 was subtracted from the red channel, which improved tracing provided by the watershed algorithm. Next, low intensity was used to remove the artefactual lines that were traced incorrectly (arrows in Fig. 3). Finally, crossing points over the traced lines were detected and removed to split them into non-connecting segments (Fig. 3c)
3. The foci on the green channel and the centromeres on the blue channel were segmented, in both cases using Otsu's algorithm [13] to determine automatically a threshold level. For the centromeres, a further step was necessary to verify that there were 23 centromeres as these correspond to the number of chromosomes. When the number of centromeres was lower, the largest and most elongated regions were split halfway [14] assuming that these corresponded to neighbouring centromeres that were not correctly segmented (Fig. 4a).
4. Using the segmented centromeres as starting points, the distance traversed along the traces delineated in step 2 was calculated (Fig. 4b). The distance from the foci could equally be calculated (Fig. 4c). In parallel, all segments were analysed for the number of contacts with foci. Fig 5 shows the quantification with different shades.

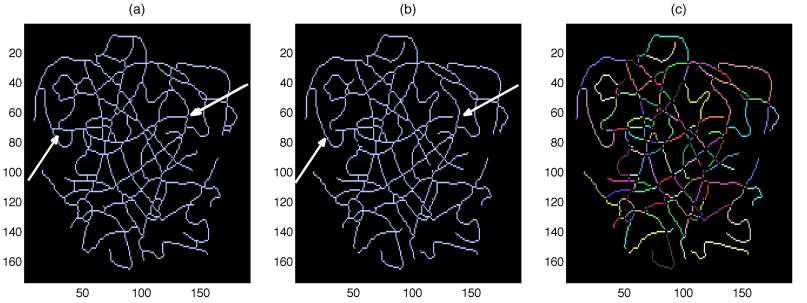

Figure 3: (a) Delineation of the ridges corresponding to the chromosomes with a watershed transform, with a modified input. Notice the lines that have been artifactually joined (arrows). (b) Deletion of low intensity regions (arrows). (c) Separation of individual segments of the chromosomes by removal of branch points. Colours have been assigned for visual discrimination.

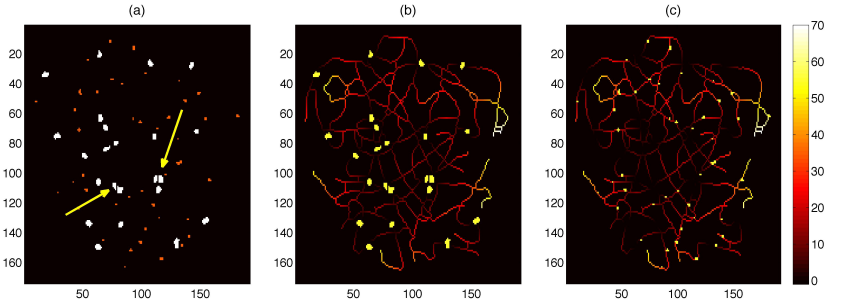

Figure 4: (a) Segmentation of centromeres (white) and foci (orange). Notice how centromeres that were too close to be segmented by intensity have been split (arrows). (b) Distance measured from the centromeres along the chromosome lines, increasing distance is denoted by a change in colour. (c) Distance measured from the foci along the chromosome lines. Colour bar indicates the distance in pixels from the centromere/foci.

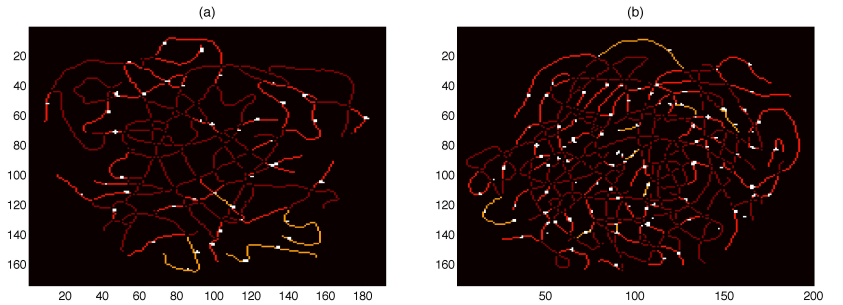

Figure 5: Quantification of number contacts between chromosomes and foci is indicated by colours: dark brown = 0 contacts, red = 1, yellow = 2.

5. Finally, statistical measurements were calculated including the number of foci, number of segments, pixels covered by the traces, and average length of segments.

3 Results

Two images of very different characteristics were used to analyse the algorithm, which has been illustrated in Figs. 2-5. In the absence of a gold standard, the images were visually analysed. The tracing revealed some artefacts, small lines in between two bright sections of chromosomes. However, it was estimated that these would have a minor impact on the measurements obtained. Statistics for the two images are presented in Table I. Some measurements can be indicative of the geometry of the chromosomes, like the number of segments, which could vary depending on the preparation of the sample. Some other measurements are indicative of the nature of the oocytes, like the number of foci or the distance from the centromeres to the foci. As an indication of the computational complexity of the algorithm, the average time to process the example images was 0.38 s running on Matlab R2013b on a PowerMac 2.5 GHz Intel Core i7 with 8GB RAM. No systematic attempt to make the code faster was made.

Measurement	Example A Fig. 1a	Example B Fig. 1b
Number of foci	43	99
Average area of foci [pixels]	3.5	3.0
Average area of centromeres [pixels]	16.5	19.5
Number of pixels covered by delineated chromosomes	2,329	3,372
Number of Segments	195	430
Average length of segment [pixels]	11.9	7.8
Average distance from centromere to foci [pixels]	19.4	17.8

Table 1: Statistical measurements extracted from the representative images.

4 Conclusion

An automatic algorithm for the segmentation and delineation of chromosomes, centromeres and other foci has been described. The nature of the images is very complex and difficult to analyse, even to expert users. Therefore, the measurements provided by this algorithm could provide elements to quantitatively analyse the chromosomes. Whilst the results presented so far are preliminary, the algorithm can in theory be applied to a variety of settings and could be further refined with a larger database. In addition, validation against a gold standard will be required. Furthermore, since the algorithm is fully automated, it is suitable for implementation in online repositories [15]. This algorithm could be of particular interest to analyse experiments of meiotic recombination [16] but its potential use extends beyond that.

It is important to note that the original arrangements of the chromosomes as well as consistent settings during image acquisition are very important. If, for instance, one image presents numerous crossings, whilst another one has chromosomes that have been disperse, those images would generate very different metrics.

References

[1] N. Preeda, T. Yanagi, K. Sone, S. Taketa, and N. Okuda, “Chromosome observation method at metaphase and pro-metaphase stages in diploid and octoploid strawberries,” *Sci. Hortic.*, vol. 114, no. 2, pp. 133–137, Oct. 2007.

[2] A. Copsey, S. Tang, P. W. Jordan, H. G. Blitzblau, S. Newcombe, A. C. Chan, L. Newnham, Z. Li, S. Gray, A. D. Herbert, P. Arumugam, A. Hochwagen, N. Hunter, and E. Hoffmann, "Smc5/6 Coordinates Formation and Resolution of Joint Molecules with Chromosome Morphology to Ensure Meiotic Divisions," *PLoS Genet*, vol. 9, no. 12, p. e1004071, Dec. 2013.

[3] S. I. Nagaoka, T. J. Hassold, and P. A. Hunt, "Human aneuploidy: mechanisms and new insights into an age-old problem," *Nat. Rev. Genet.*, vol. 13, no. 7, pp. 493–504, Jul. 2012.

[4] A. Lynn, K. E. Koehler, L. Judis, E. R. Chan, J. P. Cherry, S. Schwartz, A. Seftel, P. A. Hunt, and T. J. Hassold, "Covariation of synaptonemal complex length and mammalian meiotic exchange rates," *Science*, vol. 296, no. 5576, pp. 2222–2225, Jun. 2002.

[5] Y. Hou, W. Fan, L. Yan, R. Li, Y. Lian, J. Huang, J. Li, L. Xu, F. Tang, X. S. Xie, and J. Qiao, "Genome analyses of single human oocytes," *Cell*, vol. 155, no. 7, pp. 1492–1506, Dec. 2013.

[6] A. D. Herbert, A. M. Carr, and E. Hoffmann, "FindFoci: A Focus Detection Algorithm with Automated Parameter Training That Closely Matches Human Assignments, Reduces Human Inconsistencies and Increases Speed of Analysis," *PLoS ONE*, vol. 9, no. 12, p. e114749, Dec. 2014.

[7] C. Ottolini and et al., "MeioMaps of genome-wide recombination and chromosome segregation in human oocytes and embryos reveal selection for maternal recombination rates," *Nat. Genet.*, vol. accepted for publication.

[8] A. Kong, J. Barnard, *et al.*, "Recombination rate and reproductive success in humans," *Nat. Genet.*, vol. 36, no. 11, pp. 1203–1206, Nov. 2004.

[9] P. Lichter, T. Cremer, J. Borden, L. Manuelidis, and D. C. Ward, "Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries," *Hum. Genet.*, vol. 80, no. 3, pp. 224–234, Nov. 1988.

[10] P. Cheng, D. R. Pareddy, T.-H. Lin, J. K. Samarabandu, R. Acharya, G. Wang, and W. S. Liou, "Confocal Microscopy of Botanical Specimens," in *Multidimensional Microscopy*, P. C. Cheng, T. H. Lin, W. L. Wu, and J. L. Wu, Eds. Springer New York, 1994, pp. 339–380.

[11] E. Meijering, M. Jacob, J.-C. F. Sarria, P. Steiner, H. Hirling, and M. Unser, "Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images," *Cytom. Part J. Int. Soc. Anal. Cytol.*, vol. 58, no. 2, pp. 167–176, Apr. 2004.

[12] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.

[13] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[14] C. C. Reyes-Aldasoro, L. J. Williams, S. Akerman, C. Kanthou, and G. M. Tozer, "An automatic algorithm for the segmentation and morphological analysis of microvessels in immunostained histological tumour sections," *J. Microsc.*, vol. 242, no. 3, pp. 262–278, Jun. 2011.

[15] C. C. Reyes-Aldasoro, M. K. Griffiths, D. Savas, and G. M. Tozer, "CAIMAN: an online algorithm repository for Cancer Image Analysis," *Comput. Methods Programs Biomed.*, vol. 103, no. 2, pp. 97–103, Aug. 2011.

[16] E. R. Hoffmann and R. H. Borts, "Meiotic recombination intermediates and mismatch repair proteins," *Cytogenet. Genome Res.*, vol. 107, no. 3–4, pp. 232–248, 2004.