Phylogeny of Tumor Progression from CGH Data

Sven Bilke[1], *Qing-Rong Chen*[1], *Javed Khan*[1]

Keywords: Tumor progression, Comparative Genomic Hybridization, Microarrays

1 Introduction.

Genomic instability, gain or loss of DNA, is frequently observed in tumors [3]. While normal cells contain two copies of each chromosome (with the exception of the gender chromosomes X and Y), tumor cells often loose one or both copies or gain extra copies of specific regions of DNA. The chromosomal location of changes is not random: in many cases instable regions from different patients with the same disease cover the same locations, which are therefore called recurrent regions. It is often argued that, in the competition for nutrition and space, changes at specific locations provide an advantage and are hence more frequent (recurrent) than instabilities at neutral or adverse locations. In this 'micro-evolution' [3, 1] the patterns of genomic instabilities allow to conclude on the inheritance of genomic features and hence the 'phylogeny' of different tumor stages. In this study we use the overlap of the location of genomic instabilities shared between and unique to different phenotypes as indicators for the pathways of tumor progression. To demonstrate the feasibility of this approach we analyze cytogenetic data for three stages of Neuroblastoma the most frequent pediatric childhood tumor, and identify a progression model for this cancer by an exhaustive search in the space of progression models. The data for genomic instabilities used in this study was obtained from 32 neuroblastoma [2] specimens, 12 patients with stage 1, and 20 patients with stage 4 of which 12 were MYCN-amplified (4+) and 8 were MYCN single-copy tumors (4-). Comparative genomic hybridization on a cDNA array with 42000 elements was used to measure the relative DNA copy number. P-values for the presence of gains or losses were estimated by a sliding window method, where the distribution of genome-ordered observations in that window was compared by a t-test to the distribution observed for the full genome.

2 Results and Discussion

Under the assumption that the gains and losses are milestones in the development of the disease, the progression between different stages should be observable in the pattern of genomic instabilities. In order to develop models which can be discriminated by genomic imbalance data, we restrict our analysis to models which follow these biological principles:

1. Unobserved intermediate genotypes are possible but the model with the smallest number of genotypes (observed + unobserved) is utilized
2. All changes found in a parent genotype must be present in the offspring occurring with a similar frequency (the inheritance signature)
3. All tumor stages belonging to the same diagnostic group arise from a common ancestor (i.e. the phylogeny is a rooted tree).

The models compatible with these requirements are discriminated by the patterns of overlap of recurrent genomic instabilities, namely regions *common* to all stages, *specific* to a stage and *shared* between stages. In this feasibility study we analyze a relatively small number of three observed stages, defining seven different subsets which may either be empty or occupied. This allows for $2^7 = 128$ different observations. Not all of those are compatible with the concept of tumor progression in form of a micro-evolution with a specific genetic signature. For example, the case where none of the sets is occupied describes the case where the progression

[1]Oncogenomics Section, Pediatric Oncology Branch, Advanced Technology Center, National Cancer Institute, 8717 Grovemont Circle, Gaithersburg, MD 20877, USA. E-mail: {bilkes,chenqi,khanj}@mail.nih.gov

of neuroblastoma does not manifest itself in a specific signature of genomic instabilities. We identify ten topologically distinct models of tumor progression compatible with the biological assumptions. These models are depicted in the left panel of figure 1. As mentioned in the introduction, no single region is found to be affected in *all* tumors of a specific stage. Also, no single region is affected exclusively in one stage. However, the frequency of their occurance is often significantly different in the different stages of the disease. Therefore the frequency of a genomic instability for the different stages is the primary observable. We define a region to be *specific* to a stage if it is significantly more frequent in one stage as compared to all the other stages. If a region is significantly more frequent in two stages with respect to the third stage, this regions is called *shared*. If a region is frequent in all stages, the region is called *common*. In our analysis of the neuroblastoma-data we have verified that the final result is stable with respect to reasonable changes of the significance thresholds used in defining these regions.

We found recurrent alterations common to all three subgroups, specific for each of the subgroups and common regions of gain for stage 1 and 4- tumors. Interestingly there were no shared alterations of 4+ with 1 or 4- besides the regions common to all. The tumor progression model compatible with these findings is depicted in the right panel of figure 1.

In this study we demonstrate that the pattern of genomic instabilities in neuroblastoma can be used to conclude on the pathway of progression of the disease. The analysis of the frequency of genomic instabilities maps our data onto one of the models compatible with the assumption that 'micro-evolution' governs the progression of cancer. The selected model is in agreement with clinical evidence [4] for neuroblastoma. The identification of the progression pathways for cancer may have an important impact on the strategy for the treatment. In Neuroblastoma our result indicate that the model of a linear progression towards the more aggressive disease (figure 1 (a), sub-diagram I), which is seemingly suggested by the staging system, is not supported by our data. Instead, the best fitting model (right panel of figure 1) suggests that the final outcome of the disease is determined at an very early stage of the cancer development. The stage 4+ disease is fundamentally different from Stages 1 and 4-, also a developed stage 1 tumor does not progress to stage 4-.

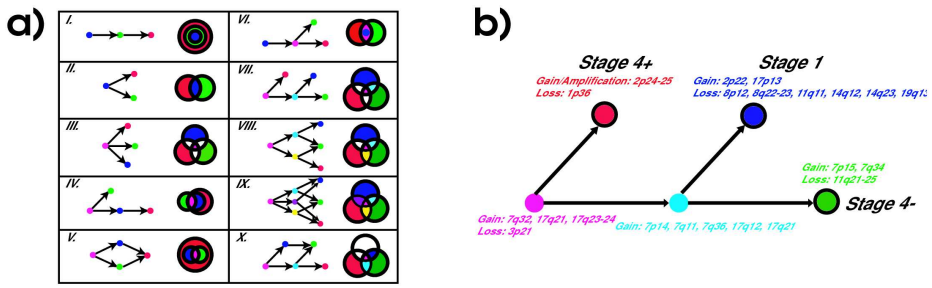

Figure 1: The ten Topologically distinct tumor progression models for three observed stages. Unobserved, intermediate states are drawn in cyan and magenta, observed stages in red, green and blue. b) The selected model summarizing the pattern of recurrent genomic instabilities in our data.

References

[1] For a review see e.g. Bogen K.T. (1989) J. Natl. Cancer Inst. 81, 267-277.

[2] For a review of Neuroblastoma biology see e.g. Brodeur, G. M. (2003) Nat Rev Cancer 3, 203-16.

[3] Rajagopalan, H, et.al. Nat Rev Cancer. 2003 Sep;3(9):695-701.

[4] Westermann, F., Schwab, M. (2002) Cancer Lett 184, 127-47.