

0. 論文

Circular binary segmentation for the analysis of array-based DNA copy number data

ADAM B. OLSHEN, E. S. VENKATRAMAN

Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA
olshena@mskcc.org

ROBERT LUCITO, MICHAEL WIGLER

Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

タイトル : Circular binary segmentation for the analysis of array-based DNA copy number data

著者 : ADAM B. OLSHEN, E. S. VENKATRAMAN, ROBERT LUCITO, MICHAEL WIGLER

arXiv投稿日 :

学会/ジャーナル : Biostatistics 2004

1. どんなもの？

- DNAコピー数変化の分析に特化した論文
- バイナリーセグメンテーションを改良してノイズに強い手法の開発
- コピー数変化が知られている細胞データと乳がん細胞データを用いて実証する

2. 先行研究

- 閾値ベースで増減を識別する手法
 - Weiss(2003)
 - 参照データと試験データは同じばらつきを持っている仮定がされている
 - この仮定が少し怪しいかもしれない
- ガウス隠れマルコフモデルを利用した手法
 - Snijders(2003)
 - ヒューリスティックな手法
 - コピー数変化の位置の最尤推定量を構築して位置特定を行い、統計的な有意性を評価する

3. コアアイデア

- バイナリーセグメンテーションの改良
- $Z_{ij} = \{1/(j-i) + 1/(n-j+i)\}^{-1/2} \{(S_j - S_i)/(j-i) - (S_n - S_j + S_i)/(n-j+i)\}$
- 上記を最大にする i, j を探し出す ($1 \leq i < j \leq n$)
- おそらくこれが棄却されなくなるまで分割し続ける
 - i, j が始点終点に近い場合は特殊な処理をする必要がある
 -

4. どうやって有効だと検証した？

- Snijdersに掲載されているaCGHデータに対してCBSを適用
 - 元論文では15の改変領域が見つかったが、提案法では12個見つかった
 - 偽陽性と思われるデータもいくつか見つかった
 - 局所的な傾向を近似しているため偽陽性が発生する

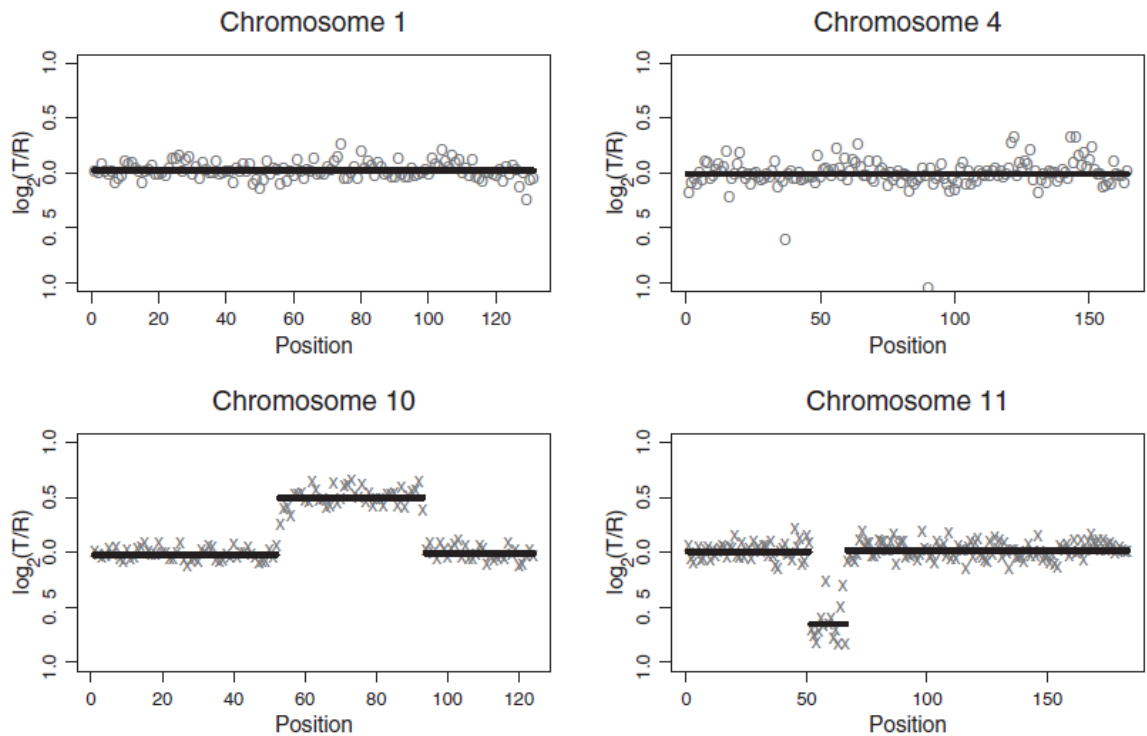


Fig. 1. A CBS analysis of the fibroblast cell line GM05296, which has known alterations only on chromosomes 10 and 11. The points are normalized log ratios, and the lines are the mean values among points in segments obtained by CBS.

-
- 乳がんのデータに適用
 - 重要だと思われる染色体に複数の変化が検出された

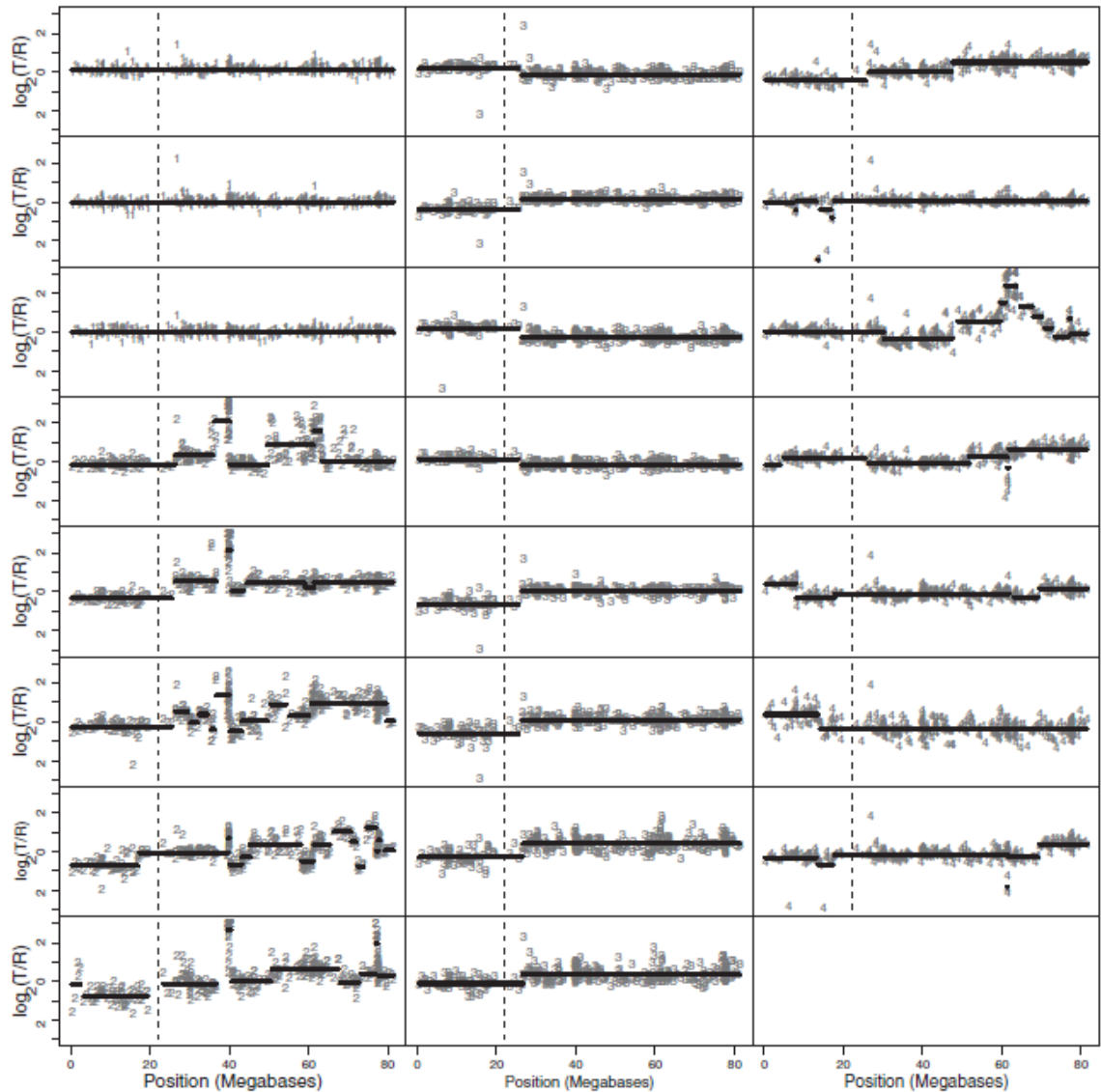


Fig. 3. CBS applied to 23 breast cancer cell lines using the ROMA technology. The plotting symbol for cell lines with no changes is 1, the symbol for those that appear to have ERB-B2 amplifications (near 40 MB) is 2, the symbols for those that appear to have whole-arm alteration is 3, and the symbol for those with other changes is 4. The dashed line is at the centromere.

○

5. データセット

- Snijders(2001)に掲載されているaCGHデータ
 - これは自由にダウンロードできる
- ROMA実験のデータ
 - こちらは乳がんのデータ

6. 疑問点

7. 次に読むべき論文は？

キーワード

- Array CGH
- バイナリセグメンテーション
- 変化点検出