

DOI: 10.3724/SP.J.1005.2008.00407

## 应用 CGH 数据和树模型探索癌症的发病机理

李小波<sup>1,2</sup>, 陈俭<sup>1</sup>, 吕炳建<sup>1</sup>, 来茂德<sup>1</sup>

1. 浙江大学医学院病理学与病理生理学系, 杭州 310058;  
2. 浙江教育学院信息学院, 杭州 310012

**摘要:** 比较基因组杂交技术(comparative genomic hybridization, CGH)主要用于检测肿瘤的染色体缺失和扩增, 迄今已积累了大量的实验数据, 为全基因组分析肿瘤的发生机制提供了可能。树模型在生物信息学领域通常被用于研究生物形成和进化的历史, 物种之间的进化关系常以系统发生树来表示。树模型同样可以作为一种有力的生物信息学工具来分析 CGH 数据, 探索癌症的发病机理。文中介绍了两种常见的树模型——分支树和距离树, 详细叙述了重建树模型的基本原理和方法, 分析了创建树模型时要注意的几个技术问题, 并对其在肿瘤研究中的应用进行了回顾和总结。肿瘤的树状模型作为单路径线性模型的泛化, 克服了以往单路径线性模型的缺点, 理论上能更加精确地概括到肿瘤的多基因、多路径、多阶段的发生发展模式, 从不同角度探讨肿瘤发生发展的分子机制。该模型除可用于分析肿瘤的 CGH 数据外, 还可用于分析其他多种类型的数据, 包括微阵列 CGH(array-CGH)技术等产生的高分辨率数据。

**关键词:** 癌症发病机理; 比较基因组杂交(CGH); 树模型; 分支树; 距离树

## Exploration of carcinogenesis based on tree models using CGH data

LI Xiao-Bo<sup>1,2</sup>, CHEN Jian<sup>1</sup>, LV Bing-Jian<sup>1</sup>, LAI Mao-De<sup>1</sup>

1. Department of Pathology & Pathophysiology, School of Medicine, Zhejiang University, Hangzhou 310058, China;  
2. School of Informatics, Zhejiang Education Institute, Hangzhou 310012, China

**Abstract:** Comparative genomic hybridization (CGH) can detect chromosomal deletions and amplifications of tumors, and various laboratories and public databases have accumulated a large number of CGH data, providing the opportunity to analyze the molecular mechanism of tumorigenesis in the whole genome. Tree models are generally used to study the history of biological formation and evolution in the field of bioinformatics, and evolutionary relationships between species are usually represented using phylogenetic tree. Tree models are also powerful bioinformatics tools to analyze CGH data and explore carcinogenesis. Two common tree models, the branching tree and the distanced-based tree, as well as their basic principles, methods are introduced detailedly, several technical problems in construction of tree models are discussed, and their applications in cancer research are reviewed systematically in this paper. As a generalization of single path linear model, tree models can more accurately conclude multigene, multistep, multipathway process of tumorigenesis, exploring the molecular mechanism of tumorigenesis from different angels. Apart from CGH data, tree models can be used to analyze various types

收稿日期: 2007-09-17; 修回日期: 2008-01-20

作者简介: 国家自然科学基金项目(编号: 30371605)资助[Supported by the National Natural Science Foundation of China (No. 30371605)]

作者简介: 李小波(1969-), 男, 浙江人, 博士研究生, 讲师, 专业方向: 医学生物信息学。Tel: 0571- 88208208; E-mail: oboaixil@126.com

通讯作者: 来茂德(1960-), 男, 浙江人, 博士, 教授, 研究方向: 肿瘤病理学。Tel: 0571- 88208208; E-mail: lmd@zju.edu.cn

of data, including high-resolution data (e.g., array-CGH data).

**Keywords:** carcinogenesis; comparative genomic hybridization; tree model; branching tree; distance-based tree

近几十年来, 分子生物学的研究表明, 肿瘤的形成和发展是基因改变的结果, 并且这些基因改变往往不只是单个基因, 而是不断累积的多基因改变。在已提出的肿瘤发生模型中, 以Vogelstein等<sup>[1]</sup>提出的单路径线性模型最为大家所接受, 然而该模型仍然存在着相当大的缺陷。实际上, 结直肠癌的发生发展除Vogelstein模型之外, 还有其他多条通路, 由于肿瘤的遗传异质性, 其他类型的肿瘤也存在类似的情况。

比较基因组杂交技术<sup>[2]</sup>(comparative genomic hybridization, CGH)能够一次测出整个染色体组所有染色体片段的缺失和扩增, 不仅快捷, 而且劳动强度低, 因此, 作为研究工具得到了广泛应用。CGH技术建立的10余年间, 各实验室和公共数据库已积累了大量的CGH数据<sup>[3]</sup>, 如何对这些数据进行有效的分析, 并用以解释肿瘤的发生机制, 已成为一个亟待解决的问题。然而, 在以往的研究中, 分析CGH数据往往只是判定某个染色体区带的改变是否为非随机事件, 并不能揭示这些重要事件之间的关联。

鉴于单路径线性模型的缺陷, Desper等<sup>[4,5]</sup>设计了树模型软件oncotrees (<http://www.ncbi.nlm.nih.gov/CBBresearch/Schaffer/>), 试图通过分析大量的CGH数据重建树模型来阐明肿瘤发生发展的分子机制。作为单路径线性模型的泛化, 树状模型能更加精确地概括到肿瘤的多基因、多路径、多阶段的发生发展模式<sup>[6]</sup>。

## 1 原理和方法

### 1.1 CGH原理

CGH是由荧光原位杂交(florescence *in situ* hybridization, FISH)技术衍生出来, 它采用了不同颜色的荧光标记方法标记样本DNA和参照DNA, 然后, 这两种标记DNA的混合物与有丝分裂中期的细胞染色体进行竞争性杂交。若样本染色体某一片段存在缺失, 则参照DNA优先与中期染色体杂交; 若样本染色体某一片段存在扩增, 则样本DNA优先与中期染色体杂交; 若样本染色体是平衡的, 即不存在缺失和扩增, 样本DNA和参照DNA等量与中期染

色体杂交。杂交形成的图像再经计算机软件处理, 通过定量精确分析染色体上的荧光分布, 就可以从全基因组范围内观察有丝分裂中期细胞的大片段DNA拷贝数的不平衡, 即DNA拷贝数的增多或减少。

### 1.2 树模型

CGH的实验结果是一系列反映全基因组DNA拷贝数差异(copy number aberrations, CNAs)的数据集, 每一个肿瘤组织产生一条数据记录。该数据集可被用来重建树模型。可以证明, 只要样本量足够大, 重建的树模型可以描述该数据集中CNAs的真实分布情况。常见的树模型有分支树和距离树两种。

#### 1.2.1 树模型的重建问题

假设树模型 $T=(V, E, r, p, L)$ 是有根树,  $V$ 表示树中各节点的集合,  $E$ 是成对事件的集合,  $r$ 表示根节点(该节点不属于 $L$ , 代表细胞的正常状态), 边 $e \in E$ ,  $p(e)$ 是该边的概率, 重要事件集合 $L$ 是节点的独立非空集合,  $L \subseteq V$ 。树模型 $T$ 在空间 $2^L$ 产生一个概率分布 $p_T$ 。树模型重建可以表述为:

输入:

- 重要事件集合 $L$
- $k$ 个样本, 在空间 $2^L$ 有概率分布 $p$

输出:

- 树模型 $T=(V, E, r, p, L)$ , 使得 $p_T$ 最接近真实的概率分布 $p$ 。

#### 1.2.2 分支树模型<sup>[4]</sup>

在分支树模型中, 树中的所有节点都必须属于集合 $L$ 。从事件 $i$ 到事件 $j$ 的有向边的权重函数可以定义为

$$w_{ij} = \log(p_{ij}) - \log(p_i + p_j) - \log(p_j)$$

假设包括根节点在内共有 $n$ 个事件, 那么共有 $n*(n-1)$ 个可能的有向边, 但最终的树模型只有 $n-1$ 个边, 重建的树模型将使得该树的所有有向边的权值之和为最大, 该树称为最大分支树。

分支树模型以一个节点为根, 而其他节点代表各个重要事件, 在同一支上, 边 $i \rightarrow j$ 表示一个假定的因果关系, 即事件 $i$ 的发生使得事件 $j$ 的发生变

得更为可能。分支树模型是Vogelstein单路径模型的泛化,从每一个节点出发有可能有多条路径发生,因而体现了肿瘤的异质性。利用创建好的分支树模型图,可能获得以下的信息:(1)在一个分支上按顺序走向的各个事件之间有可能存在因果关系,其中前一个事件的发生会引发后一个事件的发生;(2)聚集在同一分支上的事件可以作为肿瘤的亚型标记。

### 1.2.3 距离树模型<sup>[5]</sup>

在距离树模型中,集合 $L$ 中的所有节点必须是树中的叶节点,也就是说,允许一些“隐含”的、“未命名”的节点处于树中的内部节点。

如果 $x$ 和 $y$ 是树 $T$ 的两个叶节点,那么他们之间的距离可以如下公式计算

$$d_T(x,y)=-2\log p_{xy}+\log p_x+\log p_y$$

由此推测出重要事件集合 $L$ 中各个事件之间的距离矩阵 $T_L$ ,然后查找距离矩阵最接近 $T_L$ 的树。可以证明当样本集的数量足够大时,那么重建的树模型将与真实的树模型足够逼近。

距离树模型从另一角度来阐述肿瘤的发病机理。利用创建好的距离树模型图,可能获得以下的信息:(1)靠近根的事件是早期事件;(2)聚集在树的某一分支的事件可以作为肿瘤的亚型标记。树中的每条边都有长度,每条边的长度与节点间的水平距离成比例,相比于分支树,距离树更精确地计算了各事件到根节点的距离。

### 1.2.4 创建树模型时要注意的几个技术问题

在利用CGH数据创建树模型时,涉及到对CGH数据进行前期的处理,需要注意以下两个问题:

#### (1) 染色体区带的划定

在分析CGH数据时,往往需要考虑到染色体区带的划定,需从以下几个方面来考虑的:第一,鉴于CGH技术本身的特点,CGH对检测缺失的灵敏度低于扩增的灵敏度,对扩增的分辨率在2 Mb左右,而对缺失的分辨率在10~12 Mb左右<sup>[7]</sup>。因此选择的区带宽度也就必须大于这一范围;第二,也要考虑到CGH技术中的“漂移”(shift)现象和“旁观者”效应(bystander effect),宽度过小,也不能真实地反映染色体片段的DNA拷贝数差异;第三,假如把一条染色体臂上的所有缺失和扩增归并在一起,即染色体臂的任一区带的DNA拷贝数改变都被归结到该染色体臂上。这样的研究结果,增加了后期基因定位的难度,很明显,染色体区带的宽度越大,在此范围

内的基因数量越多,也就越不利于找到目标基因。因此,采取何种方案,应视具体情况而定,Jiang等<sup>[8]</sup>将区带宽度定在以染色体臂为单位,而Huang等<sup>[9]</sup>则采用了区带归并的方法。

#### (2) 重要事件的选择

CGH的数据结果记录下了所有被检测的肿瘤组织的染色体区带的变化,显然这样的结果将可能会分布在整个基因组范围。在某些区带上的扩增或缺失是经常发生的,统计学上称之为“高频率”事件或“重要”事件;另一些染色体片段没有缺失或扩增,或者有些染色体片段的扩增或缺失在所有样本集中出现的次数非常少,称之为“低频率”事件或“偶然”事件,这些事件很有可能只是随机的“噪声”或是肿瘤发生发展过程中基因的不稳定造成的,在树模型的后续分析中会放弃这些区域,而只把目光集中在“重要”事件上。“重要”事件的选择,对创建肿瘤发生发展分子机制的树模型非常重要。可以有不同的方法来选择重要事件,比如在Desper等<sup>[4]</sup>文中,采用了“团集”(Clique)启发式的方法,“团集”指的是集合中所有成对事件都相互关联的子集合,通过限制这些成对事件在同一肿瘤组织中发生的最少次数,可以得到“团集”中包含重要事件的个数。Brodeur等<sup>[10]</sup>的方法是用来选择重要非随机事件的另一方法,该方法假设在每个染色体区带,扩增和丢失的机会均等,并且其发生的概率与染色体区带宽度成正比,因此这是一个先验概率分布的方法。在Desper等<sup>[5]</sup>和Huang等<sup>[9]</sup>文中,采用了这一方法。

## 2 应用

树模型自被提出以来,被不少研究者用于探索肿瘤发生发展的分子机制。研究者利用树模型分析了包括乳腺癌<sup>[11]</sup>、膀胱癌<sup>[12]</sup>、头颈鳞状细胞癌<sup>[13]</sup>、鼻咽癌<sup>[9,14]</sup>、胸腺瘤<sup>[15]</sup>、大鼠乳腺癌<sup>[16]</sup>等肿瘤在内的CGH数据,得到了一些重要的早期改变事件,以及标志肿瘤中几个亚型的重要事件,明确了这些事件发生的顺序以及不同的通路。Simon等<sup>[17]</sup>还利用树模型分析了244例卵巢癌组织中染色体断裂点的数据。

在Desper等的工作基础上,von Heydebreck等<sup>[18]</sup>采用极大似然估计算法(Maximum likelihood estimation)建立了发病机制树模型(oncogenetic tree models),并在173例肾细胞癌的染色体变异数据中应用。Cremer等<sup>[19]</sup>结合81例FISH技术数据,应用该

模型重建了多发性骨髓瘤的同源细胞演化树模型,揭示了多发性骨髓瘤的不同亚型。Gunawan等<sup>[20]</sup>将该模型用于鉴定胃肠间质瘤细胞演化的几个通路。另外, Beerenwinkel等则提出采用混合模型(mixture models)和交叉学习方法创建基因突变树模型(mutagenetic trees), 编写了Mtreemix软件包<sup>[21]</sup>, 分别被Beerenwinkel等<sup>[22]</sup>用于HIV逆转录酶耐药性的基因突变研究, 和Rahnenführer等<sup>[23]</sup>结合杂合性丢失(LOH)数据用于成胶质细胞瘤的生存分析中。Jiang等<sup>[8]</sup>总共获取了 116 例肾细胞癌组织的CGH资料, 并利用这些数据重建了肾细胞癌的树模型。下面简要介绍Jiang等<sup>[8]</sup>的工作。

### 2.1 重要事件的选择

Jiang等<sup>[8]</sup>将区带宽度定在以染色体臂为单位, 忽略了容易产生假阳性结果和较短的染色体臂, 这样总共剩下 36 个染色体臂, 又同时对每个染色体臂考虑扩增和缺失两种情况, 因而共有 72 个可能的事件。

接着他们采用Brodeur等<sup>[10]</sup>的方法选取重要事件。应用模拟的方法, 随机产生 10 000 个所有染色体区带分布的副本, 而对于每个副本, 计算出每个事件的实际分值, 并计算出在该CGH数据结果集里发生的所有事件的最大分值。假设在 10 000 个副本中, 有 95% 的副本, 也即有 9 500 个副本其分值超过以上设定的最大分值的话, 该事件即被选为非随机事件, 也即重要事件。

### 2.2 肾细胞癌的分支树和距离树模型图分析

图 1 和图 2 分别是Jiang等<sup>[8]</sup>得到的肾细胞癌分支树和距离树模型图。

对比图 1 和图 2 的两类树模型, 至少可以有以下结论: 1)事件-13q, -9p 和-18p 预示一个亚型; 2)事件-6q, +17q, +17p 预示一个亚型; 3)事件-4q 和-4p 密切相关; 4) -3p 和-4p 是重要的早期事件。

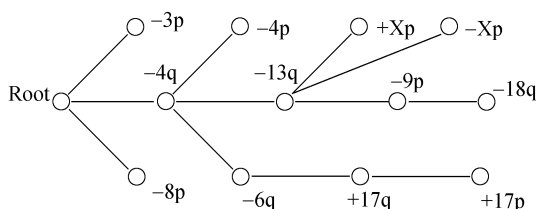


图 1 肾细胞癌的分支树模型<sup>[8]</sup>

Fig. 1 Branching tree for renal cancer data<sup>[8]</sup>

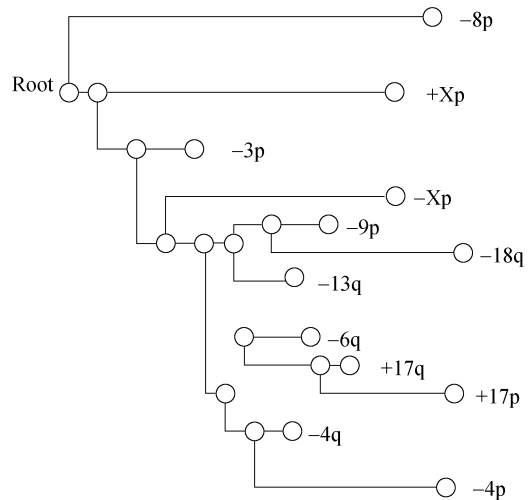


图 2 肾细胞癌的距离树模型<sup>[8]</sup>

Fig. 2 Distance-based tree for renal cancer data<sup>[8]</sup>

分支树和距离树采用了不同的数学原理, 它们之间各有优缺点。对于两类树, 早期事件都靠近根节点, 而距离树更为准确, 直接可从该节点与根节点的水平距离推断。两类树都倾向于把经常共发生的事件聚集在同一分支, 其中距离树较为准确地保存了所有成对事件共发生的信息, 而分支树则只把最为相关的成对事件放在同一分支。分支树的优点是可以直接给出在同一分支事件间的因果关系。在实际使用时, 两类树都会被创建出来, 并两两对照, 如果预测一致, 就会提高该预测的可信度, 反之则存疑。

### 3 结论和展望

Desper 等人创建的树模型, 克服了以往 Vogelstein 等人提出的结直肠癌发生发展的单路径线性模型的不足之处, 理论上能更加准确地概括肿瘤的多基因、多路径、多阶段的发生发展机制。通过距离树和分支树模型的重建, 试图回答以下问题: 有哪些和某类肿瘤相关的重要基因改变事件, 这些事件有无因果关系并遵循怎样的顺序, 哪些事件常常共发生并可以用来区分肿瘤的亚型。试图通过解答上述问题, 从不同角度探讨肿瘤发生发展的分子机制, 从而获得一些有意义的结果。

通过文献复习, 我们还发现, 即使是CGH数据, 也只有少数肿瘤被用来重建这一模型, 还有其他众多种类的肿瘤可以利用这一有力的工具进行深入研究。而且, 树模型并不局限于CGH数据分析, 比如之前描述的染色体断裂点数据、FISH技术数据、杂合



性丢失数据等, 还有包括微阵列CGH(array-CGH)技术<sup>[24]</sup>、单核苷酸多态性(SNP)芯片技术<sup>[25]</sup>等产生的高分辨率数据, 都可以用来重建树模型, 应用前景广阔。利用重建肿瘤发生发展的树模型, 可以加深对肿瘤发生发展机制的认识, 加快对肿瘤易感位点的查找和定位。同时可以结合以前相关肿瘤的研究结果和肿瘤基因表达谱<sup>[26, 27]</sup>, 进行深入研究, 发现标记物(biomarker), 从而使我们有可能会在基因水平上进行肿瘤的预防、诊断和治疗。

### 参考文献(References):

- [1] Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM, Bos JL. Genetic alterations during colorectal-tumor development. *N Engl J Med*, 1988, 319(9): 525–532.
- [2] Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, 1992, 258(5083): 818–821.[\[DOI\]](#)
- [3] Struski S, Doco-Fenzy M, Cornillet-Lefebvre P. Compilation of published comparative genomic hybridization studies. *Cancer Genet Cytogenet*, 2002, 135(1): 63–90.[\[DOI\]](#)
- [4] Desper R, Jiang F, Kallioniemi OP, Moch H, Papadimitriou CH, Schaffer AA. Inferring tree models for oncogenesis from comparative genome hybridization data. *J Comput Biol*, 1999, 6(1): 37–51.
- [5] Desper R, Jiang F, Kallioniemi OP, Moch H, Papadimitriou CH, Schaffer AA. Distance-based reconstruction of tree models for oncogenesis. *J Comput Biol*, 2000, 7(6): 789–803.[\[DOI\]](#)
- [6] YIN Zhi-Hua, HUANG Zhong-Xi, LIU Teng-Fei, LI Hong, YAO Kai-Tai. Exploration of multigene, multistep and multipathway model of nasopharyngeal and colorectal carcinogenesis. *Chinese Journal of Oncology*, 2004, 26(3): 135–138.  
尹志华, 黄仲曦, 刘腾飞, 李虹, 姚开泰. 鼻咽癌和结直肠癌多基因多阶段多途径模型的探讨. *中华肿瘤杂志*, 2004, 26(3): 135–138.
- [7] Bentz M, Plesch A, Stilgenbauer S, Dohner H, Lichter P. Minimal sizes of deletions detected by comparative genomic hybridization. *Genes Chromosomes Cancer*, 1998, 21(2): 172–175.[\[DOI\]](#)
- [8] Jiang F, Desper R, Papadimitriou CH, Schaffer AA, Kallioniemi OP, Richter J, Schraml P, Sauter G, Mihatsch MJ, Moch H. Construction of evolutionary tree models for renal cell carcinoma from comparative genomic hybridization data. *Cancer Res*, 2000, 60(22): 6503–6509.
- [9] Huang Z, Desper R, Schaffer AA, Yin Z, Li X, Yao K. Construction of tree models for pathogenesis of nasopharyngeal carcinoma. *Genes Chromosomes Cancer*, 2004, 40(4): 307–315.[\[DOI\]](#)
- [10] Brodeur GM, Tsiatis AA, Williams DL, Luthardt FW, Green AA. Statistical analysis of cytogenetic abnormalities in human cancer cells. *Cancer Genet Cytogenet*, 1982, 7(2): 137–152.[\[DOI\]](#)
- [11] Kainu T, Juo SH, Desper R, Schaffer AA, Gillanders E, Rozenblum E, Freas-Lutz D, Weaver D, Stephan D, Bailey-Wilson J, Kallioniemi OP, Tirkkonen M, Syrjakoski K, Kuukasjarvi T, Koivisto P, Karhu R, Holli K, Arason A, Johannesdottir G, Berghthorsson JT, Johannesdottir H, Egilsson V, Barkardottir RB, Johannsson O, Haraldsson K, Sandberg T, Holmberg E, Gronberg H, Olsson H, Borg A, Vehmanen P, Eerola H, Heikkila P, Pyrhonen S, Nevanlinna H. Somatic deletions in hereditary breast cancers implicate 13q21 as a putative novel breast cancer susceptibility locus. *Proc Natl Acad Sci USA*, 2000, 97(17): 9603–9608.[\[DOI\]](#)
- [12] Schaffer AA, Simon R, Desper R, Richter J, Sauter G. Tree models for dependent copy number changes in bladder cancer. *Int J Oncol*, 2001, 18(2): 349–354.
- [13] Huang Q, Yu GP, McCormick SA, Mo J, Datta B, Mahimkar M, Lazarus P, Schaffer AA, Desper R, Schantz SP. Genetic differences detected by comparative genomic hybridization in head and neck squamous cell carcinomas from different tumor sites: construction of oncogenetic trees for tumor progression. *Genes Chromosomes Cancer*, 2002, 34(2): 224–233.[\[DOI\]](#)
- [14] Wu L. Construction of evolutionary tree models for nasopharyngeal carcinoma using comparative genomic hybridization data. *Cancer Genet Cytogenet*, 2006, 168(2): 105–108.[\[DOI\]](#)
- [15] Rieker RJ, Penzel R, Aulmann S, Blaeker H, Morresi-Hauf A, Hecker E, Otto HF, Mechttersheimer G, von Heydebreck A. Oncogenetic tree models based on cytogenetic data: new insights into the development of epithelial tumors of the thymus. *Cancer Genet Cytogenet*, 2005, 158(1): 75–80.[\[DOI\]](#)
- [16] Adamovic T, Roshani L, Chen L, Schaffer BS, Helou K, Levan G, Olsson B, Shull JD. Nonrandom pattern of chromosome aberrations in 17beta-estradiol-induced rat mammary tumors: indications of distinct pathways for tumor development. *Genes Chromosomes Cancer*, 2007, 46(5): 459–469.[\[DOI\]](#)
- [17] Simon R, Desper R, Papadimitriou CH, Peng A, Alberts DS, Taetle R, Trent JM, Schaffer AA. Chromosome abnormalities in ovarian adenocarcinoma: III. Using breakpoint data to infer and test mathematical models for oncogenesis. *Genes Chromosomes Cancer*, 2000, 28(1): 106–120.[\[DOI\]](#)
- [18] von Heydebreck A, Gunawan B, Fuzesi L. Maximum likelihood estimation of oncogenetic tree models. *Biostatistics*, 2004, 5(4): 545–556.[\[DOI\]](#)
- [19] Cremer FW, Bila J, Buck I, Kartal M, Hose D, Ittrich C, Benner A, Raab MS, Theil AC, Moos M, Goldschmidt H,

- Bartram CR, Jauch A. Delineation of distinct subgroups of multiple myeloma and a model for clonal evolution based on interphase cytogenetics. *Genes Chromosomes Cancer*, 2005, 44(2): 194–203. [\[DOI\]](#)
- [20] Gunawan B, von Heydebreck A, Sander B, Schulten HJ, Haller F, Langer C, Armbrust T, Bollmann M, Gasparov S, Kovac D, Fuzesi L. An oncogenetic tree model in gastrointestinal stromal tumours (GISTs) identifies different pathways of cytogenetic evolution with prognostic implications. *J Pathol*, 2007, 211(4): 463–470. [\[DOI\]](#)
- [21] Beerenwinkel N, Rahnenfuhrer J, Kaiser R, Hoffmann D, Selbig J, Lengauer T. Mtreemix: a software package for learning and using mixture models of mutagenetic trees. *Bioinformatics*, 2005, 21(9): 2106–2107. [\[DOI\]](#)
- [22] Beerenwinkel N, Rahnenfuhrer J, Daumer M, Hoffmann D, Kaiser R, Selbig J, Lengauer T. Learning multiple evolutionary pathways from cross-sectional data. *J Comput Biol*, 2005, 12(6): 584–598. [\[DOI\]](#)
- [23] Rahnenfuhrer J, Beerenwinkel N, Schulz WA, Hartmann C, von Deimling A, Wullich B, Lengauer T. Estimating cancer survival and clinical outcome based on genetic tumor progression scores. *Bioinformatics*, 2005, 21(10): 2438–2446. [\[DOI\]](#)
- [24] Carter NP. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet*, 2007, 39(7 Suppl.): S16–21. [\[DOI\]](#)
- [25] ZHAO Guang-Rong, YANG Fan, YUAN Ying-Jin, GAO Xiu-Mei, ZHANG Jun-Ping. Progress in detection methods of single nucleotide polymorphisms. *Hereditas(Beijing)*, 2005, 27(1): 123–129.  
赵广荣, 扬帆, 元英进, 高秀梅, 张军平. 单核苷酸多态性检测方法的新进展. *遗传*, 2005, 27(1): 123–129.
- [26] Adler AS, Lin M, Horlings H, Nuyten DS, van de Vijver MJ, Chang HY. Genetic regulators of large-scale transcriptional signatures in cancer. *Nat Genet*, 2006, 38(4): 421–430. [\[DOI\]](#)
- [27] Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet*, 2006, 38(9): 1043–1048. [\[DOI\]](#)