SPANDIDOS PUBLICATIONS

Development of a phylogenetic tree model to investigate the role of genetic mutations in endometrial tumors

GUOYI ZHANG[1], BRANDON B. BECK[3], WENTAO LUO[2], FAN WU[3], STEPHEN F. KINGSMORE[4] and DONGHAI DAI[3]

Departments of [1]Mathematics and Statistics, and [2]Internal Medicine, University of New Mexico, Albuquerque, NM 87131; [3]Department of Obstetrics and Gynecology, University of Iowa Health Care, Iowa City, IA 52242; [4]National Center for Genome Resources, Santa Fe, NM 87505, USA

Received October 25, 2010; Accepted December 14, 2010

DOI: 10.3892/or.2011.1186

Abstract. With the advancement of modern genome sequencing technology, thousands of genetic mutations have been identified in human tumors. However, analysis of the role of genetic mutations in tumor development is limited by the need for prevalence information among multiple tumors and by the lack of analytic capability to define the functional contribution of genetic mutations in patients, individually and collectively. To understand the genetic basis of human endometrial cancer, the fourth most common cancer in women, transcriptome sequencing was performed on an endometrial tumor paired with normal cervical tissue. Twenty-six non-synonymous somatic mutations were validated in the tumor genome. A phylogenetic tree illustrating the mutational time-line was developed based upon the distribution of 26 mutations in 30 randomly-selected laser-captured single cells from the tumor sections. Five ubiquitous mutations were identified that are presumed to occur in the cancer founder cell of the tumor, and may collectively play critical roles in endometrial oncogenesis. However, further testing in 10 additional endometrial tumors failed to show overlapping mutations in the cancer founder cells, indicating the lack of a single common oncogenic pathway for these endometrial tumors. The effects of individual mutations in cancer cell proliferation were calculated based on descendant cell number and time span since acquiring each mutation. We have developed a phylogenetic approach to characterize individual genetic mutations in cancer cell proliferation in a single resected patient tumor. This approach provides the capability to study the tumor-specific role of genetic mutations, without relying on prevalence information from other patients.

Correspondence to: Dr Donghai Dai, Department of Obstetrics and Gynecology, University of Iowa Health Care, Iowa City, IA 52242, USA
E-mail: donghai-dai@uiowa.edu

Key words: endometrial cancer, genetic mutations, phylogenetic analysis

Introduction

Tremendous progress has been made in cancer genetics in the past several decades. A seminal model was proposed in colorectal carcinoma involving sequential occurrence of mutations in several genes critical for cellular functions as the causal events for oncogenesis (1,2). Subsequently, it has become widely accepted that most tumors are monoclonal in origin (3-5), and that the transformation of the cancer founder cell (CFC) requires multiple genetic changes (6). Recent advances in DNA sequencing technologies have identified thousands of genetic mutations in tumors (7-13). Additionally, the relative roles of these numerous mutations in oncogenesis have been increasingly recognized as etiologically complex (8). The first comprehensive study of cancer exon sequencing, reported by Sjoblom *et al*, found an average of 67 somatic mutations per breast tumor and 52 per colorectal tumor (7). This pioneering study and subsequent similar studies in pancreatic cancer and glioblastoma represent innovative efforts to substantiate the mutational and monoclonal model of oncogenesis proposed 20 years ago (2).

Endometrial cancer is the fourth most common cancer in women (14) whose development is believed to follow an oncogenic pathway similar to the paradigm established in colorectal cancer. Analysis of the role of a genetic mutation in a patient tumor often relies on the mutation prevalence in other patients and on investigation of its cellular function from *in vitro* and/or animal studies. To our knowledge, there is no analytical approach that will allow a direct analysis of the function of individual genetic mutations on cancer cell proliferation in human endometrial tumors. We hypothesize that results from a specific mathematical analysis of genetic mutations among sampled cells in individual endometrial tumors characterizes genetic aspects of oncogenesis and tumor progression. A phylogenetic tree model was developed to guide the mathematical analysis of massively parallel sequencing data that nominates candidate driver mutations in the cancer cells. Furthermore, an attempt was made to calculate a mutation's tumor-specific contribution to cancer cell proliferation based on the size of offspring produced by an ancestor cell and time span since acquiring the mutation.

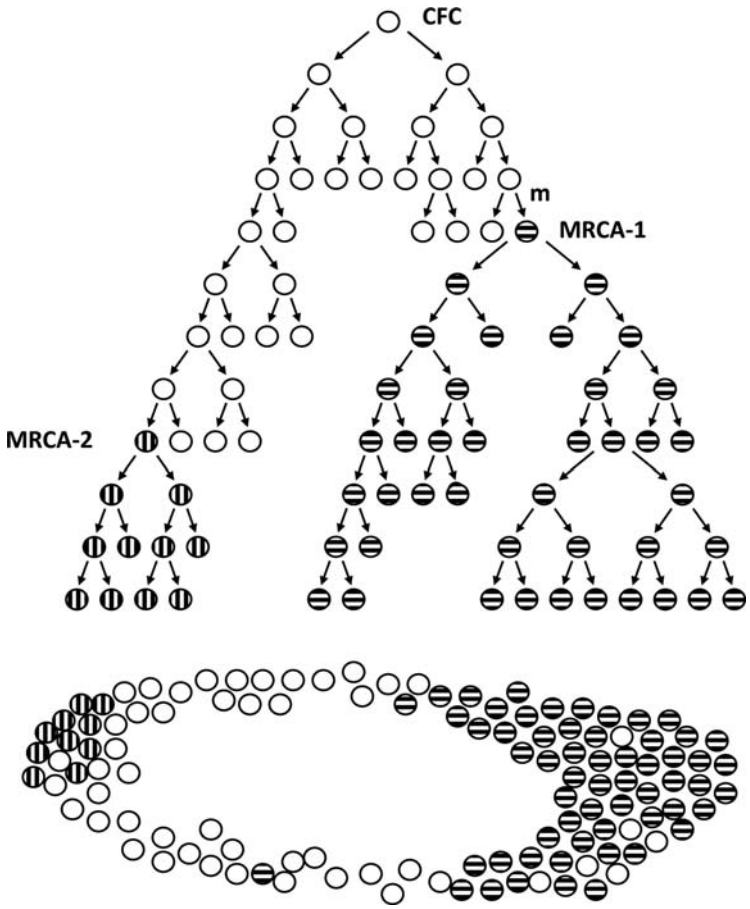

Figure 1. Schematic illustration of a tree model with imputed timeline of tumor development. The cancer founder cell (CFC) was transformed by multiple mutations. The horizontally striped cells represent a dominant clone in the primary tumor with their most recent common ancestor (MRCA-1) carrying many more mutations than the CFC, and most of these mutations are not involved in oncogenesis due to their later occurrence. Sampling of a small piece of tumor tissue (vertically striped cells) could mistake an MRCA (MRCA-2) of a small subpopulation for the CFC, resulting in reporting of many more passenger mutations as the founder mutations, as well as missing those mutations (m) important for the emergence of a dominant subclone.

Material and methods

DNA and RNA preparation. Tissue samples were obtained from anonymous, adult females using guidelines approved by HRRC at the University of New Mexico. The endometrial tumor and normal cervix were collected fresh after surgery and cut into many 5-mm pieces for snap-freezing in liquid nitrogen. mRNA was extracted using RNeasy Mini kit from Qiagen, Valencia, CA. Genomic DNA was extracted for validation of variants from genome sequencing. Frozen sections were cut at 7 μm. Single cancer cells were acquired using Arcturus PixCell IIe with guidance of hematoxylin and eosin staining of adjacent sections. Generally, we acquired 2-3 cells/section and 3-5 sections/tumor piece. Genomic DNA from single cells was amplified separately using GenomePlex Single Cell Whole Genome Amplification kit (Sigma, St. Louis, MO). PCR was completed usually with 30-35 cycles using TC-3000 Thermal Cycler (Barloworld Scientific Ltd., Burlington, NJ). Sanger sequencing was accomplished by ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA).

Transcriptome sequencing. Short-insert, paired-read libraries were generated from mRNA as described (15). Singleton 36 nucleotide reads were generated using Illumina GAII instruments as described (15). Sequences were used in analyses if average quality (Q) scores were >20, respectively. Sequences were aligned to the NCBI human reference, version 36.2, with GSNAP (16). SNVs were identified using optimized filters with the Alpheus software system (11,15,17). Putative SNVs were validated by targeted Sanger PCR and cycle sequencing. Statistical analysis was performed using JMP-Genomics (SAS Institute, Cary, NC).

Construction of the phylogenetic tree. Genetic trees are constructed according to a divisive hierarchical clustering method and verified as one of the trees generated by the maximum-parsimony method (16,18,19) using pars from the PHYLIP software package (Felsenstein, Department of Genetics, University of Washington, Seattle, Version 3.68). The clustering method is based on the effects of the ability of mutations to increase growth rates or confer selective advantage, which are evident when single cells are randomly sampled from a tumor and any reproductive advantage from a mutation is represented by the number of cells from the sample that contains the mutation. The clustering method recursively chooses the most frequent mutation among a cluster of cells and divides the cluster into sub-clusters of cells with and without that mutation.

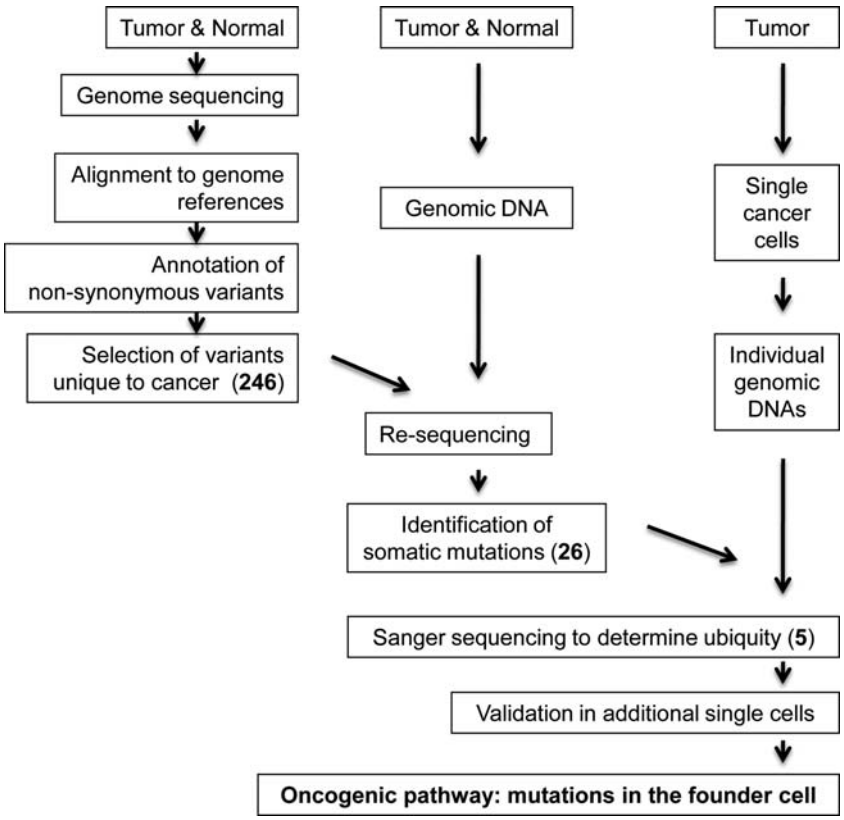

Figure 2. Schematic of cancer genetic timeline analysis. The boxes and arrows on the left indicate the discovery test in normal and tumor transcriptomes; the boxes and arrows in the middle indicate the validation test in normal and tumor genomes; and the boxes and arrows on the right indicate the distribution test in single cell genomes.

Results

A phylogenetic model is developed to describe intra-tumor mutational heterogeneity. The model (Fig. 1) is a schematic reconstruction of the classical monoclonal, step-wise mutational model of tumor progression, as originally developed for colorectal oncogenesis (1,2), assuming a central role for genetic mutations in tumor development. In the simplest form, we assume the cancer founder cell (CFC) at generation zero has acquired, for example, 5 mutations. We further assume a net constant proliferation rate for every cancer cell with acquisition of one new mutation every generation, based on the estimated net somatic mutation rate of $4.6x10^{-10}$/bp/generation (6). Therefore the presumed cell count at the n-th generation of tumor development will be 2^n with total mutations of $2^{n+1}+3$. Thirty generations of tumor development, for example, equates to over $1x10^9$ cells (~1 g of weight) and confers >30 billion mutations. Typically, however, the sensitivity to detect these mutations is ~25% of a population of cells for Sanger sequencing (6), unambiguously detecting only 1 mutant copy in the presence of 3 normal copies. Thus, Sanger sequencing using DNA from tumor homogenate can only reliably detect mutations that occurred in the CFC (5 mutations) and in the first and second generations (2 and 4 mutations, respectively). Any mutation occurring beyond the third generation will be present at <10% of the cell population and will be extremely difficult to detect by Sanger sequencing of DNA homogenates. Again, this is the simplest model assuming constant proliferation and mutation rate without considering cell death during tumor development. It should serve as a basic description before incorporating varying proliferation and mutation rates, cell death and interaction with the microenvironment.

Deep transcript sequencing identifies the mutations uniquely derived from the cancer founder cell. To overcome the limitations of Sanger sequencing, we used massively parallel transcriptome sequencing (mRNA-seq) as part of a cell-ontology-based analysis strategy, cancer genetic timeline analysis (CGTA, Fig. 2) with the intent of using deep sequencing to both detect and enumerate the frequency of expressed mutations, including those present in minority subpopulations of a tumor. RNA from matched normal and tumor specimens from an endometrial cancer patient was sequenced by mRNA-seq, yielding approximately 1 billion nucleotides of 36 bp singleton reads from each (15). Using the Alpheus pipeline, ~80% of reads aligned to the NCBI human genome reference and are available at http://citrine.ncgr.org/ (11,15,17). Bioinformatic filtering identified 246 non-synonymous single nucleotide variants (nsSNVs) in the tumor transcriptome that were not called in mRNA from coisogenic, normal tissue (11,15). These variants were re-sequenced in genomic DNA from coisogenic normal specimens and multiple tumor specimens from the same patient, and 26 were validated to be somatic mutations in the tumor genome. Next, thirty single cancer cells were acquired through

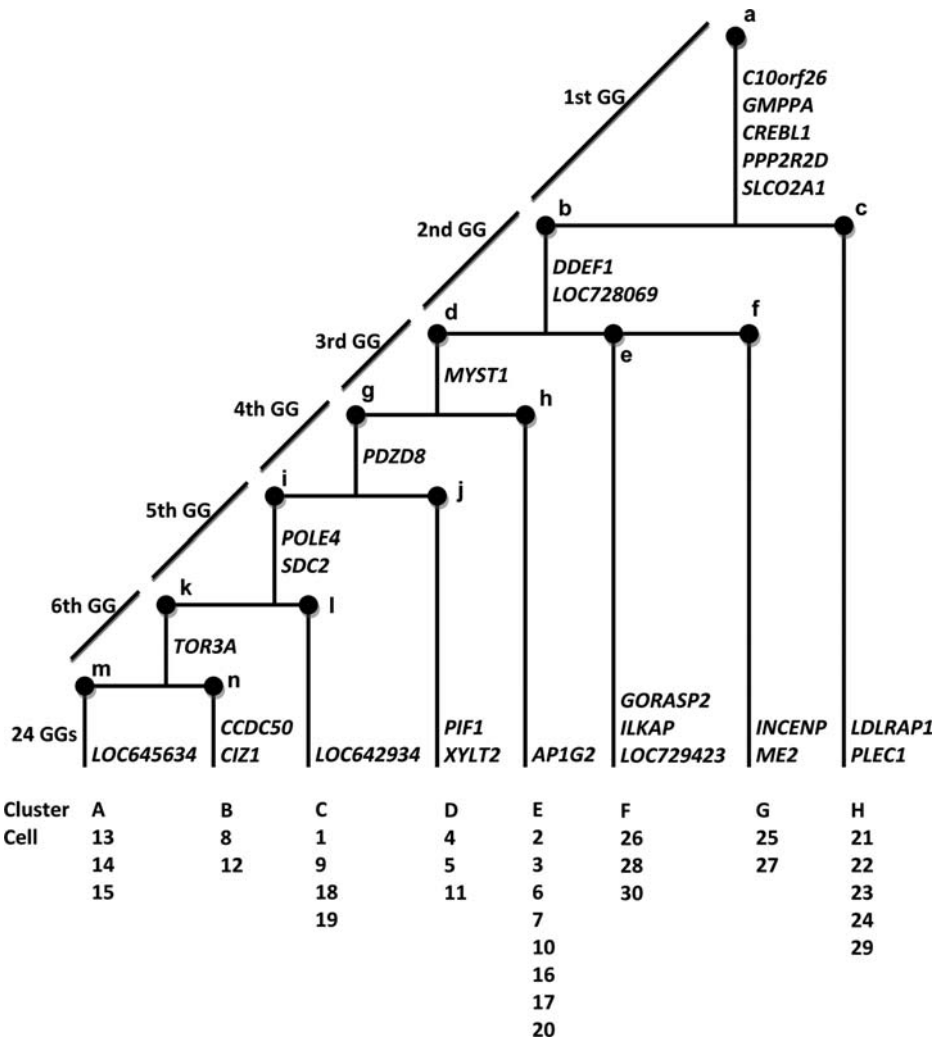

Figure 3. The phylogenetic tree of an endometrial tumor. The tree was constructed based on the distribution of mutations among individual cancer cells. See Materials and methods for details of tree construction. The mutated genes common to descendants are listed along the tree branches following the tree knots (dots, designated a through n) representing the most recent common ancestors (MRCAs) imputed to acquire these mutations at the time of their birth. The numbers at the bottom are the cell numbers we assigned when these single cancer cells were captured by LCM. Genetic generation was defined to substitute for chronological time to determine the relationship between occurrence of new mutations and descendant cell numbers.

laser-captured micro-dissection from disparate frozen sections of the tumor. The genomic DNA of each cell was extracted and amplified separately. The 26 mutated genes were re-sequenced and 5 mutations were found to be present in all 30 single cells from the tumor (Fig. 3, Table I). By virtue of ubiquity, these 5 mutations were considered to occur at the CFC of the tumor and to be responsible for early events in oncogenesis. The following arguments infer the existence of the cancer founder cell: i) the occurrence of 5 mutations in a single normal cell (i.e., passenger mutations) has an extremely low probability; ii) the likelihood of any two cancer cells acquiring the same mutation *de novo* is minimal; iii) 5 ubiquitous mutations in a specimen are strong evidence for clonality (the existence of a sole progenitor cell of the tumor); and iv) the cancer progenitor cell is likely to have had a transformed phenotype and resultant mitotic advantage caused by the 5 mutations. We suggest these 5 mutations, the set of mutations imputed to exist in the CFC, as the oncogenic pathway for the tumor. Some of these 5 mutations have been implicated in oncogenesis of other cancers (Table I). This conclusion has to rely on the assumptions that transcriptome sequencing detects all potential mutations.

Construction of the tumor's phylogenetic tree establishes the timeline of mutational events. We used data from CGTA to reconstruct a phylogenetic tree for the tumor based on the distribution of 26 genetic mutations in 30 single cancer cells in order to determine their temporal occurrence. Based upon the arguments that genetic mutations are inheritable and the likelihood of any two cancer cells acquiring the same mutation *de novo* is negligible, we established the temporal order of the 26 genetic mutations in clusters (Fig. 3). For instance, the occurrence of mutations in Cell 13 will be as such: the earliest mutations are at *C10orf26*, *GMPPA*, *CREBL1*, *PPP2R2D*, *SLCO2A1*, followed sequentially by *DDEF1*, *LOC728069*→*MYST1*→*PDZD8*→*POLE4*, *SDC2*→*TOR3A*→*LOC645634*. Based on the distribution of 26 mutations, these 30 cells were grouped into 8 clusters from A through H, each of which has a distinct phylogeny of genetic mutations with their respective imputed most recent common ancestors (MRCAs,

ist of the somatic mutations in the cancer founder cell.

Gene	Chrom[a]	Nucleotide change	AA change	Zygosity change in tumor	mRNA expression: tumor/normal[b]	Function	Association with cancer
SLCO2A1	3q21	c1269t	P→S	Het→Hom	0.62	i) Transports PGE2 and estrone 3-sulphate; ii) regulates decidualization of endometrial stroma	i) Decreased in colon cancer; ii) PGE2 promotes cancer progression
OPAL1 (C10orf26)	10q24	t1164c	S→P	Het→Hom	0.69	Not known	Prognostic factor in ALL
PPP2R2D	10q26	g1216a	G→S	Het→Hom	1.33	Modulator of TGF-ß/ activin/nodal signalling	
GMPPA	2q35	g525a	G→S	WT→Het	0.87	Converts mannose-1-phosphate to GDP-mannose during production of N-linked oligosaccharides	
ATF6B (CREBL1)	6p21	g1320c	K→N	WT→Het	1.76	Transcription factor in unfolded protein response pathway during ER stress	

[a]Chrom, chromosome. [b]Digital gene expression ratio derived from mRNA-seq.

the branch dots designated a through n). The mutation at *LOC645634* is expected to occur in Cell 13, >6 generations after the CFC. A significant increase in the depth of sequencing and the number of single cells for analysis may produce a more detailed phylogenetic tree.

Individual endometrial tumors have distinct oncogenic pathways. To investigate whether an oncogenic pathway is shared between endometrial tumors, we re-sequenced these five mutated genes from the CFC in genomic DNA from 10 additional endometrial tumors (7 endometrioid carcinoma and 3 uterine papillary serous carcinoma). No more than two of the five mutations occurred in the cancer founder cells in any of the 10 tumors, suggesting that oncogenic pathways are distinct among individual endometrial tumors. Our finding is consistent with the reports in colorectal, breast, pancreatic cancer and glioblastoma (7-10), which collectively provide strong evidence against the notion of a single oncogenic pathway for human cancers, even of a single histologic type. Thus, it remains possible that multiple oncogenic pathways may exist in endometrial cancer, and CGTA could be used as a method to identify oncogenic pathways even if they are distinct in every individual tumor.

Mathematical analysis may characterize the role of genetic mutations in tumor development. In addition to identification of the oncogenic pathway of a tumor, CGTA can utilize deep-sequencing genomic data to determine early mutations using the constructed phylogenetic tree. According to the phylogenetic tree presented in Fig. 3, CGTA infers the existence of 'genetic generations' (GG). For Cluster A, 6 GGs have passed when a cell (m) was borne with newly acquired mutations at *LOC645634*. More specifically, mutations at *DDEF1* and *LOC728069* occur when a cell (b) was borne one GG after the CFC. Likewise, mutations at *MYST1*, *PDZD8*, *POLE4* and *SDC2*, and *TOR3A* occur in cells borne 2, 3, 4 and 5 GGs after the CFC, respectively. In our case, the GGs beyond the 6th GG cannot be documented due to the limitation in sensitivity of genome sequencing. One genetic generation is defined as the shortest temporal interval from a parent to a descendant cell containing at least one unique mutation not present in the parent. Thus the value of GG can be converted into an estimate of the true physical time interval if mutation rate is known. The number of genetic generations and the number of mitotic generations (divisions) are equivalent if one or more mutations occur before each cell division, but otherwise the number of mitotic generations is greater than the number of genetic generations. We calculate the number of net mitotic generations to be an average of 30 based upon the tumor size of 1 g, which is approximately equal to 2^{30} (~10^9) cells. Again, this calculation is a backward reconstruction (from a resected tumor to the CFC) and will treat dead cells during tumor evolution as non-existent since they cannot be postulated from the surviving cancer cells in the resected tumor. The estimated mutation rate of 4.6×10^{-10} per base pair per generation (6) indicates at least one mutation per mitotic division for a genome size of 3×10^9 base pairs, and so the total number of GG is estimated to be 30 (equal to the number of mitotic generations). While the total number of cells in a tumor can be estimated according to the tumor

Table II. Proliferation index, offspring size and genetic generations of cancer cells carrying newly acquired mutations.

MRCAs	Mutant genes	Genetic generations	Offspring size (million cells)	Proliferation index	Difference
a	*C10orf26*, *GMPPA*, *CREBL1*, *PPP2R2D*, *SLCO2A1*	30	1000	1	1
b	*DDEF1*, *LOC728069*	29	833	1.022	0.025
c	*LDLRAP1*, *PLEC1*	29	167	0.942	-0.055
d	*MYST1*	28	667	1.047	0.025
e	*GORASP2*, *ILKAP*, *LOC729423*	28	100	0.949	-0.073
f	*INCENP*, *ME2*	28	67	0.929	-0.094
g	*PDZD8*	27	400	1.058	0.011
h	*AP1G2*	27	267	1.037	-0.010
i	*POLE4*, *SDC2*	26	300	1.083	0.025
j	*PIF1*, *XYL2*	26	100	1.022	-0.036
k	*TOR3A*	25	167	1.093	0.010
l	*LOC642934*	25	133	1.080	-0.004
n	*CCDC50*, *CIZ1*	24	67	1.083	-0.010
m	*LOC645634*	24	100	1.107	0.014

Values of proliferation index were presented for imputed cancer cells (MRCAs, the first column) carrying newly acquired mutations (listed in the second column). The values of genetic generations for the imputed MRCAs are defined as the number of genetic generations from the birth of an MRCA to the end-point (tumor excision). In practice, the value for an MRCA is expressed as the difference between the value of GG for the CFC and the number of GG from the birth of the CFC to the birth of the MRCA. The descendant size was the product of total cell number of the tumor, which is ~1 billion, and the percentage of descendant single cells among the total of 30 single cells. The calculation of proliferation index and difference of proliferation indices between two immediate MRCAs are described in detail in the text.

weight, the number of offspring cells from an ancestor cell cannot be directly determined in practice. In our case, the number of single cancer cells acquired through laser-captured microdissection is used as a representation of cancer cell subpopulation.

Specific mutations in an ancestor cell resulting in a indirectly measurable subpopulation over a defined period of time suggests one method for attributing increased or decreased proliferation to the mutations introduced in the ancestor cell. The discrete first derivative with respect to time of the logarithm of the subpopulation size serves as a metric of the proliferation potential (PP) of an ancestor cell dividing mitotically at regular intervals. For the time period from the birth of the ancestor cell to the end-point (time of tumor excision), an average of the PP is given by the ratio of cell number logarithm to time period. However, since the actual chronological time t cannot be determined, we substitute t with GG and provide an expression of proliferation index (P_i):

$$P_i = (\log_2 N)/g$$

where P_i is the value of P_i, and g is the number of GG from the birth of a cell to the end-point. Since the g value of a cell cannot be measured directly through experimentation due to limitations of detection of mutations in later generations (present only in small populations), the value of g of a cancer cell (from the birth of the cell to the end-point) can be expressed as the difference between the value of GG for the entire tumor (from the birth of the CFC to the end-point), which is 30 in our case, and the number of GG from the birth of the CFC to the birth of the cell, which can be calculated using the phylogenetic tree. The P_i values associated with imputed MRCAs with newly acquired mutations are given in Table II as an example of this method of quantification of proliferation potential.

Since proliferation potential is determined by intrinsic genetic alterations and a cell's interaction with the external environment, the contribution of a newly acquired genetic mutation to cell proliferation should be the difference in cell proliferation potentials between the cell and its immediate progenitor. Therefore, as shown in Table II, the 5 mutations at the CFC produced substantial acceleration of cell proliferation with an increase of 1 in P_i value (the P_i value for the normal progenitor cell is assumed to be 0). Other mutations, such as those at *DDEF1* and *LOC728069*, *MYST1*, as well as *POLE4* and *SDC2*, conferred about 0.025 increase of P_i value over the previous generation, and thus can be seen to have a positive growth effect. In this model, our calculation predicts the existence of passenger mutations, which do not contribute significantly to cancer cell proliferation (without a significant increase in P_i value), and of mutations which negatively affect cancer cell proliferation. For instance, mutations at *GORASP2*,

and *LOC729423* induce a change of P_i value at Cell e (Fig. 3, Table II). Similarly, mutations at *INCENP* and *ME2* induce change of P_i value by -0.094 at Cell f. These data suggest that mutations at these five genes may either cause partial cell death or reduced proliferation, which is consistent with views that mutations could be deleterious or advantageous to a cancer cell while most of them are essentially neutral (20).

Discussion

CGTA, using deep transcriptome sequencing and cell ontology analysis to determine cancer founder cell mutations, represents an alternate strategy to identify oncogenic pathways in individual tumors that is complementary to approaches that rely on gene prevalence information from multiple tumors. However, two major hurdles remain before realization of the full potential of CGTA. The first hurdle is the need for a rational tumor tissue collection procedure. In practice, selection of a piece of tumor for sequencing is often neither comprehensive nor random. If we assume that cancer cells do not move significantly at the primary location during tumor development, selection of a small piece of tumor represents isolation of a subclone of the cancer population. In some advanced tumors, such as endometrial and ovarian carcinoma, the entire tumor could weigh more than hundreds of grams and could be scattered in many places. Thus, a small portion of tumor could constitute a small percentage of the total cancer cell population. As illustrated in Fig. 1, such biased sampling would result in mistaking a most recent common ancestor (MRCA-2) of the small sample as the CFC of the entire tumor. We would recommend multiple biopsies of many parts of a tumor as an important and necessary first step for biospecimen banking and for application of the CGTA method, which was also recommended as a procedure to document intra-tumor heterogeneity by Merlo *et al* (20).

The second hurdle for CGTA is to achieve sufficient depth of sequencing for reconstruction of a comprehensive tree, with a resolution much higher than 6 generations presented in Fig. 3. Whole genome sequencing (13) will presumably detect all potential mutations, and will become financially feasible for many single cells as it becomes more affordable. Alternatively, direct whole genome sequencing of multiple single cells could serve the same purpose (20) although the quality of single genome remains a challenge.

Using CGTA, we have illustrated how a phylogenetic tree of a tumor can reconstruct the tumor's genetic progression and mutational distribution. However, a robust mathematical approach is needed to determine the role of individual genetic mutations in patients. We developed a criterion, proliferation potential (PP), to describe the effect of a genetic mutation on a cell's potential to produce offspring. A simple mathematical expression for PP is defined using cell number and time, and applies both to the entire tumor and to its subclones. Theoretically, cell number is well-defined and measurable, but in practice, the total cell number of a tumor or a piece of tumor can only be approximated by the tumor weight, and the relative cell number of various subclones can only be approximated by acquisition of single cells through laser-captured microdissection. The biggest challenge, however, is the accurate measurement of the other variable, chronological time, to calculate PP for a cell. In most cases, the tumor excision end-point is the only time-point available from a patient and will not help to determine the birth time (starting time) of various subclones. Thus, a surrogate or approximation has to be developed to determine the relative lifespan of cancer cell subclones, and for this reason we introduced the concept of genetic generation (GG). CGTA allows construction of a phylogenetic tree, and thus provides an objective measurement of GG. The drawback of such an approach is the limited capability to define the role of single mutations. For instance, as shown in Fig. 3, there are several branches where a genetic generation includes multiple mutations as a set. Our phylogenetic tree cannot determine whether these mutations occur during one cell division or in multiple divisions, since cell death or limited single-cell sequencing will result in the collapse of multiple GG into one GG. Thus, molecular tumor clocks are needed to translate genetic markers, such as genetic mutations and epigenetic alterations, into chronological ages (21,22). Accurate documentation of time (chronological time) will create another fundamental and objective measurement to study tumor evolution. In our case, GG is developed as a surrogate for chronological time and is expected to be useful in the determination of proliferation index (P_i), defined as the relationship between the temporal occurrence of mutations and resultant number of descendant cells. While our description of this model has obviously overemphasized the role of genetic mutations in cancer cell proliferation, other factors can be represented as well since P_i calculation can be used to estimate the net effect of the influence of multiple intrinsic and extrinsic factors. Overall, this approach provides an important and basic mathematical analysis for identification of mutations with substantial effect on cancer cell proliferation.

We have developed a phylogenetic approach to establish the timeline of mutational occurrence and characterize individual genetic mutations in cancer cell proliferation in resected patient tumor. This approach provides the capability to study the tumor-specific role of genetic mutations in cancer cell proliferation, without relying on prevalence information from other patients and functional study in cancer cell lines. Most important of all, it may have potential to study gene function in a living cancer patient.

Acknowledgements

This publication was made possible by a Grant from American Cancer Society RSG-06-105-01-CCE (DD), and a grant from the National Center for Research Resources (NCRR), a component of NIH P20 RR 016480 (SFK).

References

1. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM and Bos JL: Genetic alterations during colorectal-tumor development. N Engl J Med 319: 525-532, 1988.
2. Fearon ER and Vogelstein B: A genetic model for colorectal tumorigenesis. Cell 61: 759-767, 1990.
3. Fearon ER, Hamilton SR and Vogelstein B: Clonal analysis of human colorectal tumors. Science 238: 193-197, 1987.
4. Fialkow PJ, Gartler SM and Yoshida A: Clonal origin of chronic myelocytic leukemia in man. Proc Natl Acad Sci USA 58: 1468-1471, 1967.

5. Weinberg RA: The Biology of Cancer. Garland Science, New York, NY, 2007.
6. Jones S, Chen WD, Parmigiani G, Diehl F, Beerenwinkel N, Antal T, Traulsen A, Nowak MA, Siegel C, Velculescu VE, Kinzler KW, Vogelstein B, Willis J and Markowitz SD: Comparative lesion sequencing provides insights into tumor evolution. Proc Natl Acad Sci USA 105: 4283-4288, 2008.
7. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW and Velculescu VE: The consensus coding sequences of human breast and colorectal cancers. Science 314: 268-274, 2006.
8. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE and Vogelstein B: The genomic landscapes of human breast and colorectal cancers. Science 318: 1108, 2007.
9. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, McLendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz LA Jr, Hartigan J, Smith DR, Strausberg RL, Marie SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE and Kinzler KW: An integrated genomic analysis of human glioblastoma multiforme. Science 321: 1807-1812, 2008.
10. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, Hong SM, Fu B, Lin MT, Calhoun ES, Kamiyama M, Walter K, Nikolskaya T, Nikolsky Y, Hartigan J, Smith DR, Hidalgo M, Leach SD, Klein AP, Jaffee EM, Goggins M, Maitra A, Iacobuzio-Donahue C, Eshleman JR, Kern SE, Hruban RH, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE and Kinzler KW: Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 321: 1801-1806, 2008.
11. Sugarbaker DJ, Richards WG, Gordon GJ, Dong L, De Rienzo A, Maulik G, Glickman JN, Chirieac LR, Hartman ML, Taillon BE, Du L, Bouffard P, Kingsmore SF, Miller NA, Farmer AD, Jensen RV, Gullans SR and Bueno R: Transcriptome sequencing of malignant pleural mesothelioma tumors. Proc Natl Acad Sci USA 105: 3521-3526, 2008.
12. Cancer Genome Atlas Research Network: Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455: 1061-1068, 2008.
13. Shah S, Pishvaian MJ, Easwaran V, Brown PH and Byers SW: The role of cadherin, beta-catenin, and AP-1 in retinoid-regulated carcinoma cell differentiation and proliferation. J Biol Chem 277: 25313-25322, 2002.
14. American Cancer Society: Cancer Facts and Figures. American Cancer Society, 2010.
15. Mudge J, Miller NA, Khrebtukova I, Lindquist IE, May GD, Huntley JJ, Luo S, Zhang L, van Velkinburgh JC, Farmer AD, Lewis S, Beavis WD, Schilkey FD, Virk SM, Black CF, Myers MK, Mader LC, Langley RJ, Utsey JP, Kim RW, Roberts RC, Khalsa SK, Garcia M, Ambriz-Griffith V, Harlan R, Czika W, Martin S, Wolfinger RD, Perrone-Bizzozero NI, Schroth GP and Kingsmore SF: Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum. PLoS One 3: e3625, 2008.
16. Fitch WM: On the problem of discovering the most parsimonious tree. Amer Natur 111: 223, 1977.
17. Miller NA, Kingsmore SF, Farmer AD, Langley RJ and Mudge J: Management of high-throughput DNA sequencing projects: Alpheus. J Comp Sci Syst Biol 1: 132, 2008.
18. Edwards AWF and Cavalli-Sforza LL: The reconstruction of evolution. Ann Hum Genet 27: 104, 1963.
19. EcK RV and Dayhoff MO: Atlas of Protein Sequence and Structure. Biomedical Research Foundation, Silver Spring, MD, 1966.
20. Merlo LM, Pepper JW, Reid BJ and Maley CC: Cancer as an evolutionary and ecological process. Nat Rev Cancer 6: 924-935, 2006.
21. Shibata D: Molecular tumor clocks and dynamic phenotype. Am J Pathol 151: 643-646, 1997.
22. Shibata D and Tavare S: Counting divisions in a human somatic cell tree: how, what and why? Cell Cycle 5: 610-614, 2006.