

Unravelling the Architecture of Duplications in Tumor Genomes

Benjamin Raphael¹

Keywords: Cancer, duplication, comparative genomic hybridization (CGH), genome rearrangements, breakage-fusion-bridge cycles.

1 Introduction

Tumor cells frequently exhibit large-scale changes in their genome, including genome rearrangements (e.g. chromosome inversions and translocations), duplications, and deletions. Such genomic changes are directly implicated in the pathogenesis of cancer and can alter gene structure and regulation. While many individual rearrangements and duplications in tumors have been identified, little is known about the detailed architecture of tumor genomes.

A recently introduced technique called End Sequence Profiling (ESP) allows the identification of both rearrangements and duplications in tumor genomes at high resolution [5, 3]. In contrast, array-based techniques such as comparative genomic hybridization (CGH) are very useful in the identification of duplicated material in a tumor genome, but give little information about the organization and locations of duplicated material within the genome. For example, it is difficult to infer from CGH data whether two duplicated regions are located close together (or even on the same chromosome) in the tumor genome. In some tumors, duplicated material from several disparate regions of the human genome is co-localized [1, 6]. However, the precise architecture of these complex duplications and the mechanisms that explain their formation are not completely understood.

We develop a computational method that allow us to unravel the architecture of complex duplications in a tumor genome using ESP data. The technique relies on a model of *breakage-fusion-bridge (BFB) cycles*, a biological process that produces both rearrangements and duplications in tumor genomes. We apply our method to ESP data from the MCF7 breast tumor cell line. This work complements earlier analyses of genome rearrangements [3] and duplication by amplisome [4] using ESP data.

2 Methods and Results

An ESP experiment consists of the following steps. First, the tumor genome is split into small, overlapping pieces (clones) that vary in size from 100kb to 300kb. Second, the ends (500bp) of each clone are sequenced. Third, the resulting end sequences are mapped to the human genome sequence. Each clone whose end sequences map *uniquely* to the human genome yields a pair (x, y) of locations in the human genome corresponding to the mapped ends. We call such a pair an *end sequence pair (ES pair)*. Thus, the data from an ESP experiment consists of a set of ES pairs $(x_1, y_1), \dots, (x_n, y_n)$.

Typically, the distance between elements of an ES pair will equal the length L of a clone (100-300kb). We call such ES pairs *valid* pairs. However, since the tumor genome is a rearranged version of the human genome, there will also be a number of *invalid* pairs whose ends map far apart or have incorrect orientation. The valid and invalid pairs reveal

¹Department of Computer Science & Engineering, University of California, San Diego. E-mail: braphael@ucsd.edu

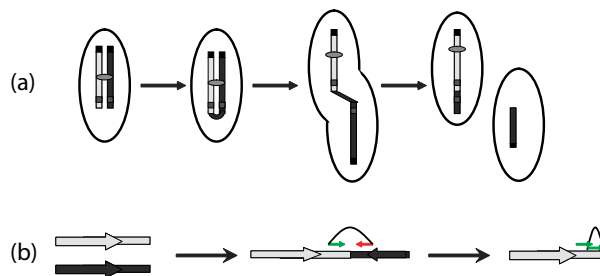


Figure 1: (a) The breakage-fusion-bridge cycle. (b) Single step of BFB results in tumor genome (middle) with mapped ES pair (right).

information about the architecture of the tumor genome. We construct the *ESP graph* from the ESP data, which is similar to the breakpoint graph used in genome rearrangement studies. Edges in the ESP graph correspond to ES pairs, and paths in the graph correspond to possible architectures of the tumor genome. Duplications in the tumor genome result in a complicated ESP graph with multiple overlapping paths due to the interleaving of different copies of duplicated regions.

To unravel this complicated structure, we use a model of the *breakage-fusion-bridge (BFB) cycle* (Figure 1a), first described in maize plants by Barbara McClintock in 1941, and later demonstrated in tumors (cf. [2]). The BFB cycle begins when a chromosome loses its telomere – a protective sequence at the end of the chromosome – by a DNA double stranded *break*. Lacking a protective telomere, the chromosome *fuses* with its sister chromatid (or another unprotected chromosome) forming a loop. During mitosis, the fused chromatids form a *bridge* as the chromatids are pulled apart into separate daughter cells, a structure that is resolved by *breakage* of the fused chromosomes. This cycle of breakage-fusion-bridge continues during subsequent cell divisions and results in a complex series of duplications and rearrangements.

BFB cycles give rise to particular subgraphs in the ESP graph. For example, the signature of one step of BFB is a short ($y - x \leq L$), invalid BES pair (Figure 1b). We use a model of BFB cycles to explain the complex duplications observed in ESP data from the MCF7 breast tumor cell line.

References

- [1] Guan, X. et al. 1994. Identification of cryptic sites of DNA sequence amplification in human breast cancer by chromosome microdissection. *Nature Genetics* 8(2), 155–161.
- [2] Murnane, J. and Sabatier, L. 2004. Chromosome rearrangements resulting from telomere dysfunction and their role in cancer. *Bioessays* 26(11): 1164–1174.
- [3] Raphael, B.J., Volik, S., Collins, C., and Pevzner, P.A. 2003. Reconstructing tumor genome architectures. *Bioinformatics* 19 Suppl 2:162–162.
- [4] Raphael, B.J. and Pevzner, P.A. 2004. Reconstructing tumor amplicomes. *Bioinformatics* 20 Suppl 1:I265–I273.
- [5] Volik, S. et al. 2003. End-sequence profiling: sequence-based analysis of aberrant genomes. *Proceedings of the National Academy of Sciences USA* 100(13), 7696–7701.
- [6] Zhu, C. et al. 2002. Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. *Cell* 109(7): 811–821.