Defining Genetic Intra-tumor Heterogeneity: A Chronological Annotation of Mutational Pathways

Wentao Luo[1], Fan Wu[2], Susan R. Atlas[3], Gavin Pickett[4], Kimberly K. Leslie[5], Donghai Dai[5]

Key Words: Cancer, chronology, genetic intra-tumor heterogeneity, single nucleotide polymorphism, cancer genome

Abstract

Tumor heterogeneity is believed to be important in tumor progression and its response to therapies. However, despite numerous mutations being reported in human tumors, genetic intra-tumor heterogeneity remains poorly defined. We have developed a novel strategy to provide a chronological annotation of mutational events in a tumor. We used an endometrial tumor from a patient and transplanted it into athymic mice to create many tumor xenografts. While the patient tumor xenografts were initially responsive to raloxifene treatment, xenografts created with cancer cell clones isolated from the same patient tumor showed dramatic differences in response to raloxifene, indicating existence of intra-tumor heterogeneity with some subpopulations inherently resistant to the drug. A 250K single nucleotide polymorphism (SNP) array from Affymetrix was used to profile genotype changes on 3 xenografts and 10 single cells from another 10 xenografts. We found 797 SNP sites containing loss of heterozygosity (LOH) common to all these specimens, indicating that genetic mutations in these regions may contain the earliest genetic events in the original patient tumor. Based upon the genotype information from the 10 single cancer cells, we developed a phylogenetic tree using neighbor-joining method. We showed that there are at least 3 distinct subpopulations in the patient tumor. Additionally, the phylogenetic tree was used to determine the order of genetic events, thus providing a chronological annotation to genetic mutations. Our approach represents an important analytic strategy for defining genetic intra-tumor heterogeneity and providing chronological annotations to the genetic landscape revealed by future whole genome sequencing in tumors.

1. Dept. of Internal Medicine, University of New Mexico, Albuquerque, NM 87131
2. Cardiovascular Research Center, University of Virginia, Charlottesville, VA 22908
3. Cancer Research and Treatment Center University of New Mexico, Albuquerque, NM 87131
4. Department of Physics & Astronomy and Center for Advanced Studies University of New Mexico, Albuquerque, NM 87131
5. Dept of Obstetrics & Gynecology, University of Iowa Health Care, Iowa City, IA 52242.

Corresponding Author: *Donghai Dai, Department of Obstetrics & Gynecology, University of Iowa Health Care, IA 52241. Email:* Donghai-Dai@uiowa.edu, *Phone: 319-384-1127*

Introduction

Cancer is considered to be a genetic disease[1] wherein an accumulation of genetic mutations allows cancer cells to acquire advantages for abnormal growth. Vogelstein et al. suggested a model of sequential genetic events involving several genes in colorectal cancer,[2] implicating a common mutational pathway among individual tumors. Loeb et al. however suggested another model with mutations in DNA stability, leading to genetic instability as an early and critical event.[3] The concept of genetic instability predicts enormous mutational heterogeneity in a tumor. A systemic and complete documentation of genetic mutations in tumors, as Cancer Genome Anatomy Project proposes, is considered an essential approach to revealing the genetic basis of malignancy. A pilot study recently demonstrated that there are on average 80-90 mutations in protein coding sequences in tumors.[4,5] However, it is believed that only a small portion of them are responsible for cancer development.[6,7] Interestingly, there seems to be no common pattern of genetic mutations in breast or colorectal tumors;[6,7] each tumor presents its own distinct mutational profile.[6] While whole genome sequencing in more tumors is necessary to establish a conclusive finding, these early results suggest that there may be several distinct pathways leading to oncogenesis among individual tumors of the same type.

The existence of several distinct genetic pathways, herein genetic heterogeneity, may not only occur between tumors but may also be present within an individual tumor. The concept of genetic intra-tumor heterogeneity (GITH) has remained an interesting and unsettled research subject for decades.[8,9] The potential for intra-tumor heterogeneity to metastasize was reported three decades ago.[8] Distinct cancer cell subpopulations presented with unique karyotypes.[9] It was proposed then that a neoplasm should not be considered as a uniform entity and that development of effective therapy should discriminate between distinct intra-tumor subpopulations.[10] In 1984, Heppner proposed defining tumor heterogeneity as tumor cell differences existing distinctly as subpopulations with the capability to expand.[11] Efforts have been made recently to define GITH through description of the differential distribution of genetic alterations inside a tumor.[12,13,14] GITH has been demonstrated in different tumor regions [12,13] and in cell populations separated by cell markers[14] using several genetic analysis tools including microsatellite analysis, single nucleotide polymorphism (SNP) arrays and serial analysis of gene expression. Losi et al. demonstrated that K-ras and p53 point mutations as well as loss of heterozygosity (LOH) on chromosomes 5q and 18q showed dramatic heterogeneity in their distribution within different regions of a tumor[13]. Interestingly, they reported that genetic heterogeneity inside a tumor decreases when tumors advance.[13] A recent study by Khalique et al. using 53 SNP markers within 15 genes described a clonal relationship between different regions within tumors.[12] Using LOH and microsatellite instability data, the authors compiled a parsimonious tree to study the evolutionary pattern of different regions

within each tumor.[12] They came to a conclusion that ovarian cancers are monoclonal in origin. Based upon distinct genetic features in cell subpopulations separated by surface markers CD24+ and CD44+, Shipitsin et al. made a molecular definition of breast tumor heterogeneity .[14] These studies and others have demonstrated the heterogeneity of genetic alterations within a tumor through description of their distinct distribution in cancer cell subpopulations. Unfortunately, these studies have all missed one critical element of genetic alterations: a temporal description of genetic events. Fundamental understanding of cancer biology – including the sequential accumulation of genetic alterations necessary for malignant transformation, tumor development through evolution and subsequent distant metastasis – requires the annotation of genetic events with a time component, the chronological order of genetic alterations. In addition, for the concept of GITH to have clinical relevance for prediction of cancer outcome or response to a therapy, a quantitative description of GITH is required. For instance, it should be possible to address questions such as what percentage of cancer cells in a tumor contain a specific mutation, or how many distinct subpopulations exist in a tumor according to the presence or absence of a set of genetic mutations. We present here a study describing a strategy to dissect a tumor and reconstruct its evolutionary history. Genetic alterations in the whole genome of single cancer cell clones were interrogated. GITH was defined through description of these alterations in distinct subpopulations. Subsequently, mutational pathways of cancer cell clones were simultaneously generated through

the construction of a phylogenetic tree. The genetic alterations common to all tumor specimens are hypothesized to be potential genetic events occurring in very early stages of carcinogenesis.

Materials and Methods

1. Collection of patient tissues and creation of xenografts in athymic mice

Collection and use of human tumor specimens was approved by the Human Research Review Committees (HRRC) at the University of New Mexico Health Sciences Center. The endometrial tumor was collected in cold sterile saline in the gross room of the Department of Pathology immediately after arrival from the operating room. Animal use protocol was approved by the University of New Mexico Health Sciences Center Institutional Animal Care and Use Committee. Six- to 8-week old female athymic Crl:NU/NU-nuBR mice were purchased from National Cancer Institute (Frederick, MD). Fresh patient tumor specimens were rinsed with cold sterile phosphate buffered saline (PBS) and grossly necrotic tissue was trimmed away. The tumor was minced and rinsed with PBS again. Approximately 100mg of tumor suspended in 100µl Dulbecco's Modified Eagle's Medium (DMEM) was injected into athymic mice subcutaneously (sc) or intraperitoneally (ip).[15,16] Xenografted tumors were harvested for tissue culture and extraction of genomic DNA.

2. Animal treatment

The xenografted tumors were measured weekly, and tumor cross-sectional areas (mm2) were calculated using the formula: length (mm) × width (mm) × /4.

Raloxifene (Eli Lilly, Indianapolis, IN) was given at 1.5mg/0.2ml/mouse through oral gavage daily (Monday through Friday) starting one week after tumor transplantation. The control mice received saline accordingly. Student's t-test was used to compare the size of tumors between the control and treatment groups.

3. Tissue culture and clone selection

Xenografts isolated from athymic mice were cultured and single cells were picked using a pipette under microscope. Genomic DNAs were extracted from one sc and 2 ip xenografts using the DNeasy Tissue Kit from QIAGEN (Valencia, CA). Ten DNA specimens from 10 single cells (each of them from one ip tumor) were also processed using a single-cell Omniplex whole genome amplification kit from Rubicon Genomics Inc. (Ann Arbor, MI).

4. Creation of sequential generations of clonal cancer cells

The endometrial cancer cell line Hec50co has been used in *in vitro* and *in vivo* studies, and was described previously.[13,14,15,16,17] They were maintained in cell culture media, Dulbecco's Modified Eagle's Medium (DMEM), with 10% FBS. Parental cells (designated as G0) were transferred to 96-well plates at 1 cell/100µl/well for clonal separation. The well with a single cell was spotted and allowed to reach subconfluence. The cell population (from the single cell) was expanded to 30 million. Some of the cells were harvested to yield genomic DNA (designated G1) and the rest were transferred to 96-well plates at 1 cell/100µl/well. Next generations (G2 and G3) were similarly created for harvest of genomic DNA.

5. Whole Genome SNP Genotyping

DNA samples were genotyped using the Affymetrix 250K Nsp Human Mapping GeneChip according to manufacturer's protocol (Affymetrix, Santa Clara, CA). Briefly, 250 ng of genomic DNA was digested with Nsp and then ligated to adapters that recognize the cohesive four base-pair (bp) overhangs. A generic primer that recognizes the adapter sequence was used to amplify adapter-ligated DNA fragments with PCR conditions optimized to preferentially amplify fragments in the 250–2,000 bp size range in a MJ Research (Waltham, Massachusetts) PTC-225 Thermocycler. After purification with a Clontech DNA Amplication (Mountain View, CA) Clean-up Kit, a total of 40 ug of PCR product was fragmented. The fragmented DNA was then labeled with biotin and hybridized to the GeneChip Mapping 250K Nsp GeneChip for 16 hr. The arrays were washed and stained using the Affymetrix fluidics Station 450 and scanned the arrays using a GeneChip Scanner 3000 G7 (Affymetrix, Santa Clara, CA). The Affymetrix GeneChip Operating Software (GCOS) collected and extracted feature data from the GeneChip Scanners. Affymetrix GeneChip Genotyping analysis software (GTYPE) version 4.1 and ArrayAssist software (Stratagene/Agilent Technologies, Santa Clara , CA) were both used to analyze feature intensity data, determine genotype calls, and LOH for each of the 262,264 genomic SNPs.

6. Generation of phylogenetic tree

We utilized the neighbor-joining (NJ) clustering method[17] as implemented in

the Geneious bioinformatics software suite[17] to reconstruct the phylogenetic tree. The four possible SNP calls (AB, AA, BB, and no call) generated by the Affymetrix analysis software were mapped onto the two nucleotides A, C, C, and C, respectively. Each clone was thus transformed into a 'DNA' character-state string or feature vector of equal length, which could then be processed by the phylogenetic methods supported by the Geneious software. A Jukes-Cantor base substitution model[18] was assumed. This represents the simplest consistent mapping of SNPs to bases: each SNP call occurs with equal frequency, and all SNP substitutions occur at the same rate. NJ is a computationally-efficient (polynomial time) phylogenetic reconstruction methodology that has been shown to perform well in numerous tests.[19] It is based on the minimum evolution criterion, and favors tree reconstructions for which the sum of the branch lengths (computed using a pairwise distance matrix) is minimized at each stage of the algorithm.[17] The final reconstructed tree can be interpreted as a phylogeny (*cf.* the time scale in Fig. 3) because the branch lengths in the model correspond to evolutionary distances.

Results

1. Development of tumor resistance to a therapy could be inherent in some cancer cells

We succeeded in establishing xenografts in athymic mice using fresh tumors from endometrial and ovarian cancer patients. One endometrial carcinoma underwent serial transplantation for more than 10 passages. Pathologic studies were conducted to compare tumor features among the original tumor and xenografts in mice and the morphology of the original patient tumors was generally preserved in xenografts.

The endometrial tumor described above was transplanted into 16 athymic mice and randomly divided into control and treatment groups, which were treated with raloxifene (Evista) at 1.5mg/0.2ml/mouse through oral gavage daily (Monday through Friday) for 6 weeks starting one week after tumor transplantation. It was expected that the tumor would be responsive to raloxifene since the tumor abundantly expressed estrogen receptors (data not shown). Indeed the tumor sizes in the treatment group were significantly smaller than those of the control group at the first 3 week during the treatment (t-test, p< 0.05, Fig. 1), indicating that this tumor was sensitive to raloxifene. However, in the 4th week after initiation of raloxifene, the tumor sizes in the two groups were not significantly different and showed the same growth rate, indicating that the tumor had developed resistance to raloxifene. To investigate whether tumor heterogeneity could be a potential mechanism of drug resistance, we sought evidence for the existence of an inherently resistant subpopulation of cancer cells in the responsive tumor.

One untreated xenograft tumor was cultured and cancer cell clones were separated into culture dishes. Nine clones were selected and each was grown to a population of 15 million cells. Each clone was injected separately into 3 athymic mice and treated with raloxifene. We found that two clones were inherently resistant to raloxifene while other clones were sensitive to raloxifene (Fig. 2) when compared to the control. This observation demonstrated the existence of cancer cell heterogeneity within a

tumor in terms of the responsiveness to hormone therapy. It suggested that intra-tumor heterogeneity needs to be addressed for formulation of effective targeted therapies. Since genetic mutations are at the root of tumor development, we next sought to determine whether intra-tumor heterogeneity can be defined according to genetic mutations in individual cancer cells.

Figure 1.
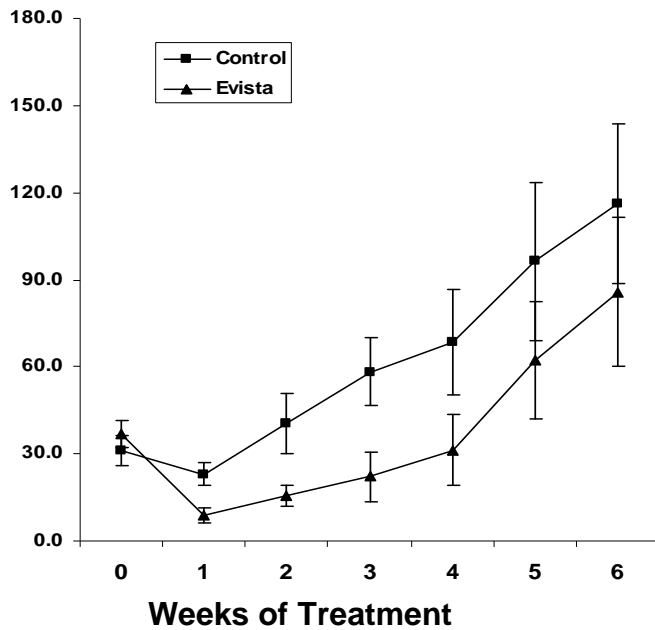
Figure 1. **Endometrial tumor response to raloxifene treatment.** Y-axis: mean tumor cross-sectional areas (mm2). X-axis: weeks after treatment. Raloxifene was given orally every day starting one week after tumor transplantation. Means (n=6-8) and standard errors are presented.

Figure 2.
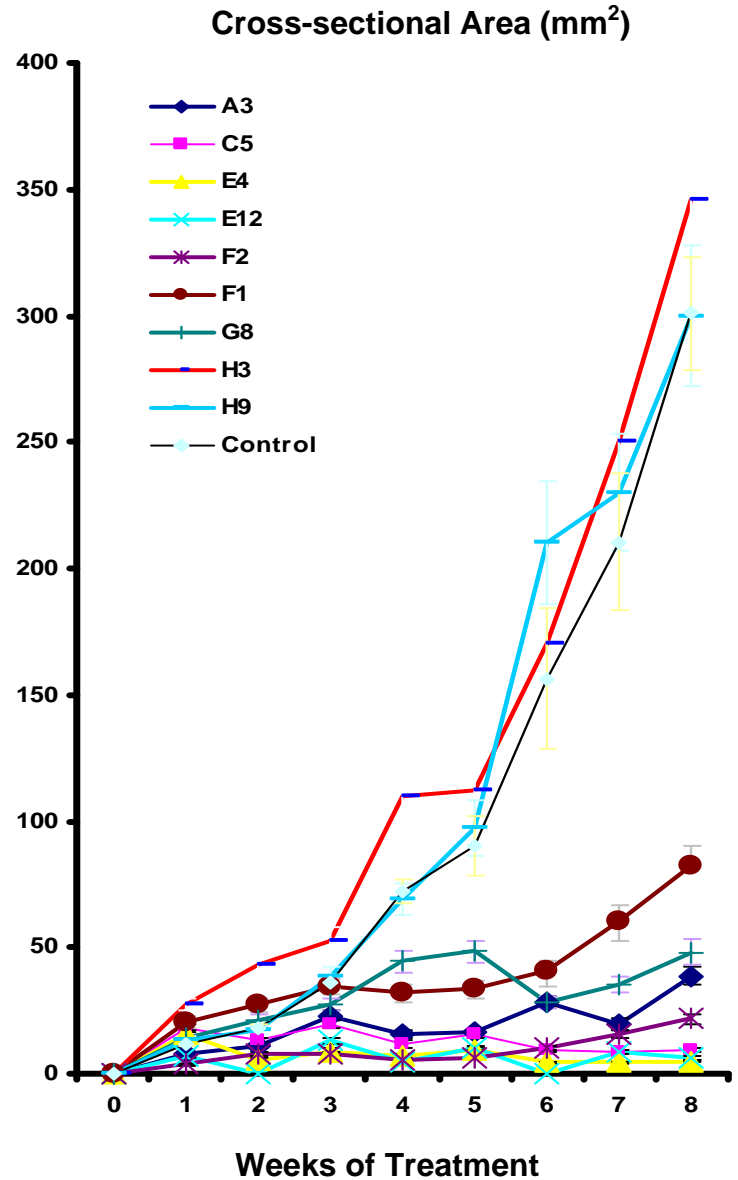
Figure 2. **Clonal tumor response to raloxifene**. Xenografts were generated by clones (A3, C5, E4, E12, F2, F1, G8, H3 and H9) from the patient tumor and treated with raloxifene as above. Y-axis: tumor cross-sectional areas (mm2). X-axis: weeks after treatment. Means (n=3) and standard errors are presented.

2. SNP array for investigation of genetic alterations in the whole genome in cancer cells

A subcutaneous (sc) or intra-peritoneal (ip) xenograft created by injection of a patient tumor homogenate is unavoidably a mixture of many cell populations. Since intra-tumor heterogeneity is our major concern, we used a cancer cell line, mixed or clonally selected, to test how a SNP array reports genotype information in a DNA from mixed cells.

We ran a test to investigate the mutation rate in a clonally expanded cancer cell population using GeneChip Human Mapping 250K Nsp Array set from Affymetrix. This gene array was able to detect changes in genotype in 262,264 SNP sites in the cancer cell genome. The endometrial cancer cell line Hec50co was grown in cell culture dishes and harvested for extraction of DNA when total cell number reached approximately 30 million. A single cell was picked among the 30 million to start a new clonal expansion. The starting parental Hec50co cell was designated as G0 and the sequential clonal passages (grown from a single cell from the previous generation) were named G1, G2 and G3, respectively (table 1). Each generation needs approximately 25 divisions to reach 30 million from a single cell ($2^{25} \approx 3 \times 10^7$), assuming no cell death, in 4 weeks (28 days). The SNP array data showed that the number of LOH increased by 1626 from G1 to G3 (two generations). Thus the rate of LOH is approximately 33 SNP sites/division or 29 SNP sites/day. This indicates that there is gradual accumulation of genetic mutations over time when cancer cells are grown *in vitro*. However, we noticed that there was a decrease of LOH from the parental (G0) to G1 (15.26% to 15.19%). Since the parental cells (G0) could be a more mixed population of cancer cells than the clonally selected descendants (G1, G2 and G3), this decrease of LOH from G0 to G1 may be a false reading by SNP array data analysis. Indeed, the SNP array on the DNA specimen created by equally mixed DNAs from G1, G2 and G3 showed that the LOH is lower than any one of the individual DNAs (14.81% vs 15.19%, 15.72% and 15.81%). Thus, a heterogeneous mixture of DNA could present an increased heterogeneous reading (AB calls) or a reduced LOH reading. After checking genotype readings in the individual SNP sites across the 3 generations and the mixture, we found that the reading in mixture DNA agrees with 90% of majority reading in G1, G2 and G3 (agrees with readings from 2 of 3 generations) and 10% goes with the minority reading (same as one of the 3 generations but not the other 2). This study suggests that we should be very careful in the interpretation of SNP array results from DNAs extracted from a tumor homogenate, which could be very heterogeneous in genetic mutations.

3. Genetic heterogeneity was demonstrated by SNP analysis in a tumor

In order to obtain clones representing subpopulations as distinct as possible, we transplanted part of the patient tumor as described in the raloxifene study into athymic mice to create many sc and ip xenografts. DNA specimens from normal tissue from the same patient, one sc, and two ip tumors, were subjected to analysis by SNP array. Ten DNA specimens from 10 single cells (each of them from one ip tumor) were also used as a pure clonal

DNA sources. They went through whole genome amplification using a commercially available single-cell Omniplex whole genome amplification kit. The LOH data from these 13 specimens are shown at Table 2. The LOH readings are markedly higher in single cells than those in tumor specimens. The difference of LOH between any two specimens varied from 0.16% to 12.76%, suggesting a very heterogeneous nature in genetic mutations among cancer cells in the tumor. This study suggests that survey of genetic alterations in the whole genome by SNP arrays could be a useful tool to document genetic profiles in individual cancer cells.

4. The common genetic mutations among all cancer cells may be the earliest mutational events in cancer evolution: a strategy to profile genetic mutations in cancer stem cells

It is generally believed that differentiated normal cells (or adult stem cells) accumulate genetic mutations and thus transform into malignant cells (or the cancer stem cells), which expand to a population of billions of cancer cells in a tumor. We reasoned that the genetic mutations in the cancer stem cell will be carried over to all their offspring, and, consequently, that the common genetic mutations shared by all cancer cells must have come from the cancer stem cell. Therefore, if we can isolate sufficient numbers of cancer cells, the common genetic mutations among them will give us a view of potential genetic mutations originating from the cancer stem cells. Among 59,582 informative SNP sites where the normal tissue had a heterozygous genotype, a matrix detailing

Table 1.

Cell Data	Number of SNP Sites			Percentage of Total SNP Sites			
	AA Call	AB Call	BB Call	AA Call	AB Call	BB Call	LOH
Ref Genomic DNA	98218	69159	94887	37.45%	26.37%	36.18%	
Hec50co cell (G0)	118491	29138	114636	45.18%	11.11%	43.71%	15.26%
Generation 1 (G1)	118438	29321	114505	45.16%	11.18%	43.66%	15.19%
Generation 2 (G2)	119435	27931	114872	45.54%	10.65%	43.80%	15.72%
Generation 3 (G3)	119330	27695	115239	45.50%	10.56%	43.94%	15.81%
Equal Mixed DNA	118150	30318	113797	45.05%	11.56%	43.39%	14.81%

SNP arrays on an endometrial cancer cell (G0), its clonally selected subsequent generations (G1, G2 and G3) and a DNA mixture equally from G1, G2 and G3. AB calls at Ref Genomic DNA was used as standard and LOH of other specimens was derived by subtraction.

genotypes of individual tumors and clones was generated for comparing the genotype in every SNP site among cancer specimens. Data analysis showed that there were 797 out of 49,582 SNP sites displaying LOH in all cancer specimens (3 tumors and 10 single cells), indicating that progenitor cells of the original patient tumor may possess genetic mutations in these regions. We found that 797 LOHs have directly affected 4 exons from 4 genes, 2 UTRs of 2 genes and 222 introns of 189 genes, and the rest are in intergenic regions. The 195 affected genes have been shown to play important roles in the following cellular functions: cellular adhesion and motility (14 genes), signal transduction (53), transcriptional regulation (21), transport (11), cellular metabolism (21), intracellular trafficking (8), RNA metabolism (7), cellular components (12), others (18) and

unknown (30). While still far from determining what genetic mutations transform stem cells, this approach potentially provides a list of earliest genetic mutations involved in the malignant transformation.

Table 2.

|	AB Calls among Total SNP Sites	LOH
Normal tissue	22.77%	
sc tumor	20.07%	2.70%
lp tumor 1	21.59%	1.18%
lp tumor 2	18.84%	3.93%
Single cell clone 1	16.85%	5.92%
Single cell clone 2	9.52%	13.25%
Single cell clone 3	8.83%	13.94%
Single cell clone 4	8.99%	13.78%
Single cell clone 5	12.25%	10.52%
Single cell clone 6	10.19%	12.58%
Single cell clone 7	9.77%	13.00%
Single cell clone 8	11.08%	11.69%
Single cell clone 9	10.75%	12.02%
Single cell clone 10	12.87%	9.90%

SNP arrays on DNA specimens from the same patient tumor. Percentage of LOH of cancer specimens was derived from subtraction of percentage of heterogeneous sites (AB calls) in tumor specimen by that in the normal tissue.

5. Construction of phylogenetic tree to define genetic intra-tumor heterogeneity

Inspired by genetic genealogies in population studies used to describe genetic relationships between individuals, we sought to construct a phylogenetic tree of cancer cells in a tumor in order to build a cancer evolutionary lineage based upon the similarity of genotypes in cancer cells. Based upon the genotype information from the 10 clones isolated from the tumor, we developed a phylogenetic tree using a neighbor joining method (Fig. 3). Nodes of the tree represent the inferred common ancestors of the genetic alterations. As shown in Fig. 3, the tumor consists of 3 distinct subpopulations with Clone 10 belonging to the first, Clone 1 to the second, and the rest belonging to the third (Fig.3, the left panel). An evolutionary history was extrapolated from identification of shared mutations. For instance, the genetic mutations in Clone 3 can be grouped into six sequential stages (Fig. 3, right panel). Any genetic event at a higher step on this pathway occurs before the genetic event at a lower step. Analysis of a sufficient number of clones will allow us to determine the time course of all significant genetic mutations and build a virtual history of genetic evolution.

Figure 3.
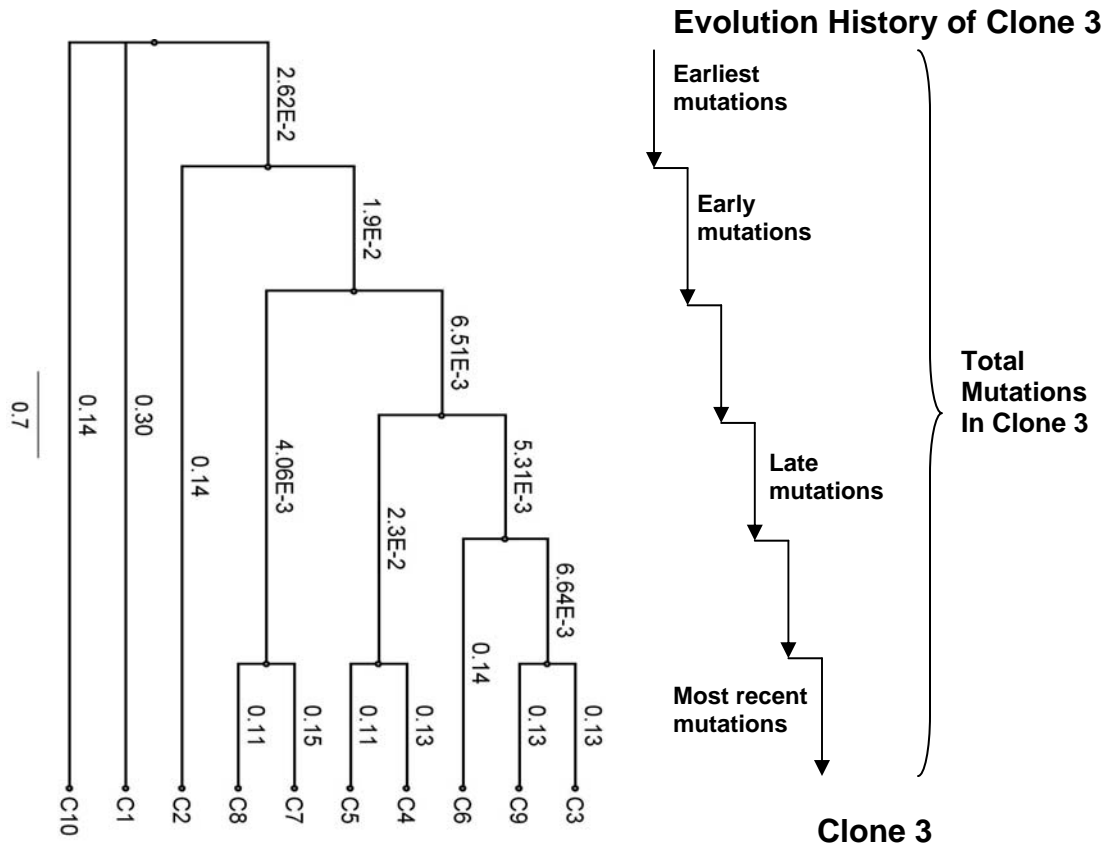
Fig. 3. Quantitative description of genetic variations in subpopulations with chronological annotations. The left panel shows a phylogenetic tree of the tumor. The right panel illustrates the chronological order of genetic events in Clone 3 with arrows pointing the flow of time from the past to the present. Any genetic event at a higher step would have to occur before the genetic events at a lower step. Addition of more clones to the tree will increase the resolution to determine the physical distribution of genetic events in clones and their evolution history.

Discussion

We have developed a genealogical approach, similar to anthropological/paleontological research, for reconstructing cancer evolutionary history and revealing the chronology of oncogenic processes. Instead of taking a tumor as one entity and using DNA from mixed cells, we isolated single cell clones to deconstruct the tumor. Theoretically, our approach will be able to differentiate genetic alterations occurring at the single cell ancestry from the new genetic mutations acquired every division as long as the cell lineage carrying a specific mutation is preserved in the subsequent evolution and cancer expansion. Our approach is logical for examining whether there is a common mutational pathway among tumors because it starts with the elucidation of the genetic pathway in an individual tumor, and is thus independent of whether there is a common genetic pathway among tumors grouped according to histology or genetic features. In another words, it may allow discovery of a common mutational pathway existing

in different types of tumors.

We reported here that the earliest genetic mutations may involve some of the 797 SNP sites. This approach will allow us to identify the earliest genetic mutations and possibly those responsible for the malignant transformation of stem cells without a need to isolate cancer stem cells, which commonly has to rely on certain cell surface markers.

However, there remain several technical hurdles to be addressed before fully realizing the potential of this strategy. The first will be the requirement of complete documentation of genetic alterations in the whole genome. Using SNP arrays has some drawbacks, even with the arrival of higher resolution chips such as 1 million SNP sites per array. The SNP array we used interrogates a base sequence every 10-20K and it will surely miss many alterations between two SNP sites such as point mutations, which are reported to be more common than other genetic alterations in cancer[6,20]. Interestingly, a recent report shows that LOH in stem cells is much more common than point mutations after exposure to a carcinogen, and LOH could be an important mechanism of carcinogen-induced oncogenesis in stem cells.[21] Additionally, LOH is a common mechanism for loss of function in tumor suppressor genes. Thus, SNP arrays to detect LOH may be particularly useful in the detection of genetic alterations in cancer stem cells.

The second technical hurdle will be the isolation of a sufficient number of pure cancer cell clones to represent the heterogeneity of the original tumor. A small piece of tumor, even isolated under microscope, may include many genetically distinct cells, which will distort the linkage analysis and generation of a phylogenetic tree. Isolation of a single cell is preferred because it has only one set of the genome without any potential heterogeneity. However, the quality of genomic DNA from a single cell may pose a problem for genomic analysis. In our study, the call rate in SNP arrays is significantly lower in the DNA from a single cell genome than that from a tumor. This may result from loss of SNP sites during amplification of a single copy genome. The potential solution will be to select a group of homogeneous cells. Here, we chose to select only one cell from every xenograft tumor in order to make these single cells as distinct from each other as possible. There is also a possibility that the genetic alterations observed in cells are acquired during tumor growth in athymic mice. There will be new genetic alterations when cancer cells proliferate. It is certain that cancer cells acquire a significant amount of genetic alterations during their xenograft growth. If the theory of genetic instability for oncogenesis bears out, there will be an exponential increase in genetic mutations and those responsible for early tumor development will be overwhelmed by new mutations. However, these more recently acquired genetic mutations should be located at the bottom part of our phylogenetic tree and the number of newly acquired genetic alterations during xenograft growth should not exceed the difference between clones 9 and 3 (or the least of differences between any two clones) since any xenograft used in the building of our phylogenetic tree was generated by injection of patient tumor homogenates.

In order to construct a phylogenetic tree using the genomic SNP data from the ten

clones, we mapped SNP character information onto corresponding DNA strings. This facilitated the use of existing phylogenetic software tools previously developed for DNA sequence data, as described below. To the best of our knowledge, this represents the first attempt to develop a clonal phylogenetic reconstruction based on SNP data reported in the literature. We have used the neighbor-joining (NJ) clustering method,[17] which is widely regarded as one of the best approaches available for computing phylogenies.[19] It is fast and has demonstrated good accuracy in many test problems. The initial reconstruction reported here can be further refined in several important ways. Continuing work is aimed at establishing the statistical robustness of the candidate reconstruction using bootstrap methods,[22] and comparing it with the results of alternative methods of phylogenetic reconstruction, such as Bayesian Markov Chain Monte Carlo (MCMC) and coalescent theory. These studies will enable us to investigate consistency and stability across alternative reconstructed clonal phylogenies and develop a final consensus reconstruction for the given set of clones, including reliability estimates for various branches, as well as providing guidance for future experiments with respect to tolerable levels of noise, and quantifying the impact of additional clonal samples on the final consensus genealogy.

Since our aim in the present study is to illustrate a general experimental/modeling strategy for the reconstruction of cancer clone genealogies, Fig. 3 should not be construed as a definitive phylogenetic reconstruction for the ten clones studied here. Further computational analysis will be needed in order to fully explore the effects on the final reconstructed cancer cell phylogeny of algorithmic choices such as the choice of distance measure or the treatment of missing data, as well as underlying model assumptions relating to the nature of character transformations, and the effects of experimental noise in the SNP data. For the NJ method, parallelized bootstrap resampling[23] may be used to compute statistical bounds and provide reliability assessments for all or parts of the reconstructed tree.[24] Other phylogenetic reconstruction methods, while algorithmically complex and computationally demanding, represent valuable complementary approaches to clustering methods such as NJ. Examples include maximum-likelihood inference using Markov Chain Monte Carlo sampling[25] and coalescent theory.[26] These methods incorporate a principled mechanism for aggregating the results of randomly sampled candidate tree structures, along with an underlying parametric evolutionary model that provides an intrinsic molecular clock. Investigation of these algorithms and their impact on the final phylogenetic reconstruction for the set of ten cancer clones is presently underway.

Our study method has introduced a new concept in cancer biology: a temporal description of genetic events in cancer development. With the development of affordable whole genome sequencing technology, we should be able to more thoroughly understand the genetic basis of cancer. Definition of GITH through quantitative description of genetic variations among subpopulations and chronological annotations of these events will significantly improve our understanding of tumor progression and assist in predicting the outcome of cancer therapies.

References

1. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;100:57-70.

2. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990;61:759-67.

3. Loeb LA, Springgate CF, Battula N. Errors in DNA replication as a basis of malignant changes. Cancer Res 1974;34:2311-21.

4. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. Science 2006;314:268-74.

5. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, et al. The Genomic Landscapes of Human Breast and Colorectal Cancers. Science 2007;318:1108.

6. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. Science 2006;314:268-74.

7. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, et al. The Genomic Landscapes of Human Breast and Colorectal Cancers. Science 2007;318:1108.

8. Fidler IJ, Kripke ML. Metastasis results from preexisting variant cells within a malignant tumor. Science 1977;197:893-5.

9. Dexter DL, Kowalski HM, Blazar BA, Fligiel Z, Vogel R, Heppner GH. Heterogeneity of tumor cells from a single mouse mammary tumor. Cancer Res 1978;38:3174-81.

10. Fidler IJ. Tumor heterogeneity and the biology of cancer invasion and metastasis. Cancer Res 1978;38:2651-60.

11. Heppner GH. Tumor heterogeneity. Cancer Res 1984;44:2259-65.

12. Khalique L, Ayhan A, Weale ME, Jacobs IJ, Ramus SJ, Gayther SA. Genetic intra-tumour heterogeneity in epithelial ovarian cancer and its implications for molecular diagnosis of tumours. J Pathol 2007;211:286-95.

13. Losi L, Baisse B, Bouzourene H, Benhattar J. Evolution of intratumoral genetic heterogeneity during colorectal cancer progression. Carcinogenesis 2005;26:916-22.

14. Shipitsin M, Campbell LL, Argani P, Weremowicz S, Bloushtain-Qimron N, Yao J, et al. Molecular definition of breast tumor heterogeneity. Cancer Cell 2007;11:259-73.

15. Dai D, Albitar L, Nguyen T, Laidler LL, Singh M, Leslie KK. A therapeutic model for advanced endometrial cancer: Systemic progestin in combination with local adenoviral-mediated progesterone receptor expression. Mol Cancer Ther 2005;4:169-75.

16. Dai D, Holmes AM, Nguyen T, Davies S, Theele DP, Verschraegen C, et al. A potential synergistic anti-cancer effect of paclitaxel and amifostine in endometrial cancer. Cancer Res 2005;65:9517.

17. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987;4:406-25.

18. Jukes TH, Cantor CR. Evolution of protein molecules. In: Munro HN, editor. Mammalian protein metabolism III. Academic, New York; 1969. p. 21-132.

19. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A 2004;101:11030-5.

20. Bielas JH, Loeb KR, Rubin BP, True LD, Loeb LA. Human cancers express a mutator phenotype. Proc Natl Acad Sci U S A 2006;103:18238-42.

21. Donahue SL, Lin Q, Cao S, Ruley HE. Carcinogens induce genome-wide loss of heterozygosity in normal stem cells without persistent chromosomal instability. Proc Natl Acad Sci U S A 2006;103:11642-6.

22. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning - data mining, inference, and prediction. Springer, New York; 2001.

23. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987;4:406-25.

24. Geneious Pro 3.0.6. Geneious Pro 3.0.6. Biomatters, Ltd, Auckland, New Zealand, www.geneious.com; 2005.

25. Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 2004;20:407-15.

26. Hein J, Schierup MH, Wiuf C. Gene genealogies, variation, and evolution - A primer in coalescent theory. Oxford, New York; 2005.