

Age-Specific Incidence Data Indicate Four Mutations Are Required for Human Testicular Cancers

James P. Brody*

Department of Biomedical Engineering, Center for Complex Biological Systems, University of California Irvine, Irvine, California, United States of America

Abstract

Normal human cells require a series of genetic alterations to undergo malignant transformation. Direct sequencing of human tumors has identified hundreds of mutations in tumors, but many of these are thought to be unnecessary and a result of, rather than a cause of, the tumor. The exact number of mutations to transform a normal human cell into a tumor cell is unknown. Here I show that male gonadal germ cell tumors, the most common form of testicular cancers, occur after four mutations. I infer this by constructing a mathematical model based upon the multi-hit hypothesis and comparing it to the age-specific incidence data. This result is consistent with the multi-hit hypothesis, and implies that these cancers are genetically or epigenetically predetermined at birth or an early age.

Citation: Brody JP (2011) Age-Specific Incidence Data Indicate Four Mutations Are Required for Human Testicular Cancers. PLoS ONE 6(10): e25978. doi:10.1371/journal.pone.0025978

Editor: Amanda Ewart Toland, Ohio State University Medical Center, United States of America

Received: June 11, 2011; **Accepted:** September 14, 2011; **Published:** October 6, 2011

Copyright: © 2011 James P. Brody. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported by the University of California. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jpbrody@uci.edu

Introduction

Tumors originate from a single cell after the cell accumulates a series of mutations [1–6], according to the multi-hit model of cancer [2,3]. These mutations can include many different types of alterations to the DNA including methylation, single base substitutions, and duplications or deletions of chromosomes. The exact number of mutations required to transform a normal human cell into a tumor cell is unknown [7].

Direct DNA sequencing of tumors has established an upper limit on the number of mutations required to transform a cell. Sequencing of breast and colorectal cancers identified about 80 mutations in a typical tumor [8]. Further statistical analysis suggested that less than 15 of those 80 are necessary [8]. A second experiment sequenced 623 known cancer-related genes in a set of 188 lung adenocarcinomas showing more than 1000 somatic mutations. Further analysis identified 26 genes that were concluded to be involved in carcinogenesis [9].

A lower limit on the number of mutations to transform a normal cell has been established in the laboratory. A human tumor cell was synthesized from normal human cells (both epithelial and fibroblast cells) by altering the expression of only three genes, which effected four biochemical pathways [6,10]. This tumor cell displayed the classic characteristics of a human tumor cell: anchorage-independent growth and formation of tumors in nude mice.

It is widely believed that colon tumors require four to six mutations [11]. This is based upon comparing the Armitage and Doll equation, $I = at^{r-1}$ [12], to age-specific incidence data. But many problems exist with this [13]: it fails to describe the data at older ages; it increases without any upper limits; and it does not incorporate clonal expansion.

Testicular gonadal germ cell cancers differ from other solid tumors in a number of ways. First, the incidence of gonadal germ

cell tumors is highest at about 30 years of age and declines to just a handful of cases diagnosed in men in their 70's. In comparison, the incidence of many other solid tumors increases with age. Second, combination chemotherapy is particularly effective against testicular gonadal germ cell tumors as compared to other solid tumors. Finally, most solid tumors originate in somatic cells, while most testicular cancers arise in germ cells.

The cause of testicular gonadal germ cell tumors is not known [14,15]. No known environmental factors affect its development [16]. A family link exists, stronger in brothers than father and sons [17]. Testicular cancers showed the third highest heritability, but most cases are sporadic [18]. Age standardized rates of testicular cancer have increased over the past few decades in the United States [19] and in other parts of the world [20].

The strongest association of testicular cancer with any other medical conditions is with cryptorchidism, where the testicles do not descend into the scrotum at birth. About 5 to 10% of those who develop testicular cancer had undescended testicles at birth, compared to about 2 to 5% in the general population [21]. It is not known whether cryptorchidism causes testicular cancers or whether both are caused by a common factor.

Two hypotheses exist for the origin of testicular cancers. The first suggests that testicular cancers are determined *in utero* or at an early age [22,23]. A second is that environmental exposure to carcinogens throughout ones lifetime leads to the development of a tumor, while genetics modifies this environmental risk [24]. Although this second hypothesis is widely believed, little evidence exists that environmental mutagens cause any of the point mutations observed in human cancers [25].

Tests of these hypotheses are mixed. A retrospective study of Swedish males found that those who underwent surgery before the age of 13 to correct undescended testicles had a slightly lower risk of developing testicular cancer than those who did not undergo

surgery [21]. This suggests that testicular cancers could not be predetermined at birth. However, a similar study containing almost twice as many subjects in Denmark found no significant change in the incidence of testicular cancer after surgery for undescended testicles [26].

Two genome wide association studies identified several mutations that predispose to the development of testicular tumors [27,28]. These mutations are located in two genes, *KITLG* and *SPRY4*, that are known to play a role in testicular development. The estimated per allele odds ratio for these are among the highest found for any genome wide association study of a cancer [29].

Previously, others have sought to understand the age-specific incidence of cancers with different approaches. Including using a Weibull distribution for lung cancer [30], analyzing the age-specific acceleration of cancers [31,32], modifying the Armitage Doll equation directly with a damping term [33–35], and using a multistage model with age-dependent behavior to estimate the number of mutations required to develop breast cancer [36]. An analysis of Danish and Norwegian cancer registries suggests that testicular cancer age-specific incidence data are best modeled with a frailty effect, where a portion of the population is non-susceptible to developing the cancer [37].

The objective of this paper is to determine how many mutations are required to develop testicular cancer. The approach is to compare the expected age-specific incidence, based upon the multi-hit model, with the measured age-specific incidence for testicular cancers.

Results

The age-specific incidence for testicular cancers is accurately described by Equation (1) with four mutations. Figure 1 shows a comparison between Equation (1) and data for all eight years. This also implies that testicular cancers develop from a single progenitor cell. Table 1 shows the parameters and error estimates, and p-value for individual years.

Figure 2 compares the best fits for models with three, four, and five mutations. The model with four mutations was the best fit. Three mutations provided a slightly worse fit in the 15–20 and 65–

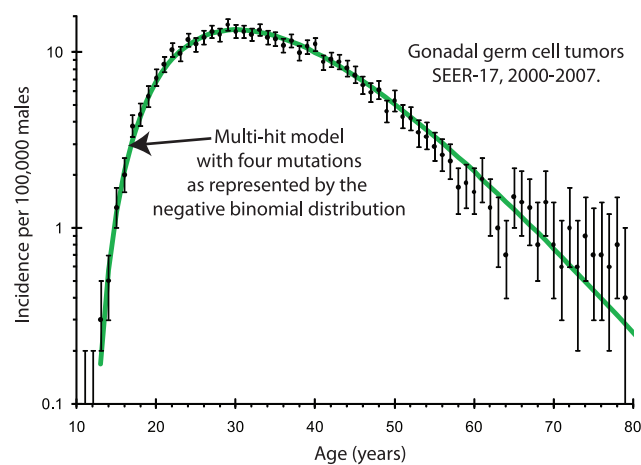


Figure 1. This is a comparison between the observed (SEER-17, 2000–2007) age-specific incidence for testicular cancers and that predicted by the multi-hit model with four mutations. The black circles represent the measured incidence, the error bars are 95% confidence intervals, and the green solid line represents the incidence predicted by the multi-hit model with four mutations. doi:10.1371/journal.pone.0025978.g001

Table 1. The best estimate of the parameters A (number of men per 100,000 who will ultimately develop testicular cancer in their lifetime) and q (the probability per year that no mutation occurs).

Year	P-value	A	q
2000	0.017	387	0.856
2001	0.043	381	0.856
2002	0.059	379	0.855
2003	0.497	370	0.854
2004	0.316	387	0.851
2005	0.004	381	0.851
2006	0.722	382	0.854
2007	0.350	376	0.846
	0.2606	380(± 5)	0.853(± 0.003)

I tested the hypothesis that the testicular cancer age-specific incidence data were derived from Equation 1. To obtain the best estimate of the parameters in Equation 1, I performed a least squares fit of Equation 1 to the age-specific incidence data for eight consecutive years (2000–2007). This table lists the best estimate of the parameters A (number of men per 100,000 who will ultimately develop testicular cancer in their lifetime) and q (the probability per year that no mutation occurs). It also lists the p-value, the probability that the hypothesis should not be rejected.

doi:10.1371/journal.pone.0025978.t001

75 age range, while five mutations provided a much worse fit in the range of 10–25 years, but was indistinguishable from 4 mutations in the 50–75 age range.

I measured the probability of advancing to the next stage at $(1 - 0.853 =)0.147$ per year. This measures only the probability of a mutation that would advance the precancerous tissue another step towards cancer, and is not directly comparable to measured mutation rates. Human germ line mutations vary across the genome by orders of magnitude [38]; a single mutation rate cannot accurately characterize the process.

Discussion

The approach presented here implicitly assumes that the probability of advancing to the next stage, which could be associated with a mutation rate, is constant. Complexity could be added to the model by modifying this assumption. At least two different mechanisms could modify the mutation rate in a pre cancerous tissue. First, the mutator phenotype hypothesis suggests that one of the first mutations on the path to a tumor must result in a higher mutation rate [39]. Second, the process of clonal expansion can expand the pool of cells at one stage, increasing the probability of advancing to the next stage [40–42]. Since the simplest assumption, a constant probability of advancing, was sufficient in this case, I did not extend the model to include a changing probability rate.

The mutations are most likely chromosomal additions or deletions, not single base alterations. Cytogenetic studies of seminoma and non-seminoma testicular cancers have shown consistent alterations to several chromosomes. In particular, amplification of a region of chromosome 12p containing several known genes is often present [43].

Although four mutations are required for the development of testicular cancers, these mutations may alter more than four genes and biochemical pathways. In addition, other mutations that are not rate limiting may occur. Non rate limiting mutations would not alter the age-specific incidence data.

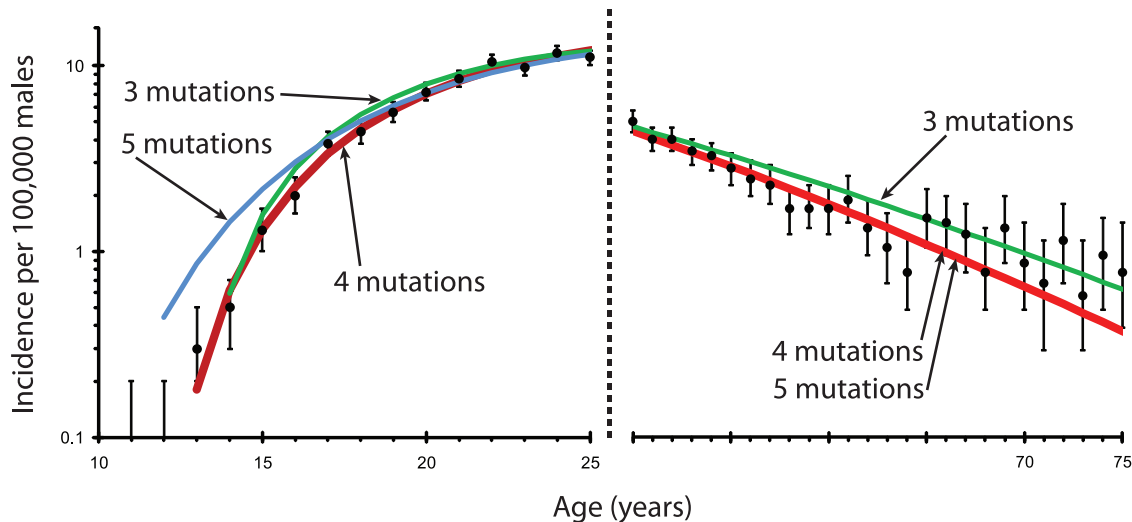


Figure 2. This figure compares the best fit models for three, four, and five mutations. To emphasize the differences, only regions before the age of 25 and after the age of 50 are shown. All three models show close fit to the data between 25 and 50 years of age. The models with four mutations and five mutations are identical after the age of 50. doi:10.1371/journal.pone.0025978.g002

One potential problem with this analysis is that it assumes no significant long term change in the rate of testicular cancers. The SEER-9 data show that the age-adjusted testicular cancer rate has increased by about 7% per year from 1973 to 2008. The standard way for dealing with temporal variation in cancer rates is to first analyze age models, then age plus drift, then age-period and/or age-cohort, and finally age-period-cohort models [44,45]. Each addition of complexity requires additional parameters and reduces the number of degrees of freedom. Since the age only model provided a good fit to the data, further complexity was avoided. However, future work on the age-specific incidence of testicular cancer should explore whether these more complex models provide alternative solutions.

Additional complexity to the model could be added in different ways. To account for inherited mutations, the model could consist of two independent terms similar to Equation (1). The first term would require $r - 1$ mutations and the second term would require r mutations. To account for multiple pathways by which testicular cancer could develop, the model could be extended by adding a second term with independent parameters from the first. Neither of these additions are necessary, but the data does not exclude the possibility of these more complex processes.

The agreement between the age-specific incidence data and Equation (1) implies that testicular cancers have a single potential progenitor cell. This contrasts with most other types of solid tumors which are thought to have many, many potential progenitor cells.

The age-specific incidence data implies that testicular cancers are pre-determined before the age of 10 and possibly at birth either through genetic or epigenetic [46] predisposition. This data is inconsistent with a hypothesis where exposure to environmental carcinogens in mature men lead to the development of a testicular tumor.

Methods

The multi-hit model describes a series of independent Bernoulli trials. A random number is drawn between zero and one. If the number is less than q , no mutation occurs; if greater than q , a mutation occurs. The process is repeated periodically. When r

mutations have occurred, a tumor begins to develop. The tumor grows, through clonal expansion, over an additional time τ until it is detected as a cancer. The time τ might also be related to normal growth and development. This process occurs in a fraction of the population, A , that lies somewhere between 0 and 100%.

Under these assumptions, the probability distribution for the age at which testicular cancer is diagnosed should be given by the solution to the series of independent Bernoulli trials, the negative binomial distribution [47]

$$p(t) = A \frac{((t-\tau)-1)!}{((t-\tau)-r)!(r-1)!} (1-q)^r q^{((t-\tau)-r)}. \quad (1)$$

The age-specific incidence measures the hazard function, which is related to Equation (1) by dividing by the ratio of the total population to the at-risk population. Males who have been previously diagnosed with testicular cancer are removed from the total population to produce the at-risk population. The effect of this is at most 0.3% on $p(t)$ and can be ignored in this case since it is overwhelmed by the predominant sampling error.

One assumption in the derivation of Equation (1) is that a single progenitor cell exists in the tissue. If many progenitor cells exist, as is thought to occur in most tissues, then cancer is diagnosed when the first of these many cells develops into a tumor. In this case, the first order statistic, or distribution of the minimum, of Equation (1) is the proper equation to compare to the age-specific incidence data. This would follow a Weibull distribution [30,48].

I tested the hypothesis that the age-specific incidence data on testicular germ cell tumors is accurately described by Equation (1). I performed a least squares fit to determine the parameters of the equation for all eight years in the dataset. Then, I calculated the reduced chi-squared value and the associated p-value, given the number of degrees of freedom, 53. The p-value, shown in Table 1, indicates the probability that the hypothesis should not be rejected.

In the United States, the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI) collects data on cancer cases. It is considered the gold-standard for data quality for cancer registries. It collects data from 17 different geographic regions that encompass just over 26% of the

Table 2. The best fit parameters for each model, along with the calculated P-value.

r	τ	A	q	P-value
3	11	386	0.884	10^{-14}
4	9	377	0.851	0.02
5	5	378	0.8388	10^{-12}

I tested models with 3, 4, and 5 rate-limiting steps fitting these models to combined 2000–2007 testicular cancer data. The best fit parameters for each model, along with the calculated P-value, are shown here.
doi:10.1371/journal.pone.0025978.t002

population of the United States [49]. This data is combined with US Census data on the population, as a function of age, in these 17 geographic areas to calculate the age-specific incidence.

I obtained testicular germ cell tumor age-specific incidence data using SEER*Stat (version 6.6.2) [50]. SEER*Stat allows one to easily query the SEER case files. I queried the database published in November 2009, the SEER 17 incidence database with single ages to 85+ [49]. This was the most recent available. I selected all reported tumors that were in males, located in the testis, and classified as germ cell or trophoblastic tumors or neoplasms of gonads, totaling 16,291 cases. I excluded testicular cancers diagnosed before the age of 13, because they were probably due to a different mechanism, 88 cases were excluded. Those

diagnosed before the age of four are probably teratoma-yolk sac tumors [43]. These account for only a small fraction (0.5%) of all testicular cancers. (From 2000–2007, the SEER-17 registries recorded 88 testicular germ cell cancers in patients under 13, 72 of these cancers were diagnosed in the first 36 months of life.)

I compared Equation (1) to age-specific incidence data collected by the SEER-17 cancer registries from 2000–2007 on the incidence of testicular cancers, both seminomas and non-seminomas. I compared these equations with the data from ages 13 to 70 years old and with the number of mutations, r , ranging from three to five. This comparison was made by minimizing the reduced chi-squared value [51] using the Generalized Reduced Gradient algorithm. This algorithm is suitable for minimizing non-linear functions. I used multiple starting points to ensure that the solution given was the global minimum and not a local minima.

I calculated error estimates by measuring the parameters A and q for eight individual years (2000–2007) and taking the standard deviation of these eight values. Uncertainty in the parameter, r , which represents the number of mutations required to develop a tumor, was measured by comparing models for three, four, and five mutations, as shown in Table 2.

Author Contributions

Conceived and designed the experiments: JB. Performed the experiments: JB. Analyzed the data: JB. Contributed reagents/materials/analysis tools: JB. Wrote the paper: JB.

References

1. Fearon ER, Hamilton SR, Vogelstein B (1987) Clonal analysis of human colorectal tumors. *Science* 238: 193–197.
2. Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. *Cell* 61: 759–767.
3. Vogelstein B, Kinzler KW (1993) The multistep nature of cancer. *Trends Genet* 9: 138–141.
4. Hanahan D, Weinberg RA (2000) The hallmarks of cancer. *Cell* 100: 57–70.
5. Hahn WC, Weinberg RA (2002) Modelling the molecular circuitry of cancer. *Nat Rev Cancer* 2: 331–341.
6. Hahn WC, Weinberg RA (2002) Rules for making human tumor cells. *N Engl J Med* 347: 1593–1603.
7. Stratton MR, Campbell PJ, Futreal PA (2009) The cancer genome. *Nature* 458: 719–724.
8. Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, et al. (2007) The genomic landscapes of human breast and colorectal cancers. *Science* 318: 1108–1113.
9. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, et al. (2008) Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 455: 1069–1075.
10. Hahn WC, Counter CM, Lundberg AS, Beijersbergen RL, Brooks MW, et al. (1999) Creation of human tumour cells with defined genetic elements. *Nature* 400: 464–468.
11. Byrne HM (2010) Dissecting cancer through mathematics: from the cell to the animal model. *Nat Rev Cancer* 10: 221–230.
12. Armitage P, Doll R (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer* 8: 1–12.
13. Moolgavkar SH (2004) Commentary: Fifty years of the multistage model: remarks on a landmark paper. *Int J Epidemiol* 33: 1182–1183.
14. Oosterhuis JW, Looijenga LHJ (2005) Testicular germ-cell tumours in a broader perspective. *Nat Rev Cancer* 5: 210–222.
15. Houldsworth J, Korkola JE, Bosl GJ, Chaganti RSK (2006) Biology and genetics of adult male germ cell tumors. *J Clin Oncol* 24: 5512–5518.
16. Manecksha RP, Fitzpatrick JM (2009) Epidemiology of testicular cancer. *BJU Int* 104: 1329–1333.
17. Westergaard T, Olsen JH, Frisch M, Kroman N, Nielsen JW, et al. (1996) Cancer risk in fathers and brothers of testicular cancer patients in Denmark: a population-based study. *Int J Cancer* 66: 627–631.
18. Czene K, Lichtenstein P, Hemminki K (2002) Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer* 99: 260–266.
19. Shah MN, Devesa SS, Zhu K, McGlynn KA (2007) Trends in testicular germ cell tumours by ethnic group in the United States. *Int J Androl* 30: 206–13; discussion 213–4.
20. Chia VM, Quraishi SM, Devesa SS, Purdue MP, Cook MB, et al. (2010) International trends in the incidence of testicular cancer, 1973–2002. *Cancer Epidemiol Biomarkers Prev* 19: 1151–1159.
21. Pettersson A, Richiardi L, Nordenskjöld A, Kaijser M, Akre O (2007) Age at surgery for undescended testis and risk of testicular cancer. *N Engl J Med* 356: 1835–1841.
22. Trichopoulos D (1990) Hypothesis: does breast cancer originate in utero? *Lancet* 335: 939–940.
23. Ekblom A (1998) Growing evidence that several human cancers may originate in utero. *Seminars in Cancer Biology* 8: 237–244.
24. Rothman N, Wacholder S, Caporaso NE, Garcia-Closas M, Buetow K, et al. (2001) The use of common genetic polymorphisms to enhance the epidemiologic study of environmental carcinogens. *Biochim Biophys Acta* 1471: C1–10.
25. Thilly WG (2003) Have environmental mutagens caused oncomutations in people? *Nat Genet* 34: 255–259.
26. Myrup C, Schnack TH, Wohlfahrt J (2007) Correction of cryptorchidism and testicular cancer. *N Engl J Med* 357: 825–7; author reply 825–7.
27. Kanetsky PA, Mitra N, Vardhanabhuti S, Li M, Vaughn DJ, et al. (2009) Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer. *Nat Genet* 41: 811–815.
28. Rapley EA, Turnbull C, Olama AAA, Dermizakis ET, Linger R, et al. (2009) A genome-wide association study of testicular germ cell tumor. *Nat Genet* 41: 807–810.
29. Chanock S (2009) High marks for GWAS. *Nat Genet* 41: 765–766.
30. Mdzinarishvili T, Sherman S (2010) Weibull-like model of cancer development in aging. *Cancer Inform* 9: 179–188.
31. Frank SA (2007) *Dynamics of Cancer: Incidence, Inheritance, and Evolution*. Princeton: Princeton University Press. pp 378.
32. Frank SA (2004) Age-specific acceleration of cancer. *Curr Biol* 14: 242–246.
33. Harding C, Pompei F, Lee EE, Wilson R (2008) Cancer suppression at old age. *Cancer Res* 68: 4465–4478.
34. Ritter G, Wilson R, Pompei F, Burmistrov D (2003) The multistage model of cancer development: some implications. *Toxicol Ind Health* 19: 125–145.
35. Pompei F, Wilson R (2002) A quantitative model of cellular senescence influence on cancer and longevity. *Toxicol Ind Health* 18: 365–376.
36. Zhang X, Simon R (2005) Estimating the number of rate limiting genomic changes for human breast cancer. *Breast Cancer Res Treat* 91: 121–124.
37. Moger TA, Aalen OO, Halvorsen TO, Storm HH, Tretli S (2004) Frailty modelling of testicular cancer incidence using scandinavian data. *Biostatistics* 5: 1–14.
38. Arnheim N, Calabrese P (2009) Understanding what determines the frequency and pattern of human germline mutations. *Nat Rev Genet* 10: 478–488.
39. Loeb LA, Loeb KR, Anderson JP (2003) Multiple mutations and cancer. *Proc Natl Acad Sci U S A* 100: 776–781.
40. Moolgavkar SH, Luebeck EG (2003) Multistage carcinogenesis and the incidence of human cancer. *Genes Chromosomes Cancer* 38: 302–306.

41. Luebeck EG, Moolgavkar SH (2002) Multistage carcinogenesis and the incidence of colorectal cancer. *Proc Natl Acad Sci U S A* 99: 15095–15100.
42. Meza R, Jeon J, Moolgavkar SH, Luebeck EG (2008) Age-specific incidence of cancer: Phases, transitions, and biological implications. *Proc Natl Acad Sci U S A* 105: 16284–16289.
43. Looijenga LH, Oosterhuis JW (1999) Pathogenesis of testicular germ cell tumours. *Rev Reprod* 4: 90–100.
44. Clayton D, Schiffers E (1987) Models for temporal variation in cancer rates. II: Age-period-cohort models. *Stat Med* 6: 469–481.
45. Clayton D, Schiffers E (1987) Models for temporal variation in cancer rates. I: Age-period and age-cohort models. *Stat Med* 6: 449–467.
46. Thornburg KL, Shannon J, Thuillier P, Turker MS (2010) In utero life and epigenetic predisposition for disease. *Adv Genet* 71: 57–78.
47. Johnson N, Kemp A, Kotz S (2005) Univariate discrete distributions. Wiley series in probability and mathematical statistics. Applied probability and statistics. Wiley.
48. Calabrese P, Tavar S, Shibata D (2004) Pretumor progression: clonal evolution of human stem cell populations. *Am J Pathol* 164: 1337–1346.
49. National Cancer Institute. Surveillance, Epidemiology, and End Results (SEER) Program (2010) SEER*Stat database: Incidence - seer 17. .
50. National Cancer Institute. Surveillance Research Program. SEER*Stat software. version 6.6.2 [Computer software].
51. Brody JP (2009) Parallel routes of human carcinoma development: implications of the age-specific incidence data. *PLoS One* 4: e7053.