A model for random genetic damage directing selection of diploid or aneuploid tumours

P. S. Bazeley*, A. L. Nestor Kalinoski†, J. A. Ways†, S.-T. Liu‡, R. S. Ramdath†, S.-i. Matsui§ and D. C. Allison†

*Program in Bioinformatics and Proteomics/Genomics, †Department of Surgery and The University of Toledo Advanced Microscopy & Imaging Center, ‡Department of Biological Sciences, The University of Toledo, Toledo, OH, USA, and §Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, NY, USA

Received 5 November 2010; revision accepted 19 December 2010

Abstract

Objectives: To test whether genetic instability may determine whether tumours become aneuploid or diploid.

Materials and methods: We have identified genes needed for cell survival or replication by combining Affymetrix gene expression array data from 12 experimental cell lines with *in silico* GEO+GNF and expO databases. Specific loss of heterozygosis (LOHs), chromosomal abnormalities (called derivative chromosomes) and numbers of normal homologues were identified by SNP and SKY analyses. Random gene losses were calculated under the assumption that bi-allelic *MMR* gene inactivation causes a 20-fold increase in rate of gene loss.

Results: There were $\sim 1.23 \times 10^4$ genes widely dispersed throughout the genome and possibly expressed by all cells for survival or proliferation, many of these genes performed housekeeping functions. Conservation of the genes may explain the complete haploid genomes found for 15 different cell types and derivative chromosomes selectively retained in aneuploid cancer cell lines after LOH formations, and normal homologue losses. Loss of cell survival/replication genes was calculated to be higher in colon stem cells of carriers of *MMR* gene mutations than carriers of *APC* gene mutations.

Conclusion: Random loss of cell survival/replication genes was calculated to be low enough for colon stem cells with *APC* gene mutations to 'select' LOH and derivative chromosome combinations favouring

tumour cell proliferation. However, cell survival/replication gene loss was calculated to be too high for colonic stem cells lacking *MMR* genes to survive chromosomal instability, explaining why *MMR* mutations only produce tumours with diploid chromosome cells.

Introduction

It is not known why 90% of epithelial tumours are aneuploid and $\sim 10\%$ have diploid chromosomal compositions (1,2). Carriers of germline DNA mismatch-repair (*MMR*) gene mutations usually develop diploid cell colon cancers while carriers of adenomatous polypolyposis coli (*APC*) gene mutations preferentially form aneuploid colon cancers (3,4). This finding implies that determination of which tumours become aneuploid and which tumours remain diploid is under some sort of hitherto unknown genetic control mechanism. It is also not clear whether development of aneuploidy plays an important role in cell transformation and tumour progression. This uncertainty is because the identifiable chromosomal abnormalities, or derivative chromosomes (5), selectively retained in aneuploid cancers (i) vary widely among different tumours, even for tumours of the same type; and (ii) the DNA segments in highly selected derivative chromosomes and loss of heterozygosis (LOHs) found in aneuploid tumours often do not align with known oncogene or tumour suppressor gene networks (6,7).

Aneuploidy develops from several genetic changes: diploid (2C) tumour precursor cells usually double their chromosome numbers to become tetraploid (4C) before becoming aneuploid (8–10). This step usually requires inactivation of p53 and other surveillance networks to allow cells with damaged DNA to enter S-phase, leading to broken chromosomes (11,12). Aneuploid cells can also recombine broken chromosome fragments into segregating chromosomal abnormalities (13), and have attenuated

Correspondence: D. C. Allison, The University of Toledo, Health Science Campus, 3000 Arlington Ave., Toledo, OH 43614-5804, USA. Tel.: 419 383 3970; Fax: 419 383 6230; E-mail: david.allison@utoledo.edu

mitotic checkpoints allowing cells with abnormal chromosomes to complete mitosis (14).

Now, we have tested and expanded a hypothesis that aneuploidy is 'selected' by the majority of solid tumours to preserve genes and/or other DNA regions essential for tumour cell growth in derivative chromosomes and to eliminate selectively, genes and/or other DNA regions retarding cell proliferation in LOHs (15). In support of this model, we found a core of $\sim 1.23 \times 10^4$ genes expressed by all cells in a screen of 885 different cell types. These included 12 different cell lines, 31 lymphomas, 349 breast cancers, 290 colon cancers and 203 ovarian cancers. These essential core genes include $\sim 1.4 \times 10^3$ housekeeping genes and $\sim 1.09 \times 10^4$ differentially expressed genes, whose expression is related to tissue and cell type and also are influenced by activation of specific transformation pathways. We also found that 15 different cell types retained at least, haploid genomes, by single nucleotide polymorphism (SNP) analysis, including aneuploid cell lines, which had undergone innumerable deletions, gene losses and gene rearrangements. This finding is also consistent with the expression of $\sim 1.23 \times 10^4$ genes located throughout the genome being necessary for survival and replication of all cells.

Herein, we describe how we arrived at the core number of cell survival genes using Affymetrix Gene Arrays and functional classifications, in both tissue and cell-line samples. We also provide evidence that haploid genomes and essential cell survival genes are preserved in highly selected derivative chromosomes following the random chromosomal breaks, LOH formations and normal homologue losses, in aneuploid cells. Application of the finding that $\sim 1.23 \times 10^4$ genes must be expressed for cell survival or replication in colon stem cells provides a plausible explanation for why carriers of *MMR* mutations form diploid cell cancers and carriers of *APC* gene mutations form aneuploid cancers.

Materials and methods

Cell culture

Four prostate cancer cell lines (DU-145, PC-3, 22Rv1, LNCap), a benign prostate cell line (RWPE), three lung cancer cell lines (A549, H520, Calu-6), two benign lung fibroblast lines (CCD-34Lu, MRC-5), a pancreatic cancer cell line (SUIT-2) and a benign skin cell line (BUD-8) made up the 12 experimental cell lines used. Each of the cell lines was obtained from the American Type Culture Collection (<http://www.atcc.org>), except for SUIT-2, which had been previously characterized (15). All cell lines were expanded as monolayer cultures. MRC-5, Calu-6, CCD-34Lu, and BUD-8 lines were grown in

MEM (modified Eagle's medium), A549 line in DMEM, RWPE line in keratinocyte-serum free medium supplemented with human recombinant EGF (5 ng/ml) and bovine pituitary extract (0.05 g/ml), LNCap and 22Rv1 lines in RPMI 1640, and SUIT-2 line in McCoy's 5A medium. Human lymphocytes were grown in RPMI 1640 with phytohaemagglutinin (Gibco Invitrogen Corporation, Carlsbad, CA) for 72 h under an IRB-approved protocol. All media were supplemented with 10% FBS (Invitrogen Corporation, Carlsbad, CA, USA).

SKY and SNP analyses

Previously reported SNP and SKY results achieved from A549, LN-18, and SUIT-2 lines (15) were combined with SNP and SKY data from PC-3 and DU145 cell lines, for in-depth analysis of five aneuploid cancer cell lines. Examined by SKY, there were 24, 29, 30, 36 and 40 karyotypes for A549, LN-18, PC-3, DU-145 and SUIT-2 cell lines respectively, at the Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, NY. Genomic DNA was extracted and sent to the Gene Expression and Genotyping Facility, supported by the Comprehensive Cancer Center of Case Western Reserve University and University Hospitals of Cleveland (P30 CA43703), for SNP mapping.

RNA extraction from cell lines

Cells were lysed using TRIzol reagent (Invitrogen Corporation) in 75 cm² flasks. RNA extraction was performed exactly as noted for cell cultures grown as monolayers or in suspension. RNA of the colourless aqueous phase was precipitated, washed and dissolved in RNase-free water.

To ensure that the RNA was of sufficient quality and free of any DNA contamination, we performed an additional RNA 'cleanup' step using Qiagen RNeasy kit (Qiagen Inc., Valencia, CA, USA). DNase digestion was not required as RNeasy silica-membrane technology efficiently removes DNA without DNase treatment. The RNeasy protocol was followed exactly as directed and yield and purity of RNA were measured using a spectrophotometer. Only RNA that had a purity ratio (A_{260}/A_{280}) of 1.9–2.1 when measured in 10 mM Tris–HCl buffer, pH 7.5 was used for gene chip experiments. Integrity of the RNA was also checked by size fractionization electrophoresis and 18S and 28S ribosomal bands were sharp and intensity ratios were as expected.

Human genome U133 plus 2.0 gene expression array

Samples of total RNA were reverse transcribed into cDNA and synthesized to generate cRNA in the presence of biotinylated ribonucleotide tri-phosphates (rNTPs). Labelled

cRNA from each sample was individually hybridized on to the Affymetrix U133 2.0 gene chip. All protocols were followed precisely as prescribed by Affymetrix (<http://www.affymetrix.com>, Affymetrix, Santa Clara, CA, USA). The Affymetrix human genome U133 Plus 2.0 expression array was used to measure expression of 53 622 gene and expressed sequence tag (EST) probe sets. Probe sets with present ($P < 0.04$) or marginal ($0.04 < P < 0.06$) expression (Affymetrix MAS 5.0 algorithm) were classified as positive expressions and probe sets with absent expressions ($0.06 < P < 1.0$) were classified as negative expressions.

In silico expression databases

The GEO+GNF cell-line database consists of samples obtained from NCBI's GEO database as well as the GNF data set, which includes expression data of NCI60 cell lines. The GNF data set is found at the 'GNF Genome Informatics Applications & Data sets' website (<http://wombat.gnf.org/>). All samples were processed with Affymetrix human genome U133A arrays and the MAS 5.0 algorithm. The expO data set (<http://expo.intgen.org/geo/home.do>) had 2035 sets of expression measurements from frozen tumour samples processed with the human genome U133 Plus 2.0 array and the MAS 5.0 algorithm available between the 6/30/08 and 10/01/08 submissions.

Gene expression screen

A total of 16 537 probes representing 12 074 genes were measured in common among the 12 experimental cell lines and the *in silico* GEO+GNF and expO databases. There were 1918 probes expressed by all 210 cell lines in the experimental and the GEO+GNF databases and 1604 probes were not expressed by any of the 210 cell lines (data not shown). Expressions of the 1918 probes expressed by all 210 cell lines were tested in the 2035 malignant tissues catalogued in the expO database; 93% (1789/1918) of these probes, representing 1384 genes, were found to be also expressed in >99% (2015/2035) of the expO tissues. These genes expressed by all cell lines and in >99% of the tissues may have actually been expressed by all cell types, as a low assay error rate of 0.009 would produce 1% or 19 false-negative measurements in 2035 samples. Finally, 14% (219/1604) of probes not expressed by any of the 210 cell lines, specific for 183 genes, were expressed in 7–100% of tissues.

Gene-expression errors

MAS 5.0 present/absent calls are commonly used to filter out noisy probe sets (16) and were used to minimize false

positive/negative calls. The 1.6% slope on the right-hand side of the congruent *in vitro* gene-expression curve of the 12 experimental cell lines and a ~0.0165% slope of congruent expression of 9622 probe sets expressed by all 12 experimental cell lines in 1–1933 expO tissues (data not shown) are inconsistent with a high rate of false-negative measurements for the widely expressed genes, possibly reflecting housekeeping genes having an approximately 30-fold increase in expression levels compared to other genes (17).

Gene/chromosome group statistics

Differences in gene numbers between gene groups were tested by chi-square analysis using Fisher's exact two-tailed statistic. Chi-square testing is appropriate for this type of data as they are count data. Alpha value of 0.001 was chosen.

Results

Defining core cell survival and replication genes

It is assumed that all human cells, regardless of tissue origin, transformation, or ploidy status, must express a core of genes for cell survival and continuous replication. These essential genes include highly conserved housekeeping genes expressed by all cell types and variable sets of differentially expressed genes, initially selected by embryonic differentiation pathways. Expression of differentially expressed gene sets can also be influenced by environmental signals and activation of specific transformation pathways. It is assumed that colonic stem cells must also express a full complement of core-cell survival and proliferation genes to constantly repopulate the colonic epithelium.

Identification of core housekeeping genes

Affymetrix U-133a microarrays were used to compare expressions of 54 675 probe sets targeting 22 734 genes in 12 benign and malignant cell lines and the 2035 malignant tissues catalogued in the *in silico* expO database. We combined these expression results also with those of the 198 cell lines catalogued in the *in silico* GEO+GNF database, although there were only 16 537 probe sets, representing 12 074 genes, measured in common between the three databases. In a combined analysis of all three databases, there were $\sim 1.4 \times 10^3$ genes, located on all chromosomes except Y, expressed by all 210 cell lines and >99% of the 2035 malignant tissues (Fig. 1a, red). The expO database tissues included 290 colon cancers derived, one assumes, from colonic stem cells.

Figure 1. Classification of gene expression.

(a) Genetic map positions of $\sim 1.4 \times 10^3$ genes expressed by all 210 cell lines and $\geq 99\%$ of tissues (red bars). These genes are found in all chromosomes except Y. (b) Numbers of genes expressed by the 12 experimental cell lines, including housekeeping genes (red) and differentially expressed genes (green). (c) Average numbers of genes expressed by the 12 experimental cell lines ($\sim 1.23 \times 10^4 \pm 2.0\%$ [%CV]), 31 lymphomas ($\sim 1.21 \times 10^4 \pm 10.1\%$), 349 breast cancers ($\sim 1.30 \times 10^4 \pm 7.2\%$), 290 colon cancers ($\sim 1.29 \times 10^4 \pm 6.6\%$) and 203 ovarian cancers ($1.30 \times 10^4 \pm 9.3\%$). The three epithelial cancers expressed an average of ~ 707 more genes than the cell lines and ~ 877 more genes than the lymphomas (white). Overall, the 885 cell types expressed $\sim 1.4 \times 10^3$ housekeeping genes (red), $\sim 1.09 \times 10^4$ differentially expressed genes (green). Also, the ~ 800 additional genes expressed by the three epithelial cancers, may be involved with, or control, tissue-specific functions (white).

We then performed a functional classification to test whether most of these 1.4×10^3 widely expressed genes (Fig. 1a, red) were conserved for performing essential housekeeping functions (18–22). Classifications were made in a coded study of functions described in the NetAffx™ annotation database (<http://www.affymetrix.com/analysis/index.affx>) as related to the probe set ID. Genes were classified as being responsible for (i) intracellular functions involving metabolism, biosynthesis, cell structure, replication and similar functions; or (ii) differentiation and developmental functions such as immune response genes, hormones, receptors and more. We found that 74% (1021/1384) of genes expressed by all cell lines and $>99\%$ of the tissues, performed intracellular functions (Table S1a). For comparison, we also classified in a coded study, functions of 183 tissue-specific genes expressed $>7\%$ of tissues in the expO database, but not by any of the 210 cell lines. We found that only 31% (56/183) of these tissue-specific genes performed intracellular functions, a significantly lower proportion than the 74% of genes expressed by almost all cell types performing intracellular functions ($P < 0.001$, Table S1).

Thus, gene classes selected solely by expression patterns have differing functional profiles. It is possible

that this estimate of $\sim 1.4 \times 10^3$ housekeeping genes expressed by all cell types, including colonic stem cells (Fig. 1a) is low, both from false-negative measurements and that expression patterns of only 12 074 genes were measured in common among the three databases, to find these genes.

Identification of differentially expressed core genes

Expressions of the 22 734 genes measured in common between the 12 cell lines and the 2035 tissues of the expO database were compared. We found that an average of $\sim 1.23 \times 10^4$ genes (range: $1.19\text{--}1.26 \times 10^4$) were expressed by the 12 cell lines (Fig. 1b), which all include the previously identified $\sim 1.4 \times 10^3$ housekeeping genes (Fig. 1a, red) and $\sim 1.09 \times 10^4$ differentially expressed genes (Fig. 1b, green). As all of these cell lines have been maintained as *in vitro* cell cultures lacking tissue structures, the average of $\sim 1.23 \times 10^4$ genes expressed by per cell line (Fig. 1b, red + green) may represent a basic core number of genes needed for survival or proliferation (17,23,24).

We then determined average numbers of genes expressed by four classes of malignancy catalogued in the

expO database. There were $\sim 1.21 \times 10^4$, $\sim 1.31 \times 10^4$, $\sim 1.30 \times 10^4$ and $\sim 1.30 \times 10^4$ genes expressed on average, by 31 lymphomas, 349 breast cancers, 290 colon cancers and 203 ovarian cancers (Fig. 1c). Again, the 290 colon cancers were all derived from colonic stem cells. Also, the three epithelial cancers expressed, on average, ~ 707 more genes than the cell lines and ~ 877 more genes than the lymphomas (Fig. 1c, white). The additional genes expressed by the epithelial tumours may have been controlling or carrying out tissue-specific functions, activated by transformation or progression, or even expressed by non-transformed cells in the tumour tissues. However, it was not possible to distinguish these putative control genes (Fig. 1c, white) from differentially expressed genes (Fig. 1c, green) also expressed in the epithelial tumour tissues.

Taken together, these results are consistent with a core of $\sim 1.23 \times 10^4$ genes, including $\sim 1.4 \times 10^3$ housekeeping genes (Fig. 1a–c, red) and $\sim 1.09 \times 10^4$ tissue-specific, or differentially expressed, genes (Fig. 1b,c, green) being required for survival and replication of all cell types *in vitro* and *in vivo*. The core housekeeping genes, at least, are located on all chromosomes except Y (Fig. 1a, red).

Haploid genomes may preserve widely dispersed, core cell survival/replication genes

SNP-maps show genetic map positions of variable base pair markers. Heterozygous SNPs have different base pairs at a genetic map position, indicating presence of distinct parental DNAs (Fig. 2, red dots). Homozygous SNPs have the same base pair at an SNP site (Fig. 2, blue dots), either due to an LOH or both parental homologues having the same base pair. Although heterozygous (red dots) and

LOH (blue dots) may appear to overlap in the SNP maps (Fig. 2), each individual SNP has a unique genetic map position. Certain genomic regions, often in the regions of centromeres, do not have SNP markers and leave gaps in chromosomal SNP-maps that are not deletions (Fig. 2). SNP-maps can be classified as: (i) heterozygous with red (heterozygous) and blue (homozygous) SNPs along the entire chromosome length, indicating presence of DNA from both parental homologues; (ii) broken-chromosome LOHs resulting from loss of a large fragment(s) from one parental homologue, with the LOH region (blue dots only) found where the DNA from only one homologue remains; and (iii) whole-chromosome LOHs, or only a haploid genome of one parental homologue remaining (blue dots only), following complete loss of the other parental homologue (Fig. 2) (15). The occasional heterozygous SNPs (red dots) found in whole-chromosome SNP-maps may be due to assay errors or to represent small, translocated DNA pieces from opposing parental homologues. Somatic chromosomes of the DU-145 prostate cancer cell line had 10 heterozygous, 10 broken-chromosome LOH and two whole-chromosome LOH SNP-maps (Fig. 2a), and the PC-3 prostate cancer cell line had 12 heterozygous, five broken-chromosome LOH and five whole-chromosome LOH SNP-maps (Fig. 2b).

We have previously found by SNP analysis that at least four cancer cell lines had haploid genomes (15). We reasoned that haploid genomes may be needed to preserve widely distributed, cell survival/replication genes (Fig. 1, red, green). We therefore extended this analysis to a total of 15 cell types, including non-transformed, short-term lymphocyte cultures from two normal individuals, four benign diploid cell lines, two near-diploid cancer cell lines and seven aneuploid cancer cell lines. We found that at

Figure 2. SNP-mapping of LOHs. (a) DU-145 cell chromosomes had 10 heterozygous, 10 broken-chromosome LOH and two whole-chromosome LOH SNP-maps. (b) PC-3 cell chromosomes had 12 heterozygous, five broken-chromosome LOH and five whole-chromosome LOH SNP-maps. See Table S2 for more detail.

least one complete haploid genome was present per somatic chromosome of each of these cell types (data not shown), regardless of the cell type's ploidy level (diploid or aneuploid), its capability to survive in long-term culture, or its transformation status (benign or malignant). Thus, retention of complete haploid genomes without large homologous deletions by all cell types seems independent of transformation status, ability to survive in long-term tissue culture, or ploidy level. This retention of complete haploid genomes by all replicating cell types can be explained by the need to retain functional alleles of $\sim 1.23 \times 10^4$ essential genes widely distributed throughout the genome (Fig. 1, red, green), as first predicted by Theodore Boveri in 1909 (25).

Highly selected chromosomal abnormalities and the preservation of haploid genomes

Identifiable chromosomal abnormalities found in >60% of karyotypes are called highly selected derivative chromosomes (5). The aneuploid DU-145 cell line had 16 highly selected derivative chromosomes, consisting of 14 translocations and two deletions, and the aneuploid PC-3 cell line had 10 highly selected derivative chromosomes, including nine translocations and one deletion (Table 1). Five PC-3 and three SUIT-2 chromosomes (15) did not have any normal homologues, but at least complete haploid genomes (Table S2): 53% (1036/1951) of DNA pieces from the highly selected, derivative chromosomes identified in the SUIT-2 and PC-3 karyotypes were from the eight chromosomes without normal homologues, compared with only 29% (73/250) of DNA pieces from the same eight chromosomes being present in derivative chromosomes with only single copies in the karyotypes ($P < 0.001$, Table 2). Thus, derivative chromosomes are specifically selected to

maintain the DNA of chromosomes lacking normal homologues, and presumably preserve genes and possibly other genetic elements essential for cell survival and replication (Fig. 1, red, green).

Derivative chromosomes may preserve core cell survival/replication genes during LOH formation

Loss of large chromosomal fragments was relatively rare in diploid cells compared to aneuploid cells. Only 11% (20/176) of the somatic chromosomes in six diploid and two near-diploid cell lines had large, broken-chromosome (18 chromosomes) or whole-chromosome LOH SNP-maps (2 chromosomes). However, 55% (85/154) of the chromosomes in seven aneuploid cancer cell lines had large LOHs, from loss of chromosome fragments or loss of whole homologues, including 55 broken-chromosome LOH and 30 whole-chromosome LOH SNP-maps ($P < 0.001$, data not shown). DNA producing genetic heterozygosity in 92% (33/36) of broken-chromosome LOH SNP-maps in the five aneuploid cancer cell lines upon which SKY was performed also originated from highly selected derivative chromosomes (Table S2b), possibly to preserve essential cell survival and replication genes (Fig. 1, red, green) (15).

Aneuploid A549, DU-145, LN-18, PC-3 and SUIT-2 cancer cell lines have many large LOHs (Table S2b,c), which have most likely, been derived from tetraploid precursor cells (15). This conclusion is drawn from the finding in combined SNP and SKY analyses that 62% (68/110) of somatic chromosomes in these aneuploid cell lines had additional normal homologues above the normal diploid number (Table S2). These extra normal homologues are more easily explained by an early tetraploidization of each of these tumours than by selective

Table 1. Karyotypes of DU-145 and PC-3 prostate cancer lines

DU-145	61, X,-Y, t(1;4), t(2;13), t(3;14), -4, t(4;10), t(5;21), del(5)×3, t(6;16), +7, t(9;11), del(9), t(10;19), -11, t(11;12), t(11;21), -13×2, t(13;19), t(14;18)×2, -15, t(15;20)×2, +17, -20, t(Y;20), +21
PC-3	52, X,-Y, -2×2, t(2;8)×2, t(2;15;17), -3×2, t(3;10)×2, t(3;10;3;10;3)×2, -4×2, t(4;6), t(4;10)×2, t(4;12), -5, del(5), +7, -8, t(8;12)×2, t(8;12;1;10), -9, -10×2, +11, -12×2, +14, -15, -17, +18, +20×2, +21×2

Table 2. Evidence that derivative chromosomes are selected to maintain haploid genomes for chromosomes without any normal homologues, by SKY analysis

	Eight chromosomes without any normal homologues by SKY analysis	Thirty-six chromosomes with normal homologues by SKY analysis	<i>P</i> -value
% (n) of DNA pieces in highly selected derivative chromosomes	53 (1036/1951)	47 (915/1951)	<0.001
% (n) of DNA pieces in unselected, single copy derivative chromosomes	29 (73/250)	71 (177/250)	

retention of many normal homologues by non-disjunctions. Specifically, it would take 21, 11, 15, 10 and 29 normal chromosomes selectively retained following 86, total, non-disjunctions to produce the extra normal homologues found in these cell lines without an early tetraploidization step having occurred (Table S2).

For example, DU-145 chromosome 16 has two normal homologues, a highly selected, t(6;16) derivative chromosome, and a broken-chromosome LOH SNP-map (Fig. 3a,b). The haploid LOH region indicates that the two homologues appearing to be normal are identical and originate from the same parent (Fig. 3a,b, blue dots). Thus, although DU-145 chromosome 16 has a diploid number of normal homologues by SKY analysis (Fig. 3a, Table 1), SNP analysis reveals an extra copy of one parental homologue (Fig. 3b). Also, mutations or other small-scale losses, of cell survival/replication genes in diploid precursor cells of this aneuploid cell tumour are still present in both of the normal appearing homologues following transition to tetraploidy (Fig. 3c, black X). As the two intact homologues are from the same parent, the chromosome 16 fragment of the t(6;16) derivative chromosome must be from the other parent to produce the region of SNP heterozygosity (Fig. 3b, red dots). The t(6;16) derivative chromosome (Fig. 3a) may have been selected to preserve functional alleles of housekeeping genes (Fig. 3c, red bars), or other differentially expressed cell survival/replication genes (not shown), which complement allelic losses occurring in diploid precursor cells (Fig. 3c).

Random loss of core cell survival/replication genes in colon stem cells with germline APC or MMR mutations

The estimate that all cells must express $\sim 1.23 \times 10^4$ core genes for survival or division (Fig. 1c, red, green) can be

applied to stem cells of the colon. This allows calculation of the cumulative random loss of these core genes during colon stem cell proliferation. Colon stem cells of *APC* gene mutation-carriers, have normal DNA repair genes and are therefore assumed to have normal rates of random gene loss. Also, colon stem cells of carriers of mutations of a single allele of the *MMR* gene would not have an increased gene loss rate until the second, normal allele of this gene was also inactivated. If normal cells have a 1×10^{-6} spontaneous rate of gene loss per cell division (26,27), then a 15-year-old normal colon stem cell, or 15-year-old colon stem cells of carriers of either germline *APC* or *MMR* gene mutations, would have lost ~ 7 cell-survival/replication genes during growth from a one-cell embryo (Tables 3a and 4a,b). Similarly, a 65-year-old colon stem cell with intact DNA repair genes would have lost ~ 29 cell-survival/replication genes (Tables 3b and 4a–c). However, gene loss rates may increase when both alleles of an *MMR* gene are functionally inactivated (3,28–30). If bi-allelic loss of an *MMR* gene leads to a 20-fold increase in gene loss rate, then a colon stem cell which lost both alleles of an *MMR* gene at 15 years of age would lose ~ 290 cell survival/replication genes by 47 years of age (Tables 3c and 4a,b,d).

Random loss of normal APC and MMR alleles in the colon stem cells of 15-year-old carriers of APC and MMR gene mutations

There are $\sim 5 \times 10^7$ stem cells in a fully developed colon (3,31), which may divide ~ 36 times a year (32). Thus, a 15-year-old colon stem cell would have undergone ~ 569 lifetime divisions after its origin from a one-cell embryo (Table 4a,b). If colon stem cells with intact DNA repair genes also have a $\sim 1 \times 10^{-6}$ rate of spontaneous gene

Figure 3. Normal homologues and derivative chromosomes producing a broken-chromosome LOH SNP-map. Two normal appearing homologues of chromosome 16 (NL) and a t(6;16) highly selected derivative chromosome (a) produce a broken-chromosome LOH SNP-map (b) for DU-145 chromosome 16. The two normal homologues (a) are from the same parent and generated by a diploid to tetraploid transition; this can be deduced from presence of a completely haploid SNP-map in the LOH region of this broken-chromosome LOH (b, blue dots). Accordingly, chromosome 16 DNA in the t(6;16) derivative chromosome (a) must originate from the opposite parental homologue to produce the region of SNP heterozygosity (b, red dots). The piece of DU-145 cell DNA in this derivative chromosome (a) carries many housekeeping genes shown in Fig. 1a (C, red bars). This DNA fragment most likely also contains many differentially expressed genes (Fig. 1c, green) also necessary for cell replication and survival (not shown). Therefore, this derivative chromosome may have been selected for gene(s) (c, red bars) complementing cell survival/replication genes lost in duplicated, normal homologues (c, X).

Table 3. Random gene loss in diploid colon stem cells

(a) Loss of cell-survival/replication genes in a 15-year-old colon stem cell with normal DNA repair genes	
Embryonic gene loss	~0.36
Birth to 15 year gene loss	~6.64
Total gene loss	~7
(b) Loss of cell survival/replication genes in a 65-year-old colon stem cell with normal DNA repair genes	
0–15 year gene loss	~7
15–65 year gene loss	~22
Total gene loss	~29
(c) Loss of cell survival/replication genes in a 47-year-old colonic stem cell in which both alleles of a DNA mismatch-repair gene were lost at 15 years of age ^a	
0–15 year gene loss	~7
15–47 year gene loss	~283
Total gene loss	~290
(d) Number of colon stem cells in 15-year-old carriers of <i>APC</i> or DNA mismatch-repair mutations in which normal alleles of these genes are also lost	
Total stem cell no.	5×10^7
No. divisions	~569
Gene loss rate	10^{-6} /cell division
No. stem cells which have lost both alleles	2.84×10^4 ($5 \times 10^7 \times 569 \times 10^{-6} = 2.84 \times 10^4$)

^aThe loss of both DNA mismatch-repair alleles is assumed to raise the random gene loss rate from 10^{-6} to 2×10^{-5} per cell division. See Table 4 for more detail.

loss per cell division, the colons of 15-year-old carriers of either germline *MMR* or *APC* mutations would have $\sim 2.84 \times 10^4$ stem cells in which both alleles of these genes would be lost (Tables 3d and 4a,b). Thus, in the adolescent colons of germline carriers of a single *APC* gene mutation, there are sufficient colon stem cells with both *APC* alleles mutationally inactivated or otherwise lost, to expand clonally and produce thousands of adenomatous polyps.

Discussion

The DNA of solid cancer cells typically have many mutations (33,34) and current estimates of mutation numbers in solid cancer cells may still be low, as detection of all mutations in heterogeneous cell populations is very difficult (35). Although most mutations in solid tumours may not have direct effects on tumour growth (6,36), many random mutations and gene losses must occur in the $\sim 1.23 \times 10^4$ essential genes for cell survival or replication (Fig. 1, red, green). Also, many aneuploid cancers have somatic copy number alterations of large, and even whole chromosome-arm length, DNA regions (34,37). Surprisingly, many of these large genomic regions with low gene densities do not contain tumour suppressor

Table 4. Random loss of cell-survival/growth genes for different colonic stem cells

(a) Loss of cell survival/replication genes <i>in utero</i>	
No. cell divisions	29
Gene loss rate/division ^a	10^{-6}
No. cell survival/replication genes	1.23×10^4 (Fig. 1d, red, green)
No. cell survival/replication genes lost (per cell) <i>in utero</i>	~ 0.36 ($29 \times (10^{-6}) \times (1.23 \times 10^4) = \sim 0.36$)
(b) Loss of cell survival/replication genes from birth to 15 years of age	
No. cell divisions per year	36^b
Total no. cell divisions	540 (36×15)
Gene loss rate/division	10^{-6}
No. cell survival/replication genes	1.23×10^4
No. cell survival/replication genes lost per cell in years 1–15	~ 6.64 ($540 \times (10^{-6}) \times (1.23 \times 10^4) = \sim 6.64$)
(c) Loss of cell survival/replication genes from 15 to 65 years of age	
No. cell divisions per year	36
Total no. cell divisions	1800 (50×36)
Gene loss rate/division	10^{-6}
No. cell survival/replication genes	1.23×10^4
No. cell survival/replication genes lost per cell in years 15–65	~ 22 ($1800 \times (10^{-6}) \times (1.23 \times 10^4) = \sim 22.14$)
(d) Loss of cell survival/replication genes from 15 to 47 years of age, following bi-allelic loss of a DNA mismatch-repair gene at 15 years of age	
No. cell divisions per year	36
Total no. cell divisions	1,152 (32×36)
Gene loss rate/division ^c	2×10^{-5}
No. cell survival/replication genes	1.23×10^4
No. cell survival/replication genes lost per cell in years 15–47	~ 283 ($1,152 \times (2 \times 10^{-5}) \times (1.23 \times 10^4) = \sim 283.39$)

^aEmbryonic loss of cell survival/replication genes at an average gene loss rate of 10^{-6} per cell division (26,27).

^bNumber of cell divisions per year (32).

^cCell survival/replication gene losses from 15 to 47 years at a loss rate of 2×10^{-5} due to loss of both alleles of a DNA mismatch-repair gene at 15 years of age (28,29,34).

genes, yet have complex genes of unknown functions (38).

Rates of random, inactivating mutations and other types of small scale gene losses vary for different cell types (26,27), possibly being $<1 \times 10^{-7}$ per cell division for embryonic stem cells, but increasing with differentiation (39). Cancer cells have higher gene loss rates from accumulated damage to DNA repair genes (27). Our analysis utilizes a hypothetical gene loss rate of 1×10^{-6} per cell division for colon stem cells, based on experimental gene loss rates found for fibroblasts and lymphoblastoid cell lines (26,27).

Ability to eliminate tumour suppressor genes by forming LOHs may drive the selection of aneuploid colon cancers secondary to *APC* gene mutations (3). Colon cancers from mutations and losses of the *APC* gene develop at 15–20 years of age in carriers of germline mutations of one *APC* allele. Also, spontaneous *APC* gene-dependent colon cancers often develop in the seventh decade for people who have inherited two normal *APC* alleles. In such a case, both *APC* alleles are usually mutationally inactivated, or otherwise lost, in precancerous polyps and in malignant colon cancers found for both types of *APC* gene-dependent tumours (3,4).

Thousands of precancerous polyps appear in the colons of *APC* gene mutation adolescent carriers. Cells of each of these polyps has the mutated *APC* allele, inherited from one parent, but also has different mutations or other

types of loss of the normal *APC* allele inherited from the other parent. Application of gene loss rate of 1×10^{-6} per cell division leads to presence of $\sim 2.4 \times 10^4$ colon stem cells with inactivating mutations or losses of both *APC* alleles in colons of *APC* gene mutation adolescent carriers (Table 3d), explaining the occurrence of thousands of precancerous polyps in these young individuals. Also, the *APC* gene is part of the mitotic spindle assembly checkpoint (40). Thus, bi-allelic mutations or losses of the *APC* gene may facilitate the diploid to tetraploid transition preceding aneuploidy (8–10), as colon stem cells with bi-allelic *APC* gene mutations and/or losses may not be able to complete mitosis and therefore must either die or become cycling tetraploid cells (14,41,42). In contrast, loss of remaining normal *MMR* genes in *MMR* mutation carriers only leads to genetic instability and does not directly produce precancerous polyps or tetraploid cells.

Colon stem cells producing *APC* gene-dependent tumours have normal DNA repair genes. Thus, the random loss of cell survival/replication genes (Fig. 1, red, green) is minimal at normal 1×10^{-6} rate of gene loss in colon stem cell precursors of *APC* gene-dependent tumours, even for colon cancers developing in the seventh decade (Table 3b). For example, it can be estimated that three cell survival/replication gene alleles are lost by chance for chromosome 1 of a diploid, 65-year-old person's colon stem cell at a 1×10^{-6} gene loss rate per cell division (Fig. 4a, black lines, Table 2a,b). These lost

Figure 4. Cell-survival/replication genes and LOH formation. (a) Hypothetical chromosome 1 homologues of a 65-year-old colon stem cell with loss of three cell survival/replication gene alleles (black lines), complemented by normal alleles on the other parental homologue (red lines). (b) A recessive, inactivating mutation of a tumour suppressor gene (yellow, cross hatched) complemented by a functional allele on the other homologue (yellow). (c) A cancer cell derived from this stem cell (b) can select a chromosome break eliminating this tumour suppressor gene (yellow) in a LOH and preserve essential cell survival/replication genes (red lines) in a derivative chromosome. (d) Hypothetical chromosome 1 homologues in a 47-year-old colon stem cell which has lost 32 cell survival/replication gene alleles (black lines) secondary to loss of both alleles of an *MMR* gene at 15 years of age. Inactivated cell survival/replication genes are complemented by functional alleles on the other homologue (red lines). (e) A recessive, inactivating mutation of a tumour suppressor gene (yellow, cross hatched) complemented by a functional allele on the other homologue (yellow). (f) A cancer cell derived from this stem cell (e) cannot select a chromosomal break for LOH elimination of the functional tumour suppressor allele (yellow) without completely eliminating many of the only functional alleles, which complement defective cell survival/replication genes (*). Number of lost cell survival and replication genes was calculated by multiplying the proportion of total chromosome 1 genes (0.111 , or $2.52 \times 10^3 / 2.2734 \times 10^4$) $\times$ total number of cell survival/replication genes lost in the stem cells (Table 3b,c).

genes are complemented by normal alleles on the other homologue (Fig. 4a, red lines). Figure 4b shows an inactivating mutation of a tumour suppressor gene on one homologue of chromosome 1 of this stem cell (yellow, cross hatched), which is complemented by a normal allele of the tumour suppressor gene on the other homologue (yellow). However, a chromosomal break can eliminate the functional tumour suppressor allele in a non-segregating chromosomal fragment while maintaining functional cell survival/replication gene alleles in a segregating derivative chromosome (Fig. 4c, red lines). As loss of crucial cell survival and division genes is relatively rare in *APC* gene-dependent tumours, the negative effects from loss of large chromosomal fragments usually can be corrected by selection of appropriate LOH and derivative-chromosome combinations, which preserve essential cell survival/replication genes and eliminate genes that result in slowing tumour cell division rates (Fig. 4c).

However, successful formation of large LOHs may still be favoured for tetraploid cells when compared even to diploid cells with intact DNA repair genes. For example, loss of certain fragments of either of diploid chromosomes shown in Fig. 4a may lead to immediate cell death from loss of alleles complementing the three defective cell survival/replication genes (Fig. 4a, red lines). LOHs formed from tetraploid precursor cells however, may be the result of many selective steps following random, and often deleterious, losses of chromosome fragments and entire homologues. However, immediate effects of such DNA fragment losses in tetraploid cells are buffered by remaining intact copies of the normal chromosomes from which DNA losses have occurred. Thus, large LOH formation in tetraploid cells is a multi-step process, requiring at least two chromosome breaks and/or mitotic segregation errors per LOH. This multi-step process may allow tetraploid cells to gradually select derivative chromosome and LOH combinations favouring cell proliferation with minimal negative effects on overall expansion of the tumour cell population, explaining why large LOHs are common in aneuploid cells when compared to diploid cells (data not shown) (30,43).

In contrast to *APC* gene-dependent colon cancers, *MMR* gene mutation carriers usually only develop diploid colon tumours. *MMR* gene-dependent colon malignancies usually appear in the fifth decade of human age and both *MMR* alleles are also lost in these neoplasms (3,4). However, both *MMR* alleles are lost or functionally inactivated in many colon stem cells of *MMR* mutation carriers by 15 years of age at a normal 1×10^{-6} gene loss rate (Table 3d). Loss of both alleles of an *MMR* gene leads to frame-shift mutations in genes with embedded microsatellites (3,30), and may increase the rate of single nucleotide changes and silencing of other genes (28,29,34,44). If loss

of both alleles of *MMR* leads to overall 20-fold increase in random gene loss, colon stem cells which have lost both alleles of *MMR* at 15 years of age will have a significant loss of cell survival/replication genes by the fifth decade (Table 3c).

This increased genetic damage to cell survival/replication genes of the stem cell precursors of colon cancers produced by *MMR* mutations may prevent LOH formation by random chromosome breaks and loss of large chromosome fragments. Figure 4d shows hypothetical chromosome 1 homologues of a 47-year-old colon stem cell, which had lost both alleles of an *MMR* gene when the carrier was 15 years of age. Genetic instability from complete loss of *MMR* has produced a complementation ladder between the two haploid genomes from loss or functional inactivation of 32 cell survival and replication genes (Fig. 4d, alternating red and black lines). Figure 4e shows one of these homologues with an inactivating tumour suppressor gene mutation (Fig. 4e, yellow, cross hatched). However, elimination of the functional tumour suppressor allele on the other homologue (Fig. 4e, yellow), by any chromosome break producing a large LOH, leads to cell death from loss of complementing cell survival/replication gene alleles (Fig. 4f, *). Thus, cancers developing in *MMR* mutation carriers cannot eliminate tumour suppressor genes in LOHs formed by loss of large chromosome fragments, or complete homologues, and still remain viable (Fig. 4d–f), explaining why only diploid colon cancers are produced by carriers of *MMR* mutations.

Epithelial tumour cells undergo a large number of cell divisions following transformation, compared to other tumour types (45) and therefore have an ongoing random loss of cell survival/replication genes from DNA replication and segregation errors (27). In a striking example, two aneuploid cancers maintained by serial transplantation in wolves/dogs and Tasmanian devils have recently been reported (46). Both tumours are estimated to be hundreds, and possibly even thousands of years of age, during which countless alleles of cell survival/replication genes must have been randomly lost.

Random production of breaks in chromosomes with fixed gene linkages, combined with selection of LOHs and derivative chromosomes alone, could not compensate for ongoing genetic damage occurring in immortalized aneuploid cells (Fig. 4f). However, additional acquisition of genetic capability to produce random, intra-chromosomal gene rearrangements (47,48) by transposable elements, the bridge-breakage-fusion cycle, and/or small somatic copy number alterations (12,38), would allow aneuploid cells to selectively retain genes needed for cell survival or replication in derivative chromosomes and to eliminate harmful, or useless, DNA regions in LOHs (6).

Two normal alleles of differentially expressed genes are often required to produce the optimal amount of protein for overall function in multi-protein pathways (49). Abnormal gene copy numbers from aneuploidy do not meet this requirement and therefore may slow proliferation of aneuploid cells. Mutations in protein degradation and folding pathways have recently been found to be selected to correct for protein stoichiometry imbalances from abnormal gene numbers in aneuploid yeast strains (50). It is possible that preferential selection of favourable derivative chromosome and LOH combinations may also be another mechanism to optimize gene dosage in aneuploid cancer cells.

In summary, we have found that at least complete haploid genomes without large homologous deletions are selectively preserved in somatic and X chromosomes in all cell types tested. These complete haploid genomes presumably retain functional copies of widely dispersed genes and/or other DNA sequences needed for survival and proliferation of all living cells, including $\sim 1.4 \times 10^3$ housekeeping genes (Fig. 1, red) and $\sim 1.09 \times 10^4$ differentially expressed genes with variable expression patterns, among different tissue types (Fig. 1, green). The need to preserve functional alleles of these essential genes and possibly other DNA regions following random mutations, or other causes of gene and DNA loss, explains the selective retention of most derivative chromosomes in aneuploid cancer cells (Fig. 3). Also, random loss of essential cell survival and replication genes or essential DNA regions may be too great in colon stem cells with *MMR* mutations (Fig. 4d) to survive chromosome instability of aneuploidy (Fig. 4e). However, random genetic damage to these essential genes and other crucial DNA regions may be minimal in colon stem cells with *APC* mutations (Fig. 4a), allowing selection of viable aneuploid cancer cells in *APC* gene-dependent colon cancers (Fig. 4c).

Acknowledgements

This study was supported by a Translational Research Stimulation Award from the University of Toledo College of Medicine and the Cancer Biology Research Fund of the University of Toledo Foundation. We thank Professor Bernard Strauss, University of Chicago, for his critical review and Kay Langenderfer for her help in preparing this manuscript.

References

- Albertson DG, Collins C, McCormick F, Gray JW (2003) Chromosome aberrations in solid tumors. *Nat. Genet.* **34**, 369–376.
- Torres EM, Williams BR, Amon A (2008) Aneuploidy: cells losing their balance. *Genetics* **179**, 737–746.
- Komarova NL, Lengauer C, Vogelstein B, Nowak MA (2002) Dynamics of genetic instability in sporadic and familial colorectal cancer. *Cancer Biol. Ther.* **1**, 685–692.
- Lengauer C, Kinzler KW, Vogelstein B (1998) Genetic instabilities in human cancers. *Nature* **396**, 643–649.
- Brothman AR, Persons DL, Shaffer LG (2009) Nomenclature evolution: changes in the ISCN from the 2005 to the 2009 Edition. *Cytogenet. Genome Res.* **127**, 1–4.
- Bignell GR, Greenman CD, Davies H, Butler AP, Edkins S, Andrews JM *et al.* (2010) Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893–898.
- Johansson B, Mertens F, Mitelman F (1996) Primary vs. secondary neoplasia-associated chromosomal abnormalities – balanced rearrangements vs. genomic imbalances? *Genes Chromosomes Cancer* **16**, 155–163.
- Ganem NJ, Storchova Z, Pellman D (2007) Tetraploidy, aneuploidy and cancer. *Curr. Opin. Genet. Dev.* **17**, 157–162.
- Holland AJ, Cleveland DW (2009) Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nat. Rev. Mol. Cell Biol.* **10**, 478–487.
- Storchova Z, Kuffer C (2008) The consequences of tetraploidy and aneuploidy. *J. Cell Sci.* **121**, 3859–3866.
- Bartek J, Bartkova J, Lukas J (2007) DNA damage signalling guards against activated oncogenes and tumour progression. *Oncogene* **26**, 7773–7779.
- Paulovich AG, Toczyski DP, Hartwell LH (1997) When checkpoints fail. *Cell* **88**, 315–321.
- Merel P, Prieur A, Pfeiffer P, Delattre O (2002) Absence of major defects in non-homologous DNA end joining in human breast cancer cell lines. *Oncogene* **21**, 5654–5659.
- Cahill DP, Lengauer C, Yu J, Riggins GJ, Willson JK, Markowitz SD *et al.* (1998) Mutations of mitotic checkpoint genes in human cancers. *Nature* **392**, 300–303.
- Nestor AL, Hollopeter SL, Matsui SI, Allison D (2007) A model for genetic complementation controlling the chromosomal abnormalities and loss of heterozygosity formation in cancer. *Cytogenet. Genome Res.* **116**, 235–247.
- Seo J, Bakay M, Chen Y-W, Hilmer S, Shneiderman B, Hoffman EP (2004) Interactively optimizing signal-to-noise ratios in expression profiling: project-specific algorithm selection and detection p-value weighting in Affymetrix microarrays. *Bioinformatics* **20**, 2534–2544.
- Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA* **101**, 6062–6067.
- Dezso Z, Nikolsky Y, Sviridov E, Shi W, Serebriyskaya T, Dosymbekov D *et al.* (2008) A comprehensive functional analysis of tissue specificity of human gene expression. *BMC Biol.* **6**, 49.
- Eisenberg E, Levanon EY (2003) Human housekeeping genes are compact. *Trends Genet.* **19**, 362–365.
- Hsiao L-L, Dangond F, Yoshida T, Hong R, Jensen RV, Misra J *et al.* (2001) A compendium of gene expression in normal human tissues. *Physiol. Genomics* **7**, 97–104.
- Tu Z, Wang L, Xu M, Zhou X, Chen T, Sun F (2006) Further understanding human disease genes by comparing with housekeeping genes and other genes. *BMC Genomics* **7**, 31.
- Warrington JA, Nair A, Mahadevappa M, Tsyganskaya M (2000) Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol. Genomics* **2**, 143–147.
- Ertel A, Verghese A, Byers S, Ochs M, Tozeren A (2006) Pathway-specific differences between tumor cell lines and normal and tumor tissue cells. *Mol. Cancer* **5**, 55.

- 24 van Staveren WCG, Solís DYW, Hébrant A, Detours V, Dumont JE, Maenhaut C (2009) Human cancer cell lines: experimental models for cancer cells in situ? For cancer stem cells? *Biochim. Biophys. Acta* **1795**, 92–103.
- 25 Baltzer F (1967) *The Chromosome Theory of Inheritance, Experiments with Doubly Fertilized Sea Urchin Eggs (Dispermy Experiments)*. Theodor Boveri, *Life and Work of a Great Biologist, 1862–1915*. Berkeley and Los Angeles: University of California Press, pp. 85–97.
- 26 Araten DJ, Golde DW, Zhang RH, Thaler HT, Gargiulo L, Notaro R *et al.* (2005) A quantitative measurement of the human somatic mutation rate. *Cancer Res.* **65**, 8111–8117.
- 27 Salk JJ, Fox EJ, Loeb LA (2010) Mutational heterogeneity in human cancers: origin and consequences. *Annu. Rev. Pathol. Mech. Dis.* **5**, 51–75.
- 28 Abdel-Rahman WM (2008) Genomic instability and carcinogenesis: an update. *Curr. Genomics* **9**, 535–541.
- 29 Eshleman JR, Markowitz SD, Donover PS, Lang EZ, Lutterbaugh JD, Li GM *et al.* (1996) Diverse hypermutability of multiple expressed sequence motifs present in a cancer with microsatellite instability. *Oncogene* **12**, 1425–1432.
- 30 Toft NJ, Arends MJ (1998) DNA mismatch repair and colorectal cancer. *J. Pathol.* **185**, 123–129.
- 31 Yatabe Y, Tavaré S, Shibata D (2001) Investigating stem cells in human colon by using methylation patterns. *Proc. Natl. Acad. Sci. USA* **98**, 10839–10844.
- 32 Bach SP, Renahan AG, Potten CS (2000) Stem cells: the intestinal stem cell as a paradigm. *Carcinogenesis* **21**, 469–476.
- 33 Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD *et al.* (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196.
- 34 Stratton MR, Campbell PJ, Futreal PA (2009) The cancer genome. *Nature* **458**, 719–724.
- 35 Tomlinson I, Sasieni P, Bodmer W (2002) How many mutations in a cancer? *Am. J. Pathol.* **160**, 755–758.
- 36 Greenman C, Stephens P, Smith R, Dalgleish GL, Hunter C, Bignell G *et al.* (2007) Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158.
- 37 Baudis M (2007) Genomic imbalances in 5918 malignant epithelial tumors: an explorative meta-analysis of chromosomal CGH data. *BMC Cancer* **7**, 226.
- 38 Beroukhi R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J *et al.* (2010) The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905.
- 39 Cervantes RB, Stringer JR, Shao C, Tischfield JA, Stambrook PJ (2002) Embryonic stem cells and somatic cells differ in mutation frequency and type. *Proc. Natl. Acad. Sci. USA* **99**, 3586–3590.
- 40 Rusan NM, Peifer M (2008) Original CIN: reviewing roles for APC in chromosome instability. *J. Cell Biol.* **181**, 719–726.
- 41 Fujiwara T, Bandi M, Nitta M, Ivanova EV, Bronson RT, Pellman D (2005) Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* **437**, 1043–1047.
- 42 Rao CV, Yamada HY, Yao Y, Dai W (2009) Enhanced genomic instabilities caused by deregulated microtubule dynamics and chromosome segregation: a perspective from genetic studies in mice. *Carcinogenesis* **30**, 1469–1474.
- 43 Sugai T, Habano W, Nakamura S-I, Sato H, Uesugi N, Takahashi H *et al.* (2000) Genetic alterations in DNA diploid, aneuploid and multiploid colorectal carcinomas identified by the crypt isolation technique. *Int. J. Cancer* **88**, 614–619.
- 44 Peltomäki P (2001) Deficient DNA mismatch repair: a common etiologic factor for colon cancer. *Hum. Mol. Genet.* **10**, 735–740.
- 45 Steel GG (1977) *Growth Kinetics of Tumours: Cell Population Kinetics in Relation to the Growth and Treatment of Cancer*. Oxford: Clarendon Press, p. 191.
- 46 Murchison EP (2009) Clonally transmissible cancers in dogs and Tasmanian devils. *Oncogene* **27**(S2), S19–S30.
- 47 Anderson GR, Brenner BM, Swede H, Chen N, Henry WM, Conroy JM *et al.* (2001) Intrachromosomal genomic instability in human sporadic colorectal cancer measured by genome-wide allelotyping and inter-(simple sequence repeat) PCR. *Cancer Res.* **61**, 8274–8283.
- 48 Stephens PJ, McBride DJ, Lin M-L, Varela I, Pleasance ED, Simpson JT *et al.* (2009) Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462**, 1005–1010.
- 49 Birchler JA, Yao H, Chudalayandi S (2007) Biological consequences of dosage dependent gene regulatory systems. *Biochim. Biophys. Acta* **1769**, 422–428.
- 50 Torres EM, Dephoure N, Panneerselvam A, Tucker CM, Whittaker CA, Gygi SP *et al.* (2010) Identification of aneuploidy-tolerating mutations. *Cell* **143**, 71–83.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 A functional analysis of the widely expressed and tissue-specific gene expression classes.

Table S2 The average numbers of DNA pieces from normal homologues, highly selected derivative chromosomes, total DNA pieces, and SNP classifications of the 110 A549, DU-145, LN-18, PC-3 and SUIT-2 somatic chromosomes. (a) Chromosomes with heterozygous SNP-maps. (b) Chromosomes with broken-chromosome LOH SNP-maps. (c) Chromosomes with whole-chromosome LOH SNP-maps. Chromosomes are sequentially listed from smallest to largest number of normal homologues in each SNP-map category.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.