

Automated Identification of Chromosome Segments Involved in Translocations by Combining Spectral Karyotyping and Banding Analysis 读书笔记

*Max Liu*¹

2020/07/07

1 文献简介

Automated Identification of Chromosome Segments Involved in Translocations by Combining Spectral Karyotyping and Banding Analysis

Benoît Legrand, Che Sau Chang, Sim-Heng Ong, Soek-Ying Neo, and Nallasivam Palanisamy

- 2008 年的老文章 [1] 了；
- 去年读过，现在想要再研究一下其中染色体密度谱曲线的计算方式；
- 通读之后觉得文章写得还是有点晦涩，很多地方表述很长，想要短句...

2 文献概括

根据之前的阅读记忆和昨天的泛读，这篇文章在已有相对充足的不同号染色体以及各自不同核型的染色体**图像的基础上**，根据这些图像计算染色体的**灰度密度谱曲线**，包括用于参照的标准密度谱和可能异常的染色体密度谱曲线，再利用 **DTW 算法**判断可能的染色体**易位位点**。

¹*Max Liu*: lzmlzm@stu.xjtu.edu.cn

3 基本概念

在多次阅读该文后，根据自己的理解阐述本文的一些基础知识：

染色体号与核型的区别

嗯上网查了一下，核型就是一个体细胞内所有染色体的排列，可以理解为（一个细胞内）一套染色体的特征。所以文中多次提到某号染色体不同核型其实就是指一堆该号染色体...

DTW 序列对齐

简单查了一下,这篇博客比较清楚了:[时间序列对齐:通过编程理解 DTW\(dynamic time warping\)](#)。但是 DTW 可以找两个不同长度序列对应的位点，却不方便直接把二者长度变为相同。

自己做实验的处理操作是直接缩放（缩放是否要根据位点，待定）。

4 具体方法

首先明确方法的实现需要计算染色体灰度密度谱，下面是重点的几种计算。

A. 染色体图像中轴确定

两种方法：

- thinning 算法确定染色体的中轴；
- 确定轴上一点后寻找到染色体两端的最短直线，文章似乎用这个。

自己做实验的话不用确定中轴了... 直接拿图像开干 $O(n_n)O$

B. 单个染色体灰度密度谱的计算

文中有中轴后，再计算平均灰度，然后归一化灰度序列。归一化直接 0 均值，方差 1 即可，可以降低不同图像本身亮度差异，以及减小着色误差。注意还有个阈值化操作。但是具体的实现细节文章确实没有给，需自行探索。

Emmmm，自己做实验的时候直接灰度平均喵喵喵

C. 用于参照的标准染色体灰度密度谱的计算

标准密度谱是为了和可能异常的染色体密度谱片段比较。

标准密度谱的确定需要考虑到正常染色体的图像质量、核型、伸长现象（我理解就是图像比例）。文中取某号正常染色体的 4 种核型，用 DTW 方法对齐后再取灰度平均，并继续平均 4 种核型。最后归一化并统一长度为 100（位点），应该是直接缩放。

D. i 号与 j 号（均正常）染色体长度比 CLR_{ij} 计算

该长度比是为了服务 scale ratio，至于为什么要有这个定义及其合理性，哥是没看明白...

$$CLR_{ij} = \frac{\bar{L}_i}{\bar{L}_j}, \text{ where } \bar{L}_i = \frac{1}{N} \sum_{k=1}^N L_{i,k}, \text{ where } L_{i,k} = \frac{l_{i,k}}{l_{total,k}}$$

这个定义真的是... 后面的 $L_{i,k}$ 是第 i 号染色体的核型 k 的标准化长度， $l_{total,k}$ 是除性染色体外核型 k 的所有染色体长度，一共 N 个核型。最后的这个标准化合理性不太明显，我也没看出来...

实现起来需要 N 个核型中所有染色体图片，自己做实验条件不够。

E. scale ratio 计算

scale ratio 就是个比例，目的是比较不同核型，因为同号染色体不同核型的同一区域可能在图片上显示出不同长度。

若同核型，则比例为 1；若同号染色体不同核型，则可能需要统一尺度，因此定义 scale ratio

$$\Gamma_i = \frac{l_i}{\hat{l}_i}$$

其中 l_i 是第 i 号标准密度谱长度，分母的 hat 对应异常核型，但是异常核型的 i 号染色体不适合用于计算 ratio，异常嘛。所以从异常核型中选一个不同号正常染色体，用 CLR_{ij} 转化，近似为异常核型 i 号染色体的潜在真实长度，因此

$$\Gamma_i = \frac{l_i}{\hat{l}_i} = \frac{l_i}{\hat{l}_j / CLR_{ji}}$$

其中不放心还可以多取几个 j 来近似✓

最后就是染色体易位位点判断，思路与我想的一致，只是手头上不方便整合两条对应的染色体，所以 DTW cost 只能有一项了。

参考文献

- [1] B. Legrand, C. S. Chang, S. Ong, S. Neo, and N. Palanisamy. Automated identification of chromosome segments involved in translocations by combining spectral karyotyping and banding analysis. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 38(6):1374–1384, 2008.