Cancer Research

Single-Molecule Genomic Data Delineate Patient-Specific Tumor Profiles and Cancer Stem Cell Organization

Andrea Sottoriva[1,2,3], Inmaculada Spiteri[2], Darryl Shibata[4], Christina Curtis[3], and Simon Tavaré[1,2,5]

Abstract

Substantial evidence supports the concept that cancers are organized in a cellular hierarchy with cancer stem cells (CSC) at the apex. To date, the primary evidence for CSCs derives from transplantation assays, which have known limitations. In particular, they are unable to report on the fate of cells within the original human tumor. Because of the difficulty in measuring tumor characteristics in patients, cellular organization and other aspects of cancer dynamics have not been quantified directly, although they likely play a fundamental role in tumor progression and therapy response. As such, new approaches to study CSCs in patient-derived tumor specimens are needed. In this study, we exploited ultradeep single-molecule genomic data derived from multiple microdissected colorectal cancer glands per tumor, along with a novel quantitative approach to measure tumor characteristics, define patient-specific tumor profiles, and infer tumor ancestral trees. We show that each cancer is unique in terms of its cellular organization, molecular heterogeneity, time from malignant transformation, and rate of mutation and apoptosis. Importantly, we estimate CSC fractions between 0.5% and 4%, indicative of a hierarchical organization responsible for long-lived CSC lineages, with variable rates of symmetric cell division. We also observed extensive molecular heterogeneity, both between and within individual cancer glands, suggesting a complex hierarchy of mitotic clones. Our framework enables the measurement of clinically relevant patient-specific characteristics *in vivo*, providing insight into the cellular organization and dynamics of tumor growth, with implications for personalized patient care. *Cancer Res; 73(1); 41–49. ©2012 AACR.*

Introduction

The cancer stem cell (CSC) paradigm posits that malignancies retain part of the stem cell organization of the tissue of origin (1, 2) and that only a subset of stem-like cancer cells have self-renewal capacity, influencing both tumor progression and therapeutic resistance (3, 4). For the vast majority of cell divisions, a CSC gives rise, through asymmetric division, to a transit-amplifying cell (TAC) with short-term replicative potential, which becomes fully differentiated after a certain number of divisions. Much less frequently, a CSC undergoes self-renewal through symmetric division, spawning a new CSC. Hence, TACs and differentiated cancer cells (DCC) are thought to comprise the bulk of the tumor mass. This model differs from the classical (also called *clonal* or *stochastic*) model of cancer, where all cells are potentially tumorigenic.

Colorectal cancer (CRC) develops through the accumulation of key mutations in the epithelial tissue of the colon (5) and it is a leading cause of cancer death (6). Evidence suggests that CSCs are present and play an important role in CRC progression and expansion (7–9). However, the most reliable evidence for the existence of CSCs in human malignancies to date comes from transplantation assays, which identify CSCs on the basis of their clonogenicity *in vitro*, using limited dilution experiments, or *in vivo* through mouse xenograft experiments. This approach shows the ability of isolated cell populations to form tumors and recapitulate the heterogeneous populations found in the primary neoplasm. A known limitation of this method is that these so-called tumor initiating cell fractions may be highly enriched for cells that are able to survive experimental manipulations, including the transplantation process and the foreign environment in which they are grown (10). Moreover, while stem cell transplantation assays can provide insight into the properties of these cells under specific experimental conditions, they cannot report on the fate of the transplanted cell in its original tumor (11). Moreover, CSCs, that are thought to be the drivers of subclonal expansions, may vary in frequency and phenotype (12). However, definitive CSC markers remain elusive in solid tissues and there is no direct evidence for CSCs capable of fueling long-term

Authors' Affiliations: [1]Department of Oncology, University of Cambridge; [2]Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge, United Kingdom; Departments of [3]Preventive Medicine, [4]Pathology, Keck School of Medicine; and [5]Biological Sciences, University of Southern California, Los Angeles, California

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Corresponding Authors: Christina Curtis, University of Southern California, 1450 Biggy street, HNRT 1517M, Los Angeles, CA 90033. Phone: 323-442-7887; Fax: 323-442-7887; E-mail: ccurtis@usc.edu and Simon Tavaré, Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, United Kingdom. Phone: 44 (0) 1223 404290; Fax: 44 1223 765900; E-mail: simon.tavare@cancer.org.uk

doi: 10.1158/0008-5472.CAN-12-2273

expansion in nonmanipulated human solid malignancies, as such new approaches are needed (10, 13, 14). Because of the difficulty in measuring tumor dynamics directly in patients, various parameters such as the CSC organization, the rate of mutation and apoptosis, and the tumor age, have not been quantified. These features play a fundamental role in tumor progression and therapy response and are likely variable among patients, with implications for personalized treatment regimes.

To interrogate the dynamics of individual tumors, one can exploit the fact that cells retain records of their past proliferative history in the form of somatic mutations [e.g., microsatellites (15, 16) or methylation (17)] that arise from errors in the replication machinery. Therefore, cells contain molecular clocks that register their phylogenetic history and can be read using genomic approaches. Intrinsic tumor characteristics directly shape such phylogenies, and consequently their corresponding molecular patterns.

We have developed a framework to measure tumor-specific features based on the underlying molecular signature of cancer cells. In previous studies, we showed the use of methylation-based molecular clocks to explore the dynamics of the human colon crypt (17–19) and CRC (20, 21), but with limited resolution. With the adoption of high-throughput sequencing we now obtain a 50-fold increase in throughput. Moreover, with the enhanced precision of the selected molecular clocks and novel computational methods that account for the spatial structure of hundreds of billions of cells, we are now able to reconstruct the ancestry of individual tumors and measure multiple clinically relevant characteristics from tumor biopsies or surgical resections. Here, we present a novel framework, referred to as Spatial Cell Ancestral Inference (SCAI), that for the first time enables the measurement of patient-specific tumor characteristics, including CSC organization, using a combination of ultradeep patient molecular data, spatial computational modeling, and statistical inference (Fig. 1A).

Materials and Methods

Spatial cell ancestral inference

SCAI is divided into 3 building blocks as shown in Fig. 1A: the patient molecular data, the mathematical/computational model, and the statistical inference method. The patient molecular data can be based on somatic point mutations, microsatellites, or neutral methylation patterns derived from clinical specimens. The second block is the mathematical model of the biologic system of interest. As in any experimental setting, a model of the studied system is required; in this case it is a computational model. This must be a faithful representation of the system, simulate tumor growth in a spatial fashion, and consider the underlying mechanisms that are thought to be most relevant, such as the mutation rate, apoptosis, and cellular organization. To fit the model to the data, we use a statistical inference technique called Approximate Bayesian Computation, or ABC (22, 23). This results in a probability distribution of the parameter values, given the data we observed. This distribution represents an indirect measurement of that parameter or characteristic in the original biologic system (the tumor) at the time the data were collected. Thus, our approach enables the measurement of clinically relevant parameters from human tumors without the need for invasive techniques that are limited to animal models.

Molecular data

The IRX2 molecular clock. A key principle of our approach is the use of neutral somatic mutations (tags) as a molecular clock. Neutrality guarantees the linear relationship between cell division and somatic errors, which would not necessarily hold for a functional genomic region under selective pressure. A molecular clock has a certain probability of *ticking* by introducing a neutral mutation at cell division. Here we use DNA methylation as a marker of cell fate because the error rate is 10,000-fold higher than that observed for nucleotide substitutions ($\sim 10^{-5}$ vs $\sim 10^{-9}$ errors per nucleotide per

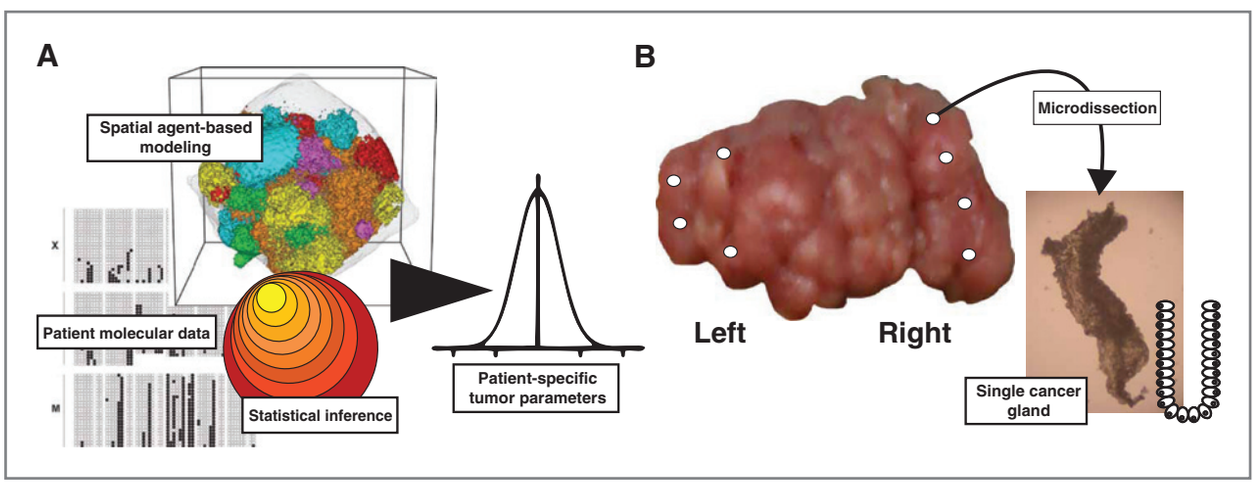

Figure 1. Schematic of the SCAI framework and sampling procedure. A, the SCAI framework makes use of patient molecular data, in this case neutral methylation haplotypes, as markers of cell mitotic history. By integrating these data with agent-based spatial computational modeling and statistical inference techniques such as Approximate Bayesian Computation, patient-specific tumor parameters can be estimated. B, 4/5 CRC glands, approximately 8,000 to 10,000 cells in size, were microdissected from each side of the tumor (left and right) and target sequenced at high coverage.

division) and hence allows for more precise measurements (24). The ticking of the clock corresponds to either a methylation or demethylation event at a CpG dinucleotide within the selected neutral locus, which is essentially unmethylated at the zygotic stage. Hence, molecular clocks based on a neutral methylation locus should display low methylation in young individuals and remain hypomethylated throughout life in nondividing tissues, whereas they should accumulate methylation errors linearly with age in dividing tissues.

In this study, we make use of the *IRX2* molecular clock, a 201 bp locus on chromosome 5. We generated high-throughput methylation data using targeted bisulfite sequencing of DNA from different normal tissue types, resected at the time of autopsy from multiple patients, to verify the neutrality of *IRX2*. Methylation data from normal colon ($n = 6$), heart ($n = 4$), cerebellum ($n = 4$), and neutrophils ($n = 2$) were analyzed. Supplementary Fig. S1 indicates that *IRX2* exhibits low methylation in nondividing tissues (heart and brain) and age-related methylation in mitotic tissues (colon). Similarly, we report low methylation in neutrophils because of the large number of slow cycling bone marrow stem cells from which they derive, and their short lifetime (~5 days). Furthermore, we report that copy number alterations of this locus in CRC are rare, as they were present in only 4 of the 1,192 arrays ($|LRR| > 0.8$) available from The Cancer Genome Atlas (TCGA) CRC study (25). As we model single-molecule information from a population of cells, it is also important to verify that PCR amplification biases, a potential problem of many second generation sequencing approaches, do not alter the composition of the patterns in our samples. The *IRX2* locus spans the rs486667 SNP, which is heterozygous in an estimated 44% of individuals of European descent, and can be used to assay allele-specific PCR biases. The majority of individuals for whom we collected normal brain and heart tissue were heterozygous for this SNP, and were therefore used to confirm the absence of significant PCR amplification biases in our approach (mean allelic frequency 0.57 ± 0.07, 95% CI). As *IRX2* spans an 8 CpG region, this locus allows for 256 unique methylation patterns, and avoids the issue of saturating the clock. These findings indicate the suitability of *IRX2* as a molecular clock.

Cancer sample collection. We collected a total of 40 tumor glands from 5 CRCs (CT, CU, CX, HA, and Z). Clinical information for each of the patients is reported in Table 1. All tumors were untreated at the time of resection. For each tumor, 2 regions approximately 0.5 cm^3 in size were sectioned from opposite sides of the tumor (referred to as *left* and *right* side). Within each region, 3 to 5 CRC glands were microdissected, each composed of 8,000 to 10,000 cells (Fig. 1B). DNA from each gland was extracted and bisulfite converted as previously described (20). The efficiency of conversion was assessed in our dataset and confirmed to be extremely high (>99.98%). Samples were then PCR amplified for the locus of interest with multiplex identifiers and sequenced with a Roche 454 GS Junior system. Amplicon sequence for *IRX2* is provided in Supplementary Table S1. We report an average throughput of more than 1,500 reads per gland.

Table 1. Patient clinical information

Patient	Age	Tumor size (cm)	Stage	MSI+
CT	53	4.5	3	N
CU	50	4.5	1	Y
CX	44	9	3	N
HA	61	7.5	3	N
Z	83	6	3	NA

The error rate for the 454 sequencing technology is in the order of 0.01 errors per nucleotide. However, the large majority of these errors are because of the presence of homopolymers in the sequence, a known problem for pyrosequencing (26). Using targeted sequencing, we avoid this problem by excluding CpGs that are in the proximity of homopolymers. Moreover, we strictly filtered out sequences containing neither a T nor a C at any of their CpG sites. We estimated those errors to occur with rate 0.004 errors/nucleotide in our analysis. Thus, the *undetectable* methylation sequencing error (where a C converts into a T or a T into a C) was approximately 0.002 errors/nucleotide. Considering also the bisulfite conversion failure rate (0.002), this results in a methylation tag sequencing fidelity of 99.6%. Hence, the low error rate achieved with this conservative filtering method, coupled with the high frequency of methylation events (in contrast with the relatively low nucleotide substitution rate), makes our approach extremely robust to sequencing errors. Sequence data analysis was done using a custom R/Perl pipeline, which extracts the methylation status from each 454 read and represents it as a binary string where 1 is methylated and 0 is unmethylated.

Modeling

Cancers are large and complex malignant tissues that acquire the ability to grow out of control, invade the surrounding tissues, and form metastases (27). During cancer progression, tumor cells undergo a large number of mitotic events and often present at diagnosis as a large mass with a diameter of a few centimeters, containing hundreds of billions of cells. Numerous mathematical and computational models have been developed to simulate tumor growth (28–30), which feature a considerable level of detail in simulating the malignant processes and shed light on various aspects of cancer dynamics. However, to do inference using molecular data, we need to simulate the whole tumor with more than 10^{11} cells, millions of times. To tackle this problem, a novel approach capable of simulating very large tumors quickly was required.

In our large-scale tumor model, we exploit the fact that colon cancer is organized into glandular structures. Cancer glands are thought to expand with different mechanisms, one of the most common being gland fission (31, 32), where a gland splits into 2 daughter glands containing about half of the original cell population each. Other mechanisms of growth have also been observed, such as top-down spread of clones (33); however, those are computationally intractable because of their complexity. For these reasons, we assume gland fission as the only mechanism of expansion in our model. We simulate the growth

and the exact 3-dimensional position of the glands at any moment in time, but do not keep track of the position of single cells within a gland. To achieve efficiency, we developed the model in 2 computationally independent parts: the calculation of the gland phylogenetic tree and the simulation of the single-cell molecular patterns within the glands (see Supplementary Materials and Methods for details). The model incorporates cell division, apoptosis, molecular mutations (methylation errors), and CSC organization by assuming that CSCs divide symmetrically with probability ψ, thus generating a new CSC, or asymmetrically with probability $1 - \psi$, and yielding a TAC. The latter can divide only a few times, G, before becoming senescent and ceasing to divide (DCC). To model the classical or clonal model of growth, it is sufficient to set $\psi = 1$, making all cells potentially tumorigenic. The cell-cycle time may be variable across the neoplasm; moreover, recent studies of normal colon crypts in mice suggest heterogeneity in the cell division rate (34, 35). In the absence of human data and for the sake of simplicity, in our model we assume an average cell cycle of 4 days (36) uniformly across the tumor. We do not make further assumptions about the growth of the tumor because its evolutionary dynamics directly follow from a chosen parameter set.

In our simulations we use a $370 \times 370 \times 370$ point lattice that contains the growth of a large carcinoma with up to 16 million glands and a total of 130 billion cells. Considering an average cell diameter of approximately 12 μm, this forms a neoplasm of about 7.5 cm in diameter.

Statistical inference

Because of the complexity of biologic systems and the numerous possible interactions between the underlying mechanisms and their parameters, the most appropriate method to address the problem is Bayesian inference (37). A powerful and elegant technique suitable for use with agent-based models is ABC (19, 22, 23). This family of statistical inference methods evolved from earlier approaches based on rejection algorithms used to construct posterior distributions (38, 39). The following ABC scheme yields a sample from an approximation of the posterior distribution $P(\theta|\rho[S\{D\},S\{D'\}] < \varepsilon)$, given the data D and a tolerance threshold ε:

1. Sample the parameter θ from the prior distribution $P(\theta)$.
2. Simulate the data D' from the computational model with input θ.
3. If $\rho(S[\mathrm{D}], S[\mathrm{D}']) < \varepsilon$ accept θ.
4. Go to 1.

In short, we accept those parameter sets that generate methylation patterns similar to the ones we observe, given a certain error threshold ε, a set of summary statistics S, and a distance measure ρ. We use the following summary statistics for each gland, which are valid for any generic set of binary strings:

- Average percent methylation S_p.
- Number of distinct methylation patterns S_d.
- Number of singleton patterns (patterns that appear only once in the gland) S_s.
- Mean pairwise distance between the patterns S_w.
- Kolmogorov distance (40) of the pattern set S_k.
- Shannon index (41) of the pattern set S_h.

Each summary statistic is normalized to have mean 0 and standard deviation 1. This ABC step produces a table where every line corresponds to a single tumor simulation. We have calculated more than 4,500 different phylogenetic trees of the glands from which we computed 8 million instances of tumor molecular patterns, each corresponding to a 130 billion cell neoplasm. We have taken prior distributions of all the parameters as uniform and validated our framework on synthetic data (see Supplementary Material and Methods and Supplementary Fig. S2). In our results, we summarize the information in each posterior distribution by its modal value.

Results

Cancer glands are heterogeneous, harboring multiple cell lineages

As the methylation profiles of neutral loci reflect the proliferative status of cells, they can be exploited to reconstruct the relationship between subclones within a sample. Intratumor heterogeneity and the spatial distribution of cancer clones are poorly understood, but are believed to follow evolutionary principles (42). We observe that almost all tumor glands (36/40) contained multiple clones and were composed of 2 to 5 distinct mitotic lineages, as shown by the phylogenetic reconstruction of the methylation patterns in Fig. 2A (see Supplementary Material and Methods for details). It is anticipated that physically proximal regions are more closely related. As approximate spatial information (right or left side of a tumor mass) was collected for each sample, the extent of heterogeneity in spatially separated tumor fragments can be examined. Figure 2B illustrates the tendency of the glands to cluster by side, indicating spatial correlation between different parts of the tumor at the molecular level, although the extent of the clustering varies between tumors. We note that occasionally the right and left glands cluster together (e.g., tumor CT) partly because of the stochasticity of the methylation patterns but also because of the heterogeneity present even between nearby glands. Thus, CRC glands are spatially related within a tumor but also exhibit heterogeneity, indicative of a complex hierarchy of mitotic clones spanning multiple spatial scales (see Supplementary Table S2).

Inference of patient-specific tumor profiles

The patient-derived molecular data we have discussed so far represent the input for our analysis framework, which returns the posterior distributions of the tumor parameter values. These parameters correspond to *measurements* of tumor characteristics directly derived from the individual patient samples. Within SCAI, the unknowns of the system are represented as parameters of the analysis. Rather than making assumptions about these parameters, we use statistical inference to infer them from the data. Specifically, we interrogated the following:

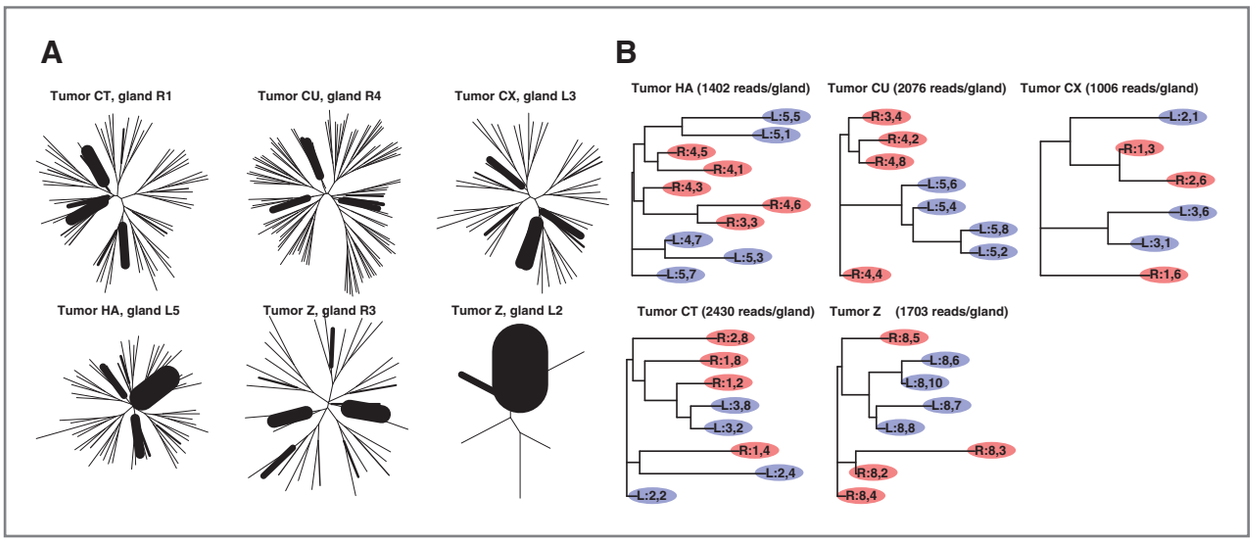

Figure 2. Intratumor heterogeneity between and within glands. A, the phylogeny of the methylation patterns shows that almost all glands (36/40) contained multiple clones, with 2 to 5 coexisting lineages (branch thickness corresponds to number of reads per pattern). An example of a monoclonal gland is shown for tumor Z, gland L2. Different clones correspond to distinct CSCs lineages that coexist within the tumor gland. B, the methylation patterns of tumor glands tend to cluster by tumor side, indicating spatial correlation between different tumor areas. Extensive intergland heterogeneity is also present, and some glands cluster together despite being located on different sides of the tumor (CT). These data reflect a complex hierarchy of mitotic clones across different scales.

- The fraction of CSCs in the tumor, ξ.
- The symmetric division rate of CSCs, ψ.
- The methylation/demethylation rate per cell division, μ.
- The context-dependent methylation factor, χ.
- The relative tumor age from malignant transformation, γ.
- The number of TAC divisions before full differentiation, G.
- The rate of apoptosis of cancer cells, a.

The parameter χ models the feed-forward effect in which the more a CpG-island is methylated, the more it tends to methylate further. Although little is known about context-dependent methylation, it is predicted that this phenomenon occurs in normal colon crypts (18). The model of cell division assumes that CSCs divide symmetrically, producing a new CSC with probability ψ and otherwise producing a new TAC that will divide only G times before becoming fully differentiated. The combination of ψ and G determines the CSC fraction ξ within a tumor, following the relation in Supplementary Fig. S3, where a high ξ corresponds to the classical or clonal model of malignancies. Hence for each pair of values ψ and G used in the simulation, there is a corresponding ξ value.

We estimated a variable CSC fraction between tumors, with values between 0.5% and 1.5% for all cases apart from CX with 4% (Fig. 3, column A). These values are consistent with the fractions determined from functional assays based on $CD133^+$ human CRC cells (43, 44), but also reflect the variability of the CSC organization in different CRC patients (45). The rate of symmetric divisions also varies between tumors, with values ranging from 1% for tumor Z to 24% for tumor CU (column B). We observed cases with a large TAC population (parameter G—column C) such as tumors CT and CU, and others that do not have a TAC compartment at all (CX and Z). However, the inference for G is not optimal because of the dim molecular signal left by such short-living TACs (see Supplementary Material and Methods and Supplementary Fig. S2). By inferring a CSC organization in human tumors, these results suggest that CRC is a CSC-driven malignancy. Moreover, for the first time we quantify variability in the cellular organization and hierarchical structure of tumors that results from differences in the (a)symmetric division rate of CSCs and the replicative potential of TACs. Our results also confirm an elevated methylation rate in tumors with respect to the normal tissue (17), varying across patients from 10^{-4} to 10^{-3} errors per CpG per division (Fig. 3D). In addition, we observe context-dependent methylation (column E) in a subset of tumors (CX, HA, and Z), where methylation tends to increase in already methylated loci. We predict a relative tumor age, defined here as the time from the emergence of the first malignant cancer cell until the time of surgical resection, between 12 and 39 months (column F). Importantly, this value does not reflect the time during which mutations accumulated before the development of the malignancy (e.g., during the adenoma stage). Finally, we obtain an apoptotic rate on the order of 0.5% per cell division in all tumors with HA and Z showing even lower rates (0.3% and 0.2%, respectively). Notably, we observe that even relatively low apoptotic rates have considerable impact on slowing tumor growth.

For each tumor, these characteristics can be summarized into a patient-specific cancer profile that illustrates the differences in the characteristics and in the cellular organization of each tumor (Fig. 4). For example, CT displays a very small CSC population and a considerable TAC compartment, accompanied by relatively high mutation and apoptotic rates and a tumor age of 12 months. Tumor CU shares similar features, but

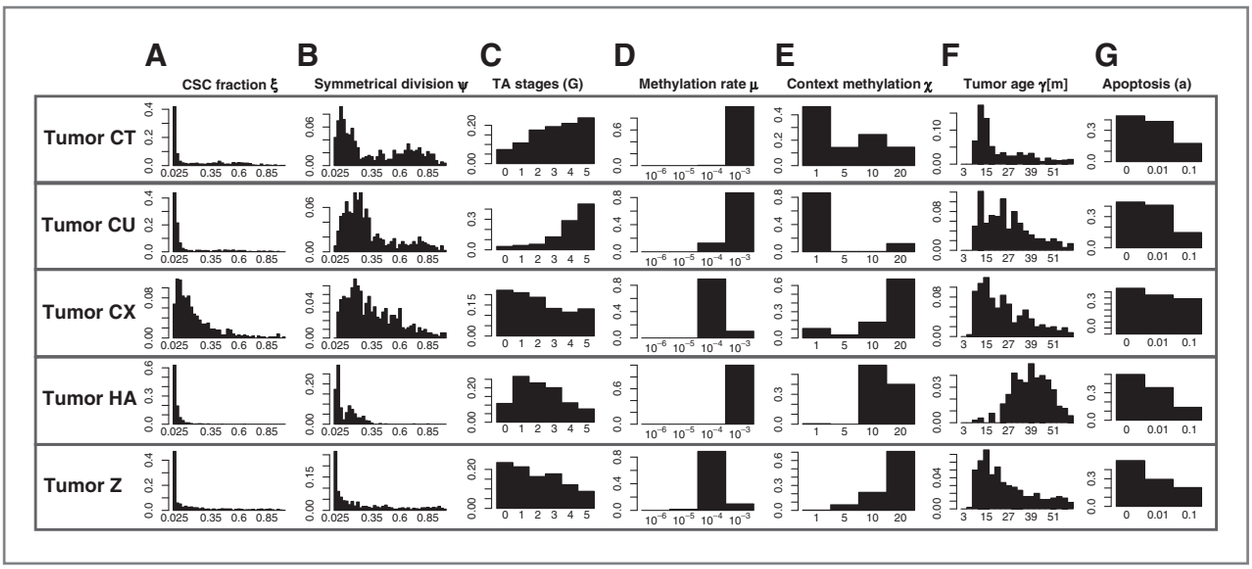

Figure 3. Distributions of clinically relevant tumor parameters measured with SCAI. We report that the tumors show a variable CSC fraction between 0.5% and 4% (A), with a symmetric division rate between 1% and 24% (B). The number of TA stages also varies between patients with only CT and CU showing a 5 stage TAC compartment (C). The methylation rates were 10- to 100-fold higher than normal (D) and context-dependent methylation was present in 3/5 cancers (E). F, the relative tumor age was predicted to range between 12 and 39 months. The apoptotic rate was low (~0.5% dead cells per cell cycle) in all tumors.

exhibits a higher symmetric division rate and hence a larger CSC fraction. CX contains a large CSC population and no TAC compartment; it features a lower mutation rate, but still develops in 12 months potentially because of the lack of TAC proliferation. This tumor is less consistent with a model where a small fraction of cells drive tumor growth. Tumor HA has a small CSC population and also a limited TAC compartment, high mutation rate, but lower apoptosis rate with an estimated tumor age of 39 months. Finally, tumor Z exhibits a small CSC fraction, no TAC compartment, and a low apoptotic rate with an estimated tumor age of 15 months. For tumor Z, using 2 additional validated molecular clocks (Supplementary Fig. S4

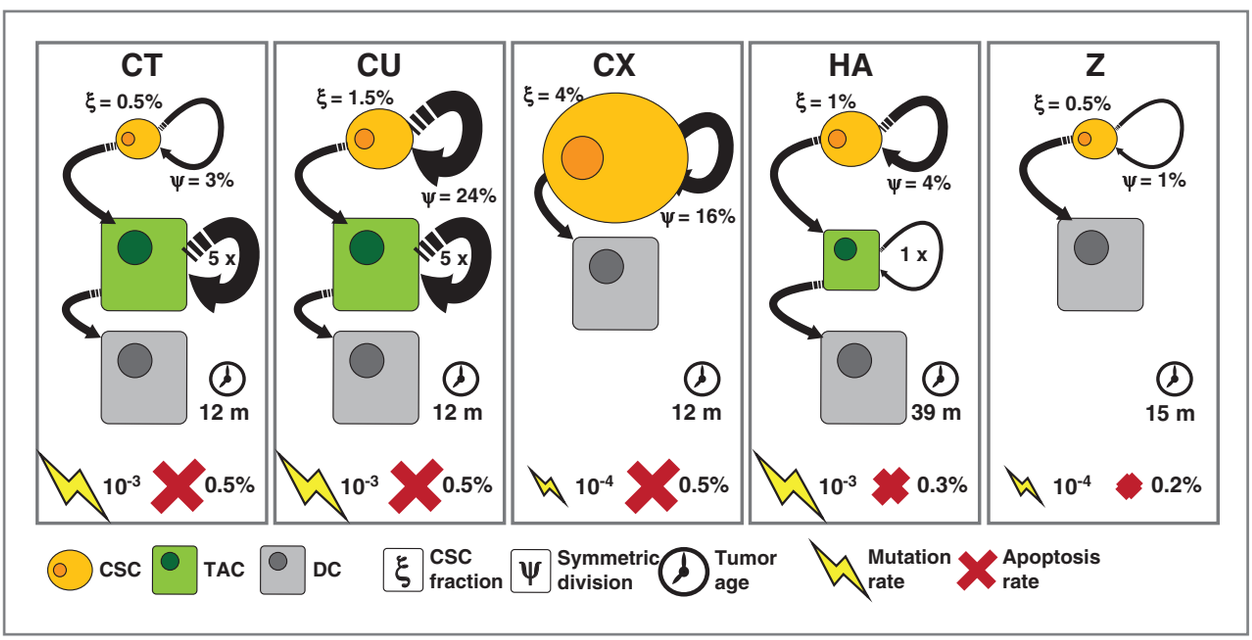

Figure 4. Comparisons of patient-specific cancer profiles. For each patient, we independently estimated a set of tumor characteristics that correspond to different cellular organizations, mutation and apoptosis rates, and tumor ages. These results reveal CSC organization in human tumors, based on the underlying molecular signature, and highlight variability in the hierarchical structure of the cancer cell population between patients.

and Table S1), we confirmed the same results (Supplementary Figs. S5 and S6 and Material and Methods).

Reconstructing tumor ancestral trees

Using the patient-specific parameters, we can retrace the ancestral tree of each tumor, providing insight into tumor evolutionary dynamics. We simulated the growth of the malignancy within our framework for each patient, virtually sampled 4 glands per tumor side and represented the phylogeny of 5 cells from each gland (Fig. 5). Each tree reflects the cellular organization of the neoplasm. Indeed, the tumors with a low CSC fraction (CT, CU, HA, and Z) show a limited number of long-lived lineages from which the cells in the glands derive. Those tumors also display an overall clustering of the glands by side. On the contrary, tumor CX shows a mixture of long-lived lineages because of the higher CSC fraction, in line with the classical interpretation of a largely tumorigenic malignancy. Tumor CX is also the largest neoplasm in our dataset (see Table 1), possibly reflecting the higher proliferative potential conferred by a larger population of CSCs.

Discussion

Deep-sequencing technologies have enabled the generation of high-throughput molecular data, which, when coupled with careful tumor sampling schemes, can provide new insight into tumorigenesis. We present a novel framework to trace the ancestries of cell populations within individual cancers and to quantify intratumor heterogeneity in primary human malignancies, while accounting for spatial processes and cellular dynamics. Importantly, this approach enables the measurement of tumor-specific characteristics from patient molecular data. Taken together, the resulting information defines a patient-specific tumor profile, which can be used to inform targeted treatment options.

Here, we interrogated intratumor heterogeneity at multiple levels of resolution, namely, between spatially separated tumor regions, between and within cancer glands, and summarized for the entire neoplasm (Supplementary Table S2). The spatial relationship between glands is evident from the molecular data, but it is accompanied by clonal heterogeneity indicative of a complex hierarchy of mitotic clones, spanning multiple spatial scales. We speculate that CSCs promote intratumor heterogeneity on the smallest scale (gland-size populations), and that clonal variation increases with spatial distance, driven by larger and larger groups of CSC lineages. As such, neighboring groups of CSCs would share a common ancestor and would have more similar genomic profiles.

Using our novel experimental and computational framework, we quantify CSC numbers in CRC without the need for xenotransplantation assays. Our results complement the evidence from transplantation experiments concerning the diversity in differentiation between cells from the same tumor. In particular, we infer a CSC fraction approximately 1%, consistent with xenotransplantation results using the $CD133^{+}$ marker (43). Moreover, the CSC fraction exhibits variability across patients, as also reported by xenotransplantation assays (45). The combination of CSC characteristics (symmetric division rate and number of transit amplifying stages) indicates radically different hierarchical organization for different tumors, which also exhibit different mutation and apoptosis rates (Fig. 4). The set of parameters for each patient represent a tumor profile that can be used to simulate the ancestral history of the sampled glands. Using our model, we simulated the entire history of a tumor using the inferred parameters (based on the modal value of the posterior) to generate a representative phylogenetic tree. We report clear differences between patients, which reflect the CSC organization responsible for neoplastic growth, with malignancies bearing a small fraction of CSC exhibiting fewer long-lived lineages (Fig. 5). The reconstructed trees and underlying data suggest that large clonal expansions are uncommon once the malignancy is initiated. Rather, the data suggest a complex and continuous interplay

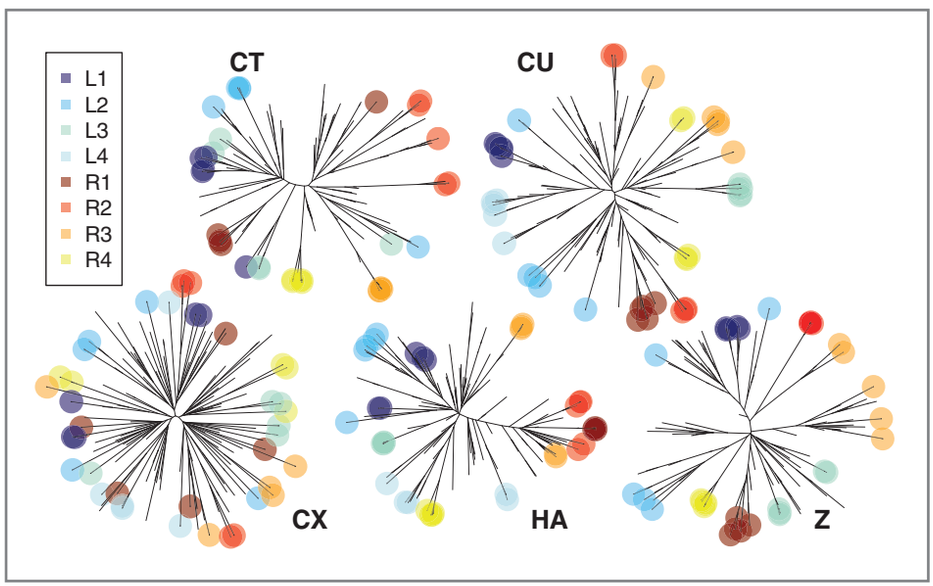

Figure 5. Reconstruction of tumor ancestral trees. We retraced the tumor phylogeny by simulating a neoplasm with the inferred parameters for each patient based on sampling 5 cells from 8 cancer glands per tumor (4 per side). In all tumors that exhibit a small CSC fraction (CT, CU, HA, and Z), the tree structure clearly underlines a small number of long-living CSC lineages. For tumor CX, with a large CSC population, the tree reflects the mixture of a large number of long-living lineages, indicating a larger population of cells with tumorigenic capacity.

between different coexisting clones, driven by CSCs. CSCs also contribute to intratumor heterogeneity by establishing the distinct cell lineages observed within the cancer glands. In particular, we observe that the majority of glands contained multiple clones, harboring 2 to 5 unique lineages. On the basis of the molecular signature of the CSC compartment, our results also indicate that CSCs are not quiescent as is sometimes suggested, but rather represent a mitotically active population that evolves and contributes to tumor growth and clonal diversity (12).

The CSC architecture may reflect features of the normal crypt stem cell structure driven by the Wnt signaling pathway (9). We have previously shown that the human colon crypt is composed of 15 to 20 stem cells, about 1% of the total 2,000. In CRC, our results suggest a similar fraction of CSCs, with values consistently on the order of 1%. In the normal crypt, the stem cell population is in homeostasis whereby 1 daughter cell remains in the niche as a stem cell whereas the other differentiates and leaves the base of the crypt. In tumors, CSCs are thought to arise because of mutations in normal stem cells (46), leading to an increase in the number of aberrant symmetric divisions and expansion of the CSC pool. The loss of asymmetric stem cell divisions was implicated in the oncogenicity of *APC* (adenomatous polyposis coli) mutations in the gut (47). Importantly, our approach provides quantitative estimates of the symmetric division rate ψ, which we observe to be low for normal colonic crypts (2.5%; ref. 17), but up to 25% in some tumors (Fig. 4, tumor CU). As we previously noted, CSC organization may have an important effect on tumor dynamics, mediated through the extent of intratumor heterogeneity (48, 49) and by conferring a selective growth advantage during tumor progression and therapy (50).

Our multisampling scheme, coupled with ultradeep sequencing of an informative molecular clock enables the interrogation of geographically separated areas of the tumor, providing a panoramic view of the heterogeneity within the neoplasm. Using a novel computational framework and a multisampling scheme, this approach elucidates on the cellular dynamics that regulate tumor growth in individual patients, with implications for personalizing patient treatment and potentially informing prognosis. Our framework is based on data collected from readily available biopsy or resection specimens, and allows for the generation of patient-specific cancer profiles within days from the time of collection.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: A. Sottoriva, D. Shibata, C. Curtis, S. Tavaré
Development of methodology: A. Sottoriva, I. Spiteri, D. Shibata, C. Curtis, S. Tavaré
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): I. Spiteri, D. Shibata, C. Curtis
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A. Sottoriva, I. Spiteri, D. Shibata, C. Curtis, S. Tavaré
Writing, review, and/or revision of the manuscript: A. Sottoriva, I. Spiteri, D. Shibata, C. Curtis, S. Tavaré
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): I. Spiteri
Study supervision: C. Curtis, S. Tavaré

Acknowledgments

The authors acknowledge the support of the University of Cambridge, Hutchinson Whampoa, and the University of Southern California. The authors also thank Louis Vermeulen and Daniel Andrews for useful discussions.

Grant Support

This research was funded by Cancer Research UK.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 13, 2012; revised August 20, 2012; accepted October 3, 2012; published OnlineFirst October 22, 2012.

References

1. Medema JP, Vermeulen L. Microenvironmental regulation of stem cells in intestinal homeostasis and cancer. Nature 2011;474:318–26.
2. Soltanian S, Matin MM. Cancer stem cells and cancer therapy. Tumour Biol 2011;32:425–40.
3. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature 2001;414:105–11.
4. Vermeulen L, Sprick MR, Kemper K, Stassi G, Medema JP. Cancer stem cells—old concepts, new insights. Cell Death Differ 2008;15:947–58.
5. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990;61:759–67.
6. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin 2010;60:277–300.
7. Todaro M, Alea MP, Di Stefano AB, Cammareri P, Vermeulen L, Iovino F, et al. Colon cancer stem cells dictate tumor growth and resist cell death by production of interleukin-4. Cell Stem Cell 2007;1:389–402.
8. Pohl A, Lurje G, Kahn M, Lenz HJ. Stem cells in colon cancer. Clin Colorectal Cancer 2008;7:92–8.
9. Vermeulen L, De Sousa EMF, van der Heijden M, Cameron K, de Jong JH, Borovski T, et al. Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. Nat Cell Biol 2010;12:468–76.
10. Clevers H. The cancer stem cell: premises, promises and challenges. Nat Med 2011;17:313–9.
11. Shackleton M, Quintana E, Fearon ER, Morrison SJ. Heterogeneity in cancer: cancer stem cells versus clonal evolution. Cell 2009;138:822–9.
12. Greaves M, Maley CC. Clonal evolution in cancer. Nature 2012;481:306–13.
13. Driessens G, Beck B, Caauwe A, Simons BD, Blanpain C. Defining the mode of tumour growth by clonal analysis. Nature 2012;488:527–31.
14. Schepers AG, Snippert HJ, Stange DE, van den Born M, van Es JH, van de Wetering M, et al. Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. Science 2012;337:730–5.
15. Tsao JL, Yatabe Y, Salovaara R, Jarvinen HJ, Mecklin JP, Aaltonen LA, et al. Genetic reconstruction of individual colorectal tumor histories. Proc Natl Acad Sci U S A 2000;97:1236–41.
16. Frumkin D, Wasserstrom A, Kaplan S, Feige U, Shapiro E. Genomic variability within an organism exposes its cell lineage tree. PLoS Comput Biol 2005;1:e50.
17. Yatabe Y, Tavaré S, Shibata D. Investigating stem cells in human colon by using methylation patterns. Proc Natl Acad Sci U S A 2001;98:10839–44.
18. Nicolas P, Kim KM, Shibata D, Tavaré S. The stem cell population of the human colon crypt: analysis via methylation patterns. PLoS Comput Biol 2007;3:e28.

19. Sottoriva A, Tavaré S. Integrating Approximate Bayesian Computation with complex agent-based models for cancer research. In:Saporta G, Lechevallier Y, editors, COMPSTAT 2010—Proceedings in computational statistics. Berlin: Springer, Physica Verlag 2010;57–66.
20. Siegmund KD, Marjoram P, Woo YJ, Tavaré S, Shibata D. Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. Proc Natl Acad Sci U S A 2009;106:4828–33.
21. Siegmund KD, Marjoram P, Tavaré S, Shibata D. High DNA methylation pattern intratumoral diversity implies weak selection in many human colorectal cancers. PLoS ONE 2011;6:e21657.
22. Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. Genetics 2002;162:2025–35.
23. Marjoram P, Tavaré S. Modern computational approaches for analysing molecular genetic variation data. Nat Rev Genet 2006;7:759–70.
24. Shibata D. Mutation and epigenetic molecular clocks in cancer. Carcinogenesis 2011;32:123–8.
25. TCGA Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330–7.
26. Gilles A, Meglécz E, Pech N, Ferreira S, Malausa T, Martin J-F. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC Genomics 2011;12:245.
27. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;100: 57–70.
28. Dormann S, Deutsch A. Modeling of self-organized avascular tumor growth with a hybrid cellular automaton. In Silico Biol 2002;2:393–406.
29. Jiang Y, Pjesivac-Grbovic J, Cantrell C, Freyer JP. A multiscale model for avascular tumor growth. Biophys J 2005;89:3884–94.
30. Anderson AR, Weaver AM, Cummings PT, Quaranta V. Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. Cell 2006;127:905–15.
31. Wong W-M, Mandir N, Goodlad RA, Wong BCY, Garcia SB, Lam S-K, et al. Histogenesis of human colorectal adenomas and hyperplastic polyps: the role of cell proliferation and crypt fission. Gut 2002;50: 212–7.
32. Wright NA, Poulsom R. Top down or bottom up? Competing management structures in the morphogenesis of colorectal neoplasms. Gut 2002;51:306–8.
33. Shih I-M, Wang T-L, Traverso G, Romans K, Hamilton SR, Ben-Sasson S, et al. Top-down morphogenesis of colorectal tumors. PNAS 2001;98:2640–5.
34. Barker N, van Es JH, Kuipers J, Kujala P, van den Born M, Cozijnsen M, et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature 2007;449:1003–7.
35. Lopez-Garcia C, Klein AM, Simons BD, Winton DJ. Intestinal stem cell replacement follows a pattern of neutral drift. Science 2010; 330:822–5.
36. Rew DA, Wilson GD, Taylor I, Weaver PC, Rew DA, Wilson GD, et al. Proliferation characteristics of human colorectal carcinomas measured *in vivo*, Proliferation characteristics of human colorectal carcinomas measured *in vivo*. Br J Surg 1991;78:60–6.
37. Beaumont MA, Rannala B. The Bayesian revolution in genetics. Nat Rev Genet 2004;5:251–61.
38. Tavare S, Balding DJ, Griffiths RC, Donnelly P. Inferring coalescence times from DNA sequence data. Genetics 1997;145:505–18.
39. Fu YX, Li WH. Estimating the age of the common ancestor of a sample of DNA sequences. Mol Biol Evol 1997;14:195–9.
40. Li M, Vitányi PMB. An introduction to Kolmogorov complexity and its applications. New York: Springer; 2008.
41. Shannon CE. A mathematical theory of communication. Bell Syst Tech J 1948;27:379–423.
42. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. Biochim Biophys Acta 2010;1805:105–17.
43. Ricci-Vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, Peschle C, et al. Identification and expansion of human colon-cancer-initiating cells. Nature 2007;445:111–5.
44. O'Brien CA, Pollett A, Gallinger S, Dick JE. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature 2007;445:106–10.
45. Dieter SM, Ball CR, Hoffmann CM, Nowrouzi A, Herbst F, Zavidij O, et al. Distinct types of tumor-initiating cells form human colon cancer tumors and metastases. Cell Stem Cell 2011;9:357–65.
46. Barker N, Ridgway RA, van Es JH, van de Wetering M, Begthel H, van den Born M, et al. Crypt stem cells as the cells-of-origin of intestinal cancer. Nature 2009;457:608–11.
47. Quyn AJ, Appleton PL, Carey FA, Steele RJC, Barker N, Clevers H, et al. Spindle orientation bias in gut epithelial stem cell compartments is lost in precancerous tissue. Cell Stem Cell 2010;6:175–81.
48. Sottoriva A, Verhoeff JJ, Borovski T, McWeeney SK, Naumov L, Medema JP, et al. Cancer stem cell tumor model reveals invasive morphology and increased phenotypical heterogeneity. Cancer Res 2010;70:46–56.
49. Sottoriva A, Sloot PM, Medema JP, Vermeulen L. Exploring cancer stem cell niche directed tumor growth. Cell Cycle 2010; 9:1472–9.
50. Sottoriva A, Vermeulen L, Tavaré S. Modeling evolutionary dynamics of epigenetic mutations in hierarchically organized tumors. PLoS Comput Biol 2011;7:e1001132.