Power Law Distribution of Chromosome Aberrations in Cancer

Attila Frigyesi, David Gisselsson, Felix Mitelman, and Mattias Höglund

Centre for Mathematical Sciences, Mathematical Statistics, Lund University [A. F.], and Department of Clinical Genetics, University Hospital, Lund [D. G., F. M., M. H.], Sweden

Abstract

Cancer cells are characterized by having aberrant chromosomes. The number of aberrations and the specific chromosomes affected are correlated with tumor progression. We show that for breast, colorectal, and renal cell carcinomas the distribution of the number of such aberrations per tumor follow a power law distribution with an exponent close to unity. We present two stochastic models that in simulation experiments result in power law distributions of the number of changes per tumor. The first model is based on a multiplicative fluctuation process and the second on a preferential attachment principle linked to an observation process, i.e., a tumor detection and treatment process. Because almost identical power law distributions are seen in breast, colorectal, and renal cell carcinomas we suggest that the obtained distributions are consequences of a common mechanism operating in malignant epithelial tumors.

Introduction

A hallmark in of cancer development is the accumulation of genetic lesions. Some of these may be seen in the form of chromosomal aberrations when the cells enter mitosis. During the past decades a wealth of information on chromosomal changes in cancer cells have accumulated, and based on the collected data two major categories of tumors may be discerned, those with simple and those with complex karyotypes. The former category is dominated by leukemias, the latter by carcinomas. Carcinomas also show a high degree of variability, both regarding the number of changes per tumor and the spectrum of recurrent imbalances (1). Only rarely are balanced primary, disease-specific aberrations seen. An acquired chromosomal instability has been suggested to generate the complex characteristics of these karyotypes (2). In recent investigations (3–5) we have used several statistical tools (6) to analyze the patterns of chromosomal changes in carcinomas. By these methods we were able to identify several general characteristics of carcinoma karyotypes, and it was found that the distributions of the number of imbalances per tumor follow monotonously decreasing functions, indicating that a common process causes most levels of karyotypic complexity. In these investigations we limited the analysis to chromosomal imbalances present above a given frequency. Although this strategy simplified the interpretations, it will, however, affect the shape of the distributions by excluding cases with many changes; the maximum number of imbalances present in a case being limited by the number of imbalances included in the investigation. To eliminate this inadequacy we retrieved all karyotypes of BCs, CCs, and RCCs, from the Mitelman Database of Chromosome Aberrations in Cancer and scored the NAPT by counting the number of entries in each karyotype. We show that the distribution of the number of chromosome aberrations per tumor in all three of the tumor types follows a power law distribution with an exponent close to unity, suggesting that the obtained distributions are generated by a common mechanism.

Materials and Methods

Data, Distribution Fitting, and Simulations. The NAPT was estimated by scoring the number of entries in each karyotype. Although some entries may represent more than one event, e.g., three-way translocations, such cases are rare, and the NAPT measure gives a good estimate of the number of changes present in each tumor. The data set included 667 BC karyotypes with a total of 4,564 aberrations, 533 CC karyotypes with a total of 3,553 aberrations, and 673 RCC karyotypes with a total of 4,574 aberrations. Lognormal, geometric, exponential, simple power law with the probability density function (p.d.f.) \( a \times x^{-a} \) and the Zipf-Mandelbrot distribution with the p.d.f. \( a \times (x+b)^{-a} \) were fitted to the BC, CC, and the RCC data, respectively. Maximum likelihood estimators were calculated for the geometric, exponential, and lognormal distributions, whereas the power law and the Zipf-Mandelbrot distributions were fitted in a least squares sense. The multiplicative fluctuation process was simulated for 100,000 tumors using the model \( \text{NAPT}(t+1) = \text{NAPT}(t) \times \text{Exp}(0.1) \), i.e., the number of aberrations at generation \( t+1 \) is equal to the number of aberrations at generation \( t \) multiplied by a decaying random factor \( \geq 1 \). The preferential attachment process with an observational process was simulated for 100,000 tumors with \( \text{P}[\text{NAPT}(t+1) = \text{NAPT}(t)+1] = \text{NAPT}(t)/30 \), i.e., a tumor progressing from generation \( t \) to generation \( t+1 \) acquires an additional aberration with a probability directly proportional to the number of aberrations at generation \( t \). An exponentially decreasing number of processes were observed and killed with each generation, simulating the clinical course of tumor detection, surgical removal, and cytogenetic analysis. The mean of this exponential process was \( 400 \).

Temporal Analysis. To obtain a value for the time of appearance of a chromosomal change, all of the tumors with the given change were selected and the distribution of the number of changes per tumor plotted. An aberration that is seen frequently in low-complex karyotypes, and, thus, would be defined as early in the karyotypic evolution, will produce distributions with peak frequencies at low values of the number of changes per tumor, whereas changes occurring late in the karyotypic evolution would produce peak frequencies at higher values. Because these frequency distributions often are skewed, the mean is not a good estimate of the TO. Instead the peak values, the modes of these distributions were used as the TO (6). TO is, thus, a function of karyotypic complexity. To estimate how restrained a given aberration is to the estimated TO value, a measure for TO variability would be needed. To obtain such a measure, the selected distributions were resampled with replacement (bootstrapped) 1000 times and the mode scored after each resampling. The 25th and 75th percentiles of the bootstrapped modal values were then used as a measure of TO variability.

Results and Discussion

The relative frequencies of the NAPT values are plotted for each tumor type in Fig. 1. Because the obtained distributions may be described by monotonously decreasing functions with heavy tails, the data were tested for meeting the requirements of an exponential distribution, a geometric distribution, a simple power law distribution, a generalized power law distribution (the Zipf-Mandelbrot distribution), and a lognormal distribution, respectively. The observed data suggesting that the obtained distributions are consequences of a common mechanism.
showed a best fit to the simple power law and the Zipf-Mandelbrot distributions. The estimated $\alpha$ values were 1.25, 1.05, and 0.86, for BC, CC, and RCC, respectively, and the corresponding SDs 0.25, 0.20, and 0.26. Hence, assuming normality, the estimated values for $\alpha$ did not differ significantly. The BC, CC, and RCC data sets were subsequently pooled and the obtained NAPT distribution used to produce a better estimation of $\alpha$. The pooled data showed the best fit to a simple power law distribution with $\alpha \approx 1.05$. These results strongly suggest that the distribution of the number of acquired changes in the carcinomas studied follows power law statistics, $P(NAPT) \sim NAPT^{-\alpha}$ where $\alpha$ is close to unity.

Many seemingly unrelated phenomena show power law distributions, e.g., the size distributions of earthquakes measured by the Gutenberg-Richter scale (7), species extinctions in geological time (8), and commercial firm sizes (9). To explain some of these phenomena, Bak et al. (10) put forward the hypothesis of SOC. According to this idea, complex systems evolve toward a state of maintained criticality at which they are able to propagate perturbations on all of the possible length and size scales, often described as avalanches. The idea that complex karyotypes are caused by a perturbation of a system that has reached a state of SOC through the accumulation of mutations is attractive from a tumorigenetic point of view, with the initiating perturbation causing the avalanche/genome rearrangements being a crucial genetic alteration e.g., telomere dysfunction (11). An alternative explanation may be that cells, and evolving systems in general, are prone to show HOT (12). In this setting, systems are tuned, through selection, to highly structured and efficiently operating states within a given environment. This is partly achieved by generating barriers that reduce the effects of cascading failures caused by perturbations encountered frequently during the evolutionary process. However, rare disturbances may lead to dramatic consequences. The results of such failures show power law distributions with exponents close to unity (12). In this context, the complex karyotype would be caused by a cascading failure in a system showing HOT. However, the avalanche dynamics of the systems usually associated with SOC and HOT are rapid, compared with the driving processes. That is, the system dynamics are slow but will eventually result in a state where a rare single event may, at a single time, cause a large effect i.e., a large number of aberrations. This makes the suggested models poorly compatible with the current notion of tumor development; tumor progression rather implicates growth and the successive acquisition of changes that together form a functional pathogenic system (13, 14).

On the other hand, a variety of stochastic growth processes have been shown to converge to power law distributions (15–17). One such process is multiplicative fluctuations (18). In this setting each karyotype passes through a series of updating steps (tumor progression steps) at which the NAPT changes with a fraction $\alpha$ of its value, i.e., $NAPT(t+1) = \xi NAPT(t)$, where $\xi$ is a positive definite random variable larger than one. In a logarithmic space this equates to $\log[NAPT(t+1)] = \log(\xi) + \log[NAPT(t)]$. This implies that for $t \to \infty$ 

Fig. 2. Results from the simulation of (A) a multiplicative fluctuation process and (B) a geometrical growth process caused by preferential attachment and an exponentially distributed observation/killing process. The simulated data have a best fit to a power law distribution with $\alpha = 1.11$ (A) and with $\alpha = 1.07$ (B). The insets show the respective data in log-log plots.

Fig. 1. The distributions of the NAPT in (A) breast carcinomas, (B) colorectal carcinomas, (C) renal cell carcinomas, and (D) in the pooled data set. The insets show the respective data in log-log plots.
the distribution of log[NAPT] approaches a uniform distribution, and transforming back to linear space gives \( f(P(\log[NAPT])) \) = \( C / NAPT \) \( -1 \) \( dNAPT \), where \( C \) represents the normalization factor. This gives \( P(\log[NAPT]) \sim \) \( NAPT \) \(-1\) for the distribution of NAPT. Simulations from such a multiplicative fluctuation model do indeed show power law behavior with \( \alpha \) close to unity (Fig. 2A).

An alternative process is exponential or geometrical growth that is observed in a stochastic fashion (19). One such geometrical growth process is preferential attachment (20). In the present context preferential attachment would mean that the probability to acquire a chromosomal aberration increases with the number of aberrations already present. In conjunction with tumor detection (observation), this process is expected to produce a power law distribution (19). Indeed, simulations using a linear increase in probability to acquire a new aberration and an exponentially distributed observation process do produce power law distributions (Fig. 2B).

There are at least two possible interpretations of preferential attachment in the context of karyotype evolution. One interpretation implies that the frequency by which the tumor cell is presented with new aberrations increases with the number of changes already present but that the probability to accept a given aberration is constant, and, thus, entails an increasing genomic instability. Biologically this interpretation would correspond to a mutator phenotype (21) with an increasing mutation rate during tumor progression. The second interpretation implies that the frequency by which the cell is presented with new aberrations is constant, but that the probability that a given aberration is accepted as part of the karyotype increases with the number of changes already present, i.e., that an increasing number of alternative changes may be accepted as part of the karyotype. This interpretation does not necessarily include the notion of a mutator phenotype (22) but rather an evolving and an increasingly permissive tumor microenvironment.

The two alternatives of preferential attachment both include the concept of a stepwise addition of chromosomal changes. Given this process, it would be possible to determine whether a specified aberration is early or late in the karyotypic evolution by investigating its frequency in karyotypes with different numbers of aberrations. In Fig. 3A we have investigated the lateness of the loss of chromosome 9 (–9) in nonpapillary RCC by producing the distribution of the number of changes per tumor of the cases with this imbalance. We have used previously the modal values of such distributions as an approximation of the TO of that particular change (3–6). By estimating the TO for all of the changes seen in a tumor type it is possible to model the temporal order by which the aberrations are acquired (3, 6). To obtain more reliable estimates of TO, the original distributions were resampled 1000 times with replacement and the TO calculated after each resampling. The 25th and 75th percentiles of the bootstrapped TO values were then used to determine at which karyotypic progression steps a change is most likely to occur (Fig. 3B). The temporal orders established in this way have been shown to correlate well with histopathological data and are, thus, biologically relevant (4, 5). TO values that overlap in Fig. 3B signify that the respective change may occur at the same progression steps. By examining the number of alternative changes that may occur at each step in nonpapillary RCC (Fig. 3B) it can be concluded that the increase from one to six abnormalities involve 0, 1, 3, 5, 9, and 10 alternative changes at each subsequent step. In Fig. 3C these data have been plotted together with data obtained in a similar way for the cytogenetic pathway in papillary RCC (3), the major pathway operating in BC (4), and the two cytogenetic pathways operating in CC (5). It may be seen that an increasing number of alternative changes may occur at each subsequent step, supporting the concept that the frequency by which the cell is challenged with a new aberration is constant, but that the criteria for keeping any given aberration becomes less stringent when the number of imbalances increases.

The presented results point to an important aspect of carcinogenesis, the power law distribution of the number of chromosomal changes per tumor. We suggest that the process generating this distribution is based on a common mechanism as it may be seen in more than one tumor type. Furthermore, we have presented two simple stochastic models, with biologically plausible interpretations, that may account for the described scale-free behavior of the number of acquired chromosome changes in cancer.

Acknowledgments
We thank Prof. Nils Mandahl for valuable comments on this article.

References
3. Höglund, M., Gisselsson, D., Mandahl, N., Johansson, B., Mertens, F., Mitelman, F., and Sáll, T. Multivariate analyses of genomic imbalances in solid tumors reveal...
Power Law Distribution of Chromosome Aberrations in Cancer

Attila Frigyesi, David Gisselsson, Felix Mitelman, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/63/21/7094

Cited articles
This article cites 20 articles, 7 of which you can access for free at:
http://cancerres.aacrjournals.org/content/63/21/7094.full#ref-list-1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.