Perspective

Counting Divisions in a Human Somatic Cell Tree
How, What and Why?

Darryl Shibata[1],*

Simon Tavaré[2,3]

[1]Department of Pathology; University of Southern California Keck School of Medicine; Los Angeles, California, USA

[2]Department of Biological Sciences; University of Southern California; Los Angeles, California USA

[3]Department of Oncology; University of Cambridge; Cambridge, UK

*Correspondence to: Darryl Shibata; Department of Pathology; University of Southern California, Keck School of Medicine; 1200 North State Street; Unit I, Room 2428; Los Angeles, California 90033 USA; Tel.: 323.226.7067; Fax: 323.226.2686; Email: dshibata@usc.edu

ABSTRACT

The billions of cells within an individual can be organized by genealogy into a single somatic cell tree that starts from the zygote and ends with present day cells. In theory, this tree can be reconstructed from replication errors that surreptitiously record divisions and ancestry. Such a molecular clock approach is currently impractical because somatic mutations are rare, but more feasible measurements are possible by substituting instead the 5' to 3' order of epigenetic modifications such as CpG methylation. Epigenetic somatic errors are readily detected as age-related changes in methylation, which suggests certain adult stem cells divide frequently and "compete" for survival within niches. Potentially the genealogy of any human cell may be reconstructed without prior experimental manipulation by merely reading histories recorded in their genomes.

INTRODUCTION

Genomes are almost identical copies of copies. The basis of this Perspective is the "almost identical" duplication of genomes—the idea that replication errors inevitably occur and surreptitiously record cell divisions. In theory, any genome can be traced from errors that accumulate as it is copied and physically passed from generation to generation. Molecular clock approaches compare genomes between species,[1] which are many copies removed from the original or source sequence present in a common ancestor (Fig. 1A). Greater numbers of sequence differences correlate with greater times or copies since a common ancestor.

Ancestral trees have been reconstructed from DNA or protein sequences for a wide variety of organisms. By analogy, it is easy to imagine a human somatic cell tree that starts from the zygote and ends with present day cells (Fig. 1B). There are billions of somatic cells so there will be billions of branches, which become progressively fewer further back in time because all cells are related to each other. Much of the broad structure of a human somatic cell tree is known because the embryonic origins of most tissues are understood. Many details remain uncertain because experimental methods commonly employed to trace cell fates are generally impractical in humans. However, it may still be possible to reconstruct human cell fates without prior manipulation by merely reading histories surreptitiously recorded within their genomes.

HOW TO COUNT LIFETIME NUMBERS OF HUMAN CELL DIVISIONS

A human somatic cell tree expands with age, with longer branches in older individuals (Fig. 1B). Cells divide at different rates, and therefore it is important to distinguish between chronological age (time since birth) and mitotic age (lifetime numbers of divisions). Cells within most individuals have identical chronological ages in the sense that they are offspring of the same zygote, but mitotic ages likely differ between tissues.

With few exceptions such as lymphocytes, essentially all somatic cells have sequences almost identical to the zygote. In theory, cells with greater mitotic ages should contain greater numbers of replication errors or somatic mutations. However, testing this prediction with sequences is currently impractical because normal replication fidelity is high. Somatic mutation frequencies are typically one per million bases in cancer genomes.[2] One approach to move from theory to feasible experimentation is to substitute the 5' to 3' order of bases with the 5' to 3' order of epigenetic modifications, which also exhibit somatic inheritance but less replication fidelity.

Figure 1. Ancestral trees physically connect genomes. (A) Life is thought to originate from a most recent ultimate common ancestor that existed billions of years ago. A species "tree of life" reconstructed from sequences indicates ancestry and how long ago common ancestors existed. (B) By analogy, cells within an individual are related by a single somatic cell tree that starts with the zygote and ends with present day cells. Branch lengths may be drawn proportional to numbers of divisions since the zygote (mitotic age). With aging, branches lengthen for mitotically active lineages. Common ancestors (circles) exist in all trees. The zygote is the universal common ancestor, and adult stem cells are more recent common ancestors.

Epigenetic modifications can encode information without changing base sequences. For example, methylation of cytosine may occur at CpG dinucleotides and is correlated with promoter activity,[3] which can allow for many different cell types from the same genome. This covalent modification is not copied during DNA replication, and a newly synthesized DNA molecule is hemi-methylated. DNA methyltransferases can add a new methyl group opposite a hemi-methylated CpG, and replicate the original pattern. However, replication cannot be perfect and methylation patterns, especially in CpG rich regions lacking function, may degrade with age.

Most CpG islands lack methylation at birth.[3] Therefore, replication errors would tend to increase methylation. Recent studies illustrate that methylation increases with age in certain tissues and CpG rich regions.[4,5] One can translate age-related methylation into ordered 5' to 3' binary ("0" for unmethylated and "1" for methylated) information strings suitable for a molecular clock analysis by cloning PCR products of bisulfite treated DNA and sequencing individual bacterial clones.[6] Starting from an unmethylated state, each methylated site represents at least one error (although demethylation or back errors are also possible). In this way, one can sample the 5' to 3' order of methylation at a given series of CpG sites, "count" the number of changes or errors since the zygote (mitotic age), and determine whether patterns are similar within a population of human cells (ancestry).

Different methylation patterns are present between and within individual intestinal crypts[6,7] or endometrial glands[8] from the same individual (Fig. 2). Therefore, although average methylation increases with age (Fig. 3), mechanisms underlying age-related methylation appear to be stochastic and consistent with random replication errors. Age-related methylation appears unrelated to function. For example, one epigenetic molecular clock contains 8 CpG sites in the 3' untranslated region of the CSX locus (Fig. 2). This gene is primarily expressed in the heart and therefore methylation is unlikely to confer selection in other tissues. Replication error rates differ between CpG islands because age-related changes range from undetectable (most CpG sites) to varying degrees of methylation in older individuals. For the CSX epigenetic molecular clock, the error rate has been estimated at ~2×10^{-5} per CpG site per division,[6] which is about 10,000 times greater than base replication errors (~10^{-9} per division). Age-related methylation does not appear secondary to a catastrophic loss of methylation fidelity because within one gland or crypt, one allele or locus may be methylated and the other completely unmethylated, which is more consistent with stochastic errors.
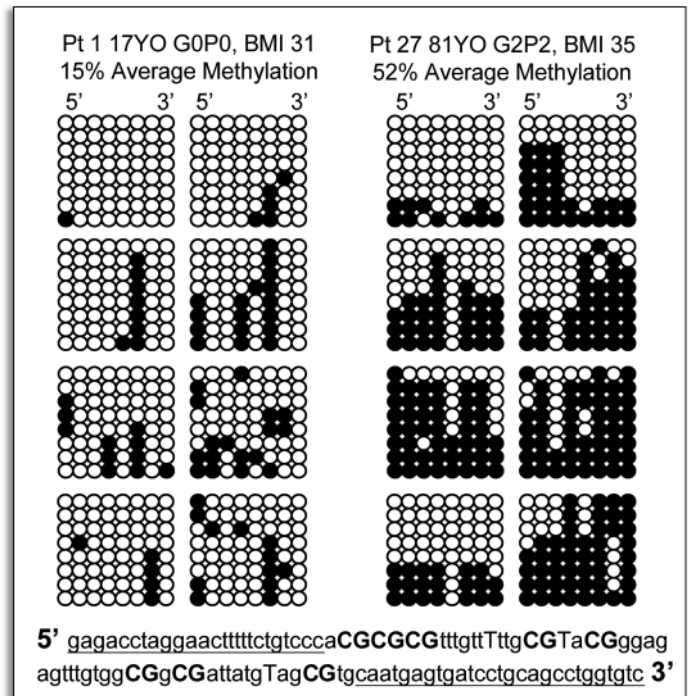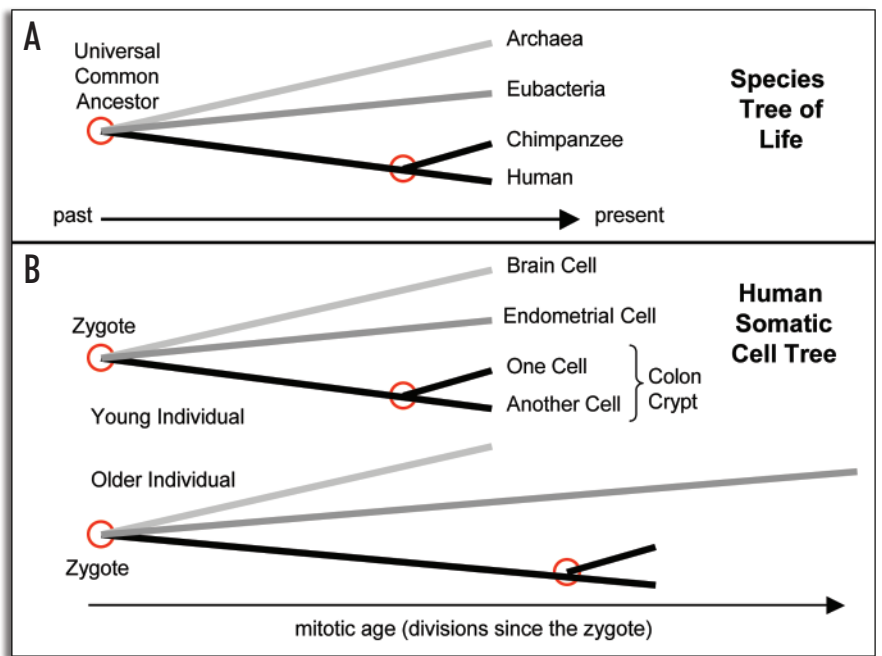

Figure 2. Epigenetic patterns sampled from single endometrial glands. There are eight glands each from two individuals. Eight alleles are sampled from each gland (64 total sites examined per uterus). Methylation patterns are represented horizontally in a 5' to 3' order (unmethylated sites are open circles and methylated sites are filled circles). The bisulfite converted CSX sequence is illustrated below. PCR primer sites are underlined, CpG sites are in bold, and C's at nonCpG sites converted by bisulfite are T's. Methylation is presumably absent at birth and more methylation is generally present in older individuals. However, stochastic replication errors can produce many different patterns—all of the cells within a gland or uterus have presumably divided the same number of times, even though some CSX alleles are more methylated. This is sort of like gamblers who start the same, play equivalent numbers of games, but finish with different winnings.
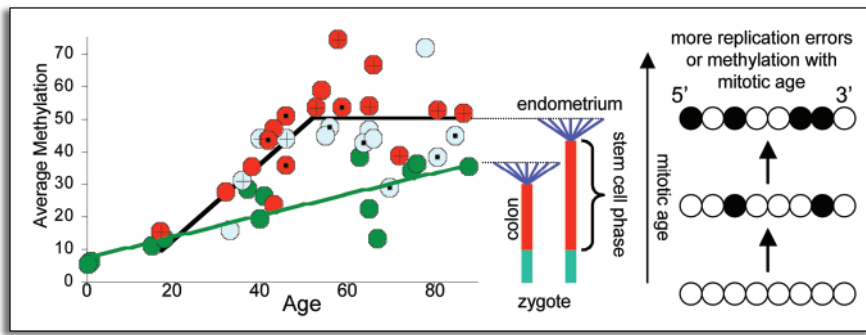
Figure 3. Average crypt or gland CSX methylation increases with age in the colon (green) and endometrium (black). Colon errors (methylation) increase linearly with age,[6] but endometrial methylation appears related to menstrual cycles—it increases between menarche and menopause, but is constant when menstruation ceases.[8] Red circles indicate two or fewer children, blue circles three or more children, plus marks indicate obesity (body mass index >30), and dots lean women (body mass index <24). Relatively more methylation is observed in women with fewer children or obesity. Genealogies at the right indicate that many, short lived, differentiated cells originate from a small number of stem cells, whose lineages ultimately originate from the zygote. Current methylation patterns reflect replication errors since the zygote (no CSX methylation), and age-related replication errors primarily reflect stem cell divisions (see Fig. 4).
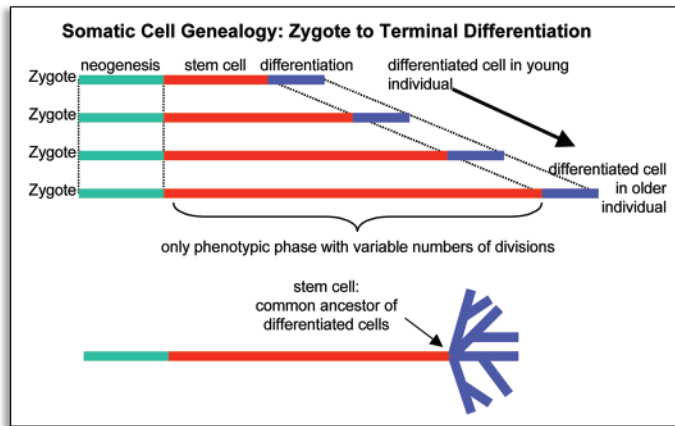

Figure 4. Stem cells are common ancestors because more numerous differentiated epithelial cells are progeny of stem cells. Somatic cell genealogies can be divided into three phases—neogenesis (from the zygote to the stem cell), stem cell latency, and differentiation. Age-related changes reflect stem cell divisions because development and differentiation are relatively short, fixed intervals, which are similar regardless of chronological age. The stem cell phase may lengthen with age because stem cell lineages are long-lived mitotic lineages.

Proving that tissue methylation (Fig. 2) is secondary to random replication errors is difficult because direct observations are impractical. Indeed, species molecular clocks have not been experimentally validated and exactly how to reconstruct the past from sequences remains controversial.[1] Molecular clocks depend on circumstantial evidence organized by the rigor of mathematical models. Generally molecular clocks require that almost identical copies are inherited and that rates of change are proportion to time or numbers of copies. The first criterion can be demonstrated by examining methylation patterns within and between individual clonal units such as single colon crypts. A single human colon crypt is a clonal structure and contains about 2,000 cells, most of which are differentiated progeny from a small number of stem cells.[9] Cells and their sequences should be more alike within a crypt compared to cells in other crypts in the

same colon. Consistent with methylation patterns representing copies of copies, methylation patterns are more alike within crypts than between crypts.[6]

Demonstrating that numbers of errors are proportional to mitotic age is more difficult because division is typically correlated with chronological time. For example, intestinal cell division is likely constant and therefore intimately related to time. Although division appears required for methylation in vitro,[10] age-related methylation may also include errors that accumulate independent of division. Endometrium provides an opportunity to disentangle chronological time from mitotic activity because women of the same age may have different numbers of menstrual cycles and therefore different mitotic ages. For example, women with more children should have fewer menstrual cycles, and obese women should have greater endometrial divisions because of chronic estrogen stimulation.[11] Age-related methylation occurs in human endometrial glands[8]—the numbers of errors (or percent CSX methylation) increases between menarche (around age 12 years) and menopause (around age 52 years), and is essentially constant after menopause, when menstruation ceases (Fig. 3). Moreover, numbers of errors or methylation are significantly higher in older women with fewer children compared to similar women with more children, and significantly higher in obese compared to lean women.[8] Therefore, age-related endometrial methylation appears more correlated with mitotic rather than chronological age.

A similar disentanglement between chronological age and mitotic age is also possible by measuring methylation in nonmitotic tissues such as the brain or heart. Preliminary data (not shown) demonstrate brain and heart CSX methylation is low before birth and higher in adults. However, average methylation is similar in young and old adults. Therefore, CSX clock methylation does not appear to change much in the absence of cell division.

WHAT DOES MITOTIC AGE MEASURE?

Mitotic age assigns a copy number to every cell. For example, the first two cells after the zygote have "copy number one". Subsequent cells would be copy number 2, 3, 4…… and so on. Potentially a "new" cell from a new division could have copy number ~25,000 in a 70 year old if prior cells divided everyday. Not every cell is endowed with the ability to make new copies because most differentiated cells cannot divide, and die within defined time periods. Stem cells have the ability to divide throughout life and replication errors can only accumulate in long-lived cells such as stem cells.

Mitotic tissues can be conceptually organized by genealogy into three distinct phenotypic phases—neogenesis or development between the zygote and tissue, stem cell latency, and differentiation (Fig. 4). Stem cells are largely uncharacterized because they are few in number and difficult to identify. However, stem cells are common ancestors in a somatic cell tree and produce large numbers of differentiated progeny, which are only a few copies removed from their stem cells (Fig. 4). Therefore, it is relatively easy to infer a stem cell genome by sampling its more numerous progeny.

Age-related methylation is evidence for stem cell mitotic activity because neogenesis and differentiation intervals are relatively short

and fixed—differentiated cells regardless of chronological age have similar numbers of neogenesis and differentiation divisions. The only phenotypic phase that can have variable numbers of divisions is the stem cell phase, and age-related methylation provides an indication of how often tissue stem cells divide (Fig. 4). Given this logic, endometrial stem cells are more mitotic compared to intestinal stem cells (Fig. 3).

WHY IS MITOTIC AGE IMPORTANT?

Mitotic age is likely a fundamental parameter in aging and cancer. Younger individuals with cells of lower copy numbers are generally healthier than older individuals composed of cells with much higher copy numbers. Genetic and environmental factors may modulate aging through cell division. Cancer frequencies increase with chronological age, and many somatic mutations that contribute to a malignant phenotype may represent replication errors. Epidemiology strongly suggests cell division is oncogenic, with greater cancer risks in individuals with more divisions.[12] Many carcinogens are mitogens.[13] However, it has been difficult to relate directly division with aging and cancer because it has been impractical to directly count lifetime divisions. Epigenetic molecular clock studies are consistent with greater divisions in women with higher risk factors (nulliparity and obesity) for endometrial cancer.[8] The ability to more directly infer tissue mitotic ages from easy- to- measure replication errors provides new opportunities to experimentally link cell division with human aging and diseases such as cancer.

Age-related methylation also implies stem cell evolution. Mammalian stem cells are thought to reside in niches that extrinsically define stem cells by location.[14] Niches are potentially small evolutionary crucibles because numbers of progeny are physically limited by the size of the niche (Fig. 5). Stem cells almost always divide asymmetrically to produce one stem cell daughter, and one daughter that leaves the niche and differentiates. However, stem cells can also divide symmetrically, and potentially one stem cell can produce two stem cell daughters, but only if another stem cell lineage becomes extinct by producing two differentiated daughters (Fig. 5). Rare symmetric divisions will eventually result in the loss of all current stem cell lineages except one, or niche clonal evolution.[14]

Imagining evolution can be difficult when changes occur slowly or without visible manifestations. Genomes, which accumulate changes, provide some of the best evidence for evolution because they are similar but different between individuals and species. Similarly, niche stem cell evolution is imperceptible because stem cells cannot be distinguished from each other, but methylation pattern differences between and within crypts of the same individual suggest stem cells divide and die during repeated cycles of clonal evolution.[6] The clonal evolution of tumor progression can be considered an abnormal visual manifestation of normal stem cell rhythm.

The mechanisms responsible for niche clonal evolution are uncertain, but may again reflect imperfections in any replication process. Stem cells almost always divide asymmetrically (>90% of
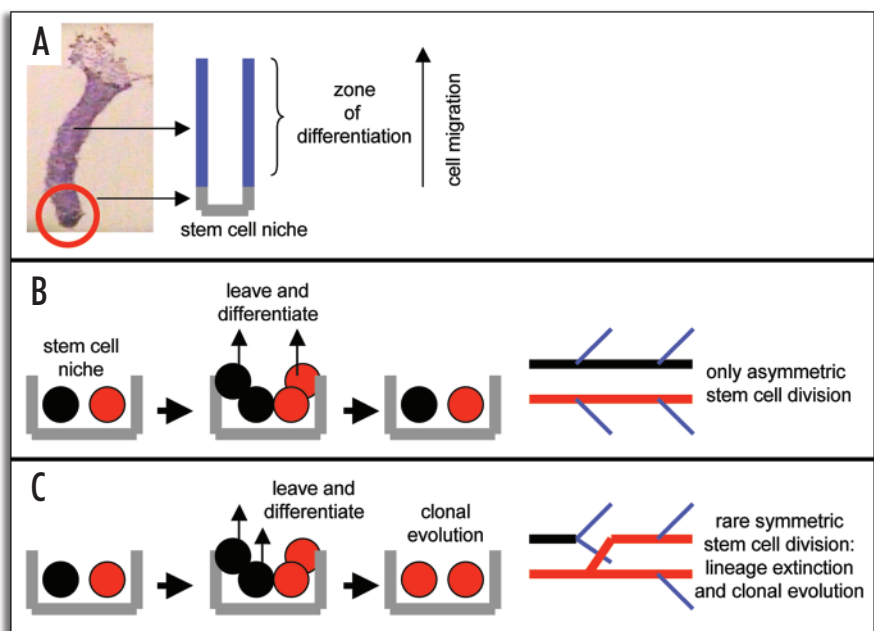

Figure 5. Stem cell niches are small evolutionary crucibles because stem cells divide and only a limited number of cells remain within the niche. (A) Approximate location of a niche in a free floating single colon crypt isolated from a fresh colon.[6] Stem cells lie within the niche, and cells that migrate out of the niche eventually differentiate and die. (B) A niche with two stem cells is illustrated (colon crypts may have more than two stem cells).[6,9] Stem cells most often divide asymmetrically to produce one stem cell daughter and one daughter that leaves the niche. If stem cells always divide asymmetrically, stem cell lineages survive a lifetime. (C) Symmetrical divisions produce two similar daughters, but result in lineage extinction or niche clonal evolution. Niche clonal evolution is normally imperceptible because one cannot distinguish between stem cells. However, numbers of divisions and ancestry are surreptitiously recorded by replication errors, and methylation patterns in adult colons indicate rare symmetrical divisions occur, with clonal evolution recurring on average every eight years.[6]

the time),[6,16] but even rare symmetrical divisions imply stem cell lineage extinction (Fig. 5). Niche clonal evolution appears to recur approximately every eight years in normal human colon crypts.[6] Niche stem cell dominance is likely due to chance or drift, but specific mutations, such as those in APC commonly found in cancers,[17] may also enhance niche persistence and dominance over normal stem cells. Evolution appears to operate whenever numbers of almost identical copies are limited—between individuals and the cells within an individual.

SUMMARY

Billions of human cells can be organized by genealogy into a single somatic cell tree (Fig. 1). Stem cells are common ancestors in a somatic cell tree and stem cell lineages connect present day cells with the past. Rather than static, guaranteed permanence, many human stem cells appear to divide frequently and "compete" for survival in niches. Stem cell division and evolution may underlie how and why we change with age, and help explain some of the diseases associated with aging. Simply substituting epigenetic modifications for sequences provides new opportunities to systematically apply modern molecular evolution evolutionary methods to human aging and diseases. Easy-to-measure epigenetic replication errors potentially allow genealogy reconstruction for any human cell without prior manipulations. This Perspective has emphasized the logic of epigenetic molecular clocks and neglected the mathematical and statistical modeling[6] that

is absolutely essential to reconstruct the past, especially when a multitude of scattered outcomes are consistent with stochastic errors (see legend to Fig. 2). There are other possible mechanisms for age-related methylation, but replication errors and stem cell evolution provide plausible explanations for seemingly random genome patterns that change with age. Further studies are needed to test how well our studies capture histories recorded by replication errors,[18] but in theory all genomes contain ancestral information waiting to be read by the literate.

References

1. Bromham L, Penny D. The modern molecular clock. Nat Rev Genet 2003; 4:216-24.
2. Wang TL, Rago C, Silliman N, Ptak J, Markowitz S, Willson JK, Parmigiani G, Kinzler KW, Vogelstein B, Velculescu VE. Prevalence of somatic alterations in the colorectal cancer cell genome. Proc Natl Acad Sci USA 2002; 99:3076-80.
3. Bird A. *DNA* methylation patterns and epigenetic memory. Genes Dev 2002; 16:6-21.
4. Ahuja N, Li Q, Mohan AL, Baylin SB, Issa JP. Aging and DNA methylation in colorectal mucosa and cancer. Cancer Res 1998; 58:5489-94.
5. Issa JP. CpG-island methylation in aging and cancer. Curr Top Microbiol Immunol 2000; 249:101-18.
6. Yatabe Y, Tavaré S, Shibata D. Investigating stem cells in human colon by using methylation patterns. Proc Natl Acad Sci USA 2001; 98:10839-44.
7. Kim JY, Siegmund KD, Tavaré S, Shibata D. Age-related human small intestine methylation: Evidence for stem cell niches. BMC Med 2005; 3:10.
8. Kim JY, Tavaré S, Shibata D. Counting human somatic cell replications: Methylation mirrors endometrial stem cell divisions. Proc Natl Acad Sci USA 2005; 102:17739-44.
9. Potten CS, Loeffler M. Stem cells: Attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. Development 1990; 110:1001-20.
10. Velicescu M, Weisenberger DJ, Gonzales FA, Tsai YC, Nguyen CT, Jones PA. Cell division is required for de novo methylation of CpG islands in bladder cancer cells. Cancer Res 2002; 62:2378-84.
11. Pike MC, Pearce CL, Wu AH. Prevention of cancers of the breast, endometrium and ovary. Oncogene 2004; 23:6379-91.
12. Preston-Martin S, Pike MC, Ross RK, Jones PA, Henderson BE. Increased cell division as a cause of human cancer. Cancer Res 1990; 50:7415-21.
13. Ames BN, Gold LS. Too many rodent carcinogens: Mitogenesis increases mutagenesis. Science 1990; 249:970-1.
14. Spradling A, Drummond-Barbosa D, Kai T. Stem cells find their niche. Nature 2001; 414:98-104.
15. Calabrese P, Tavaré S, Shibata D. Pre-tumor progression: Clonal evolution of human stem cell populations. Am J Pathol 2004; 164:1337-46.
16. Marshman E, Booth C, Potten CS. The intestinal epithelial stem cell. Bioessays 2002; 24:91-8.
17. Kim KM, Calabrese P, Tavaré S, Shibata D. Enhanced stem cell survival in familial adenomatous polyposis. Am J Pathol 2004; 164:1369-77.
18. Ro S, Rannala B. Methylation patterns and mathematical models reveal dynamics of stem cell turnover in the human colon. Proc Natl Acad Sci USA 2001; 98:10519-21.