nature genetics

Genetic clonal diversity predicts progression to esophageal adenocarcinoma

Carlo C Maley[1], Patricia C Galipeau[2,3], Jennifer C Finley[4], V Jon Wongsurawat[5], Xiaohong Li[2,3], Carissa A Sanchez[2,3], Thomas G Paulson[2,3], Patricia L Blount[2,3,5], Rosa-Ana Risques[4], Peter S Rabinovitch[2–4] & Brian J Reid[2,3,5,6]

Neoplasms are thought to progress to cancer through genetic instability generating cellular diversity[1,2] and clonal expansions driven by selection for mutations in cancer genes[3,4]. Despite advances in the study of molecular biology of cancer genes[5], relatively little is known about evolutionary mechanisms that drive neoplastic progression. It is unknown, for example, which may be more predictive of future progression of a neoplasm: genetic homogenization of the neoplasm, possibly caused by a clonal expansion, or the accumulation of clonal diversity. Here, in a prospective study, we show that clonal diversity measures adapted from ecology and evolution can predict progression to adenocarcinoma in the premalignant condition known as Barrett's esophagus, even when controlling for established genetic risk factors, including lesions in *TP53* (*p53*; ref. 6) and ploidy abnormalities[7]. Progression to cancer through accumulation of clonal diversity, on which natural selection acts, may be a fundamental principle of neoplasia with important clinical implications.

Much research in cancer biology has focused on genetic instability, tumor suppressor genes and oncogenes[5] within cells. However, few studies have addressed genetic effects at the level of populations of cells and evolutionary mechanisms that drive neoplastic progression[3,8]. It is unknown whether the state of tumor suppressor and oncogenes and the presence of genetic instability provide all relevant information for the prognosis of a neoplasm or whether clonal diversity provides additional prognostic information. Little attention has been paid to genetic heterogeneity within a neoplasm as identified by lesions in evolutionarily neutral loci[9,10], except for an investigation of the role of neutral loci in microsatellite instability[11]. To our knowledge, clonal diversity has never been used to predict progression to cancer.

Barrett's esophagus is a premalignant condition uniquely suited for prospective study of neoplastic clonal evolution[6,7,12,13]. It has recently been recognized as a neoplasm[10] because it is generally clonal[10,13] and hyperproliferative and carries over 30-fold increased risk of progressing to cancer[14]. In a prospective study, Barrett's segment length was a weak predictor of progression to esophageal adenocarcinoma[15]. Genetic instability occurs at high frequency even in the earliest stages of Barrett's esophagus[13,16], yet only a minority (<5%) of individuals with Barrett's esophagus progress to esophageal adenocarcinoma in their lifetimes[17,18]. Genetic lesions important in Barrett's esophagus progression, including aneuploidy[7], tetraploidy[7] and loss of heterozygosity (LOH) in *CDKN2A* (also known as *p16* or *INK4A*)[13] and TP53 (ref. 6), are common to many neoplasms[19–21]. Periodic biopsy surveillance of Barrett's esophagus for early detection of esophageal adenocarcinoma[22] allows for analysis of clonal diversity over time and space *in vivo*.

We have adapted diversity measures from ecology[23] and evolution[24,25] to quantify clonal diversity in Barrett's esophagus. Here, each sample is not a single organism but consists of thousands of cells from a purified portion of a biopsy. A simple measure of diversity is the number of clones in a neoplasm. Ecological measures of diversity typically integrate both number and abundance of clones[23]. The Shannon diversity index (H) is

$$H = -\sum_i p_i \ln(p_i)$$

where p_i is the frequency of clone i in the neoplasm. The Shannon diversity index is preferable to other common diversity measures such as Simpson's index because it is not dominated by the most frequent clone. Results were not significantly different when we used a correction for small sample numbers[23]. Genetic divergence, similar to taxonomic distinctness[24], is the number of loci showing differences by LOH divided by the number of informative (heterozygous in normal) loci. The mean genetic divergence between all samples within a neoplasm is not dependent on the number of samples or clones[24] but rather on their accumulation of genetic differences since they diverged from a common ancestor. We calculated diversity measures within each neoplasm at baseline in a cohort of 268 individuals with Barrett's esophagus using a uniform sampling of one biopsy for every 1 to 2 cm of the Barrett's segment (**Fig. 1**). Systematic sampling across space is necessary because clones in Barrett's esophagus neoplasms are

[1]The Wistar Institute, 3601 Spruce St., Philadelphia, Pennsylvania 19104, USA. Divisions of [2]Human Biology and [3]Public Health Sciences, Fred Hutchinson Cancer Research Center, P.O. Box 19024, Seattle, Washington 98109, USA. [4]Department of Pathology, [5]Department of Medicine and [6]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. Correspondence should be addressed to C.M. (cmaley@alum.mit.edu).

Received 11 March 2005; accepted 23 February 2006; published online 26 March 2006; doi:10.1038/ng1768

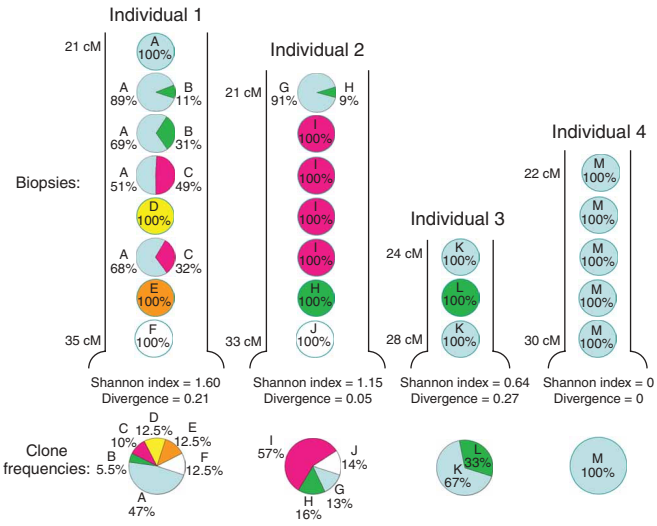

Figure 1 Clonal diversity in four Barrett's segments. Genetically distinct fractions of a biopsy are shown at each level with their frequencies. Clones and total frequencies within a segment (larger pie charts below) were used to calculate Shannon diversity and mean pairwise genetic divergence scores. The apparently noncontiguous clones in Individuals 1–3 may be attributable to interleaving of contiguous clones in two dimensions[26], biopsies taken from different quadrants in different levels or cell turnover and clonal expansion. Individual 1 has a high Shannon index and high divergence (clone A, diploid, LOH in *TP53* and D9S1121; B, aneuploid (3.7*N*), LOH in *TP53* and D9S1121; C, tetraploid, LOH in *TP53* and D9S1121; D, diploid, LOH in *TP53*, D9S925, D9S1121 and D9S1118; E, diploid, LOH in *TP53*; F, diploid with no LOH). Individual 2 has a high Shannon index and low divergence (clone G, diploid, LOH at D9S2169 and D9S1121; H, diploid, LOH at D9S2169, D9S935 and D9S1118; I, diploid, LOH at D9S2169, D9S935, D9S925 and D9S1118; J, diploid with no LOH). Individual 3 has a low Shannon index and high divergence (clone K, diploid, LOH at D9S935, D9S925 and D9S1118; L, diploid, LOH in *TP53*). Individual 4 has a low Shannon index and low divergence (clone M, diploid, LOH in all markers on chromosome 9p). An example of divergence can be calculated in Individual 3 based on six samples (15 pairwise comparisons). Clones K and L differ by two LOH events in four informative loci, resulting in a divergence of 0.27.

spatially heterogeneous and are not all represented in a single biopsy (**Fig. 2**)[12,26]. We identified clones by any difference in flow cytometric DNA content (for differences $>0.2N$), LOH (**Fig. 2**), microsatellite shifts (new alleles) and *CDKN2A* or *TP53* sequence mutations. We monitored participants with serial endoscopies to determine cancer outcome.

Participants in the cohort (**Table 1**) were followed for an average of 4.4 years (range 0.1–8.4). Thirty-seven esophageal adenocarcinomas developed during this period. Neither age (relative risk (RR) = 1.00, 95% confidence interval (c.i.): 0.97–1.02) nor sex (RR = 1.22, 95% c.i.: 0.57–2.60) was significantly associated with cancer outcome. Barrett's segment length and number of samples were not as strong predictors of progression as diversity measures were and were not statistically significant when combined with other factors (**Fig. 3**, **Supplementary Fig. 1** and **Supplementary Note** online).

Our data set encompasses a variety of molecular lesions, including LOH and microsatellite shifts at 19 loci spanning chromosomes 9 and 17, as well as sequence mutations in *TP53* and *CDKN2A*. Statistically significant diversity indices based on LOH alone were better predictors of progression (higher RR values and lower *P* values) than indices based on LOH, shifts and sequence mutations (**Table 2**). The upper quartiles of the number of clones, Shannon index and genetic divergence based on LOH were strongly predictive of increased progression to esophageal adenocarcinoma (**Fig. 3**; Kaplan-Meier log rank test, $P < 0.001$ for all), showing that neoplasms with greater clonal diversity at baseline were more likely to progress to cancer than neoplasms with lesser clonal diversity.

Controlling for 17p (*TP53*) LOH, tetraploidy and aneuploidy, which are known to be associated with genomic instability, we found that number of clones, divergence and Shannon index, all based on LOH, individually predict cancer outcome (**Table 2**, right column). All three diversity measures significantly improve cancer outcome prediction (likelihood ratio test, $P < 0.05$) when added to a multivariate Cox model including 17p (*TP53*) LOH, aneuploidy and

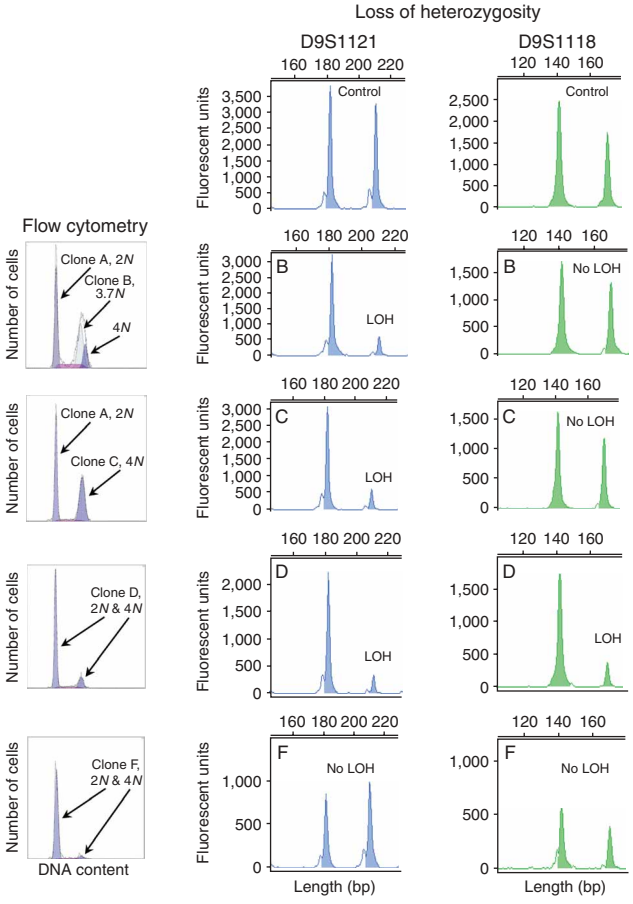

Figure 2 Primary flow cytometric and microsatellite LOH data from clones B, C, D and F in Individual 1 (**Fig. 1**). The left column shows flow cytometry histograms with DNA content on the *x*-axis and frequency on the *y*-axis. Cell-cycle fractions (filled curves) have been fit to the raw data (jagged lines) by Multicycle software (Phoenix Flow Systems). Diploid, Ki67-negative cells were excluded owing to the presence of stroma and other nonepithelial cells. Clone A is diploid, clone B is composed of the cells in the aneuploid fraction (3.7*N*) and clone C is the tetraploid (increased 4*N*) fraction. The two columns at right show microsatellite data, with fluorescent units on the *y*-axis and length of microsatellites (in bp) on the *x*-axis for loci D9S1121 and D9S1118 on chromosome 9p. The top row shows the values for gastric biopsies from the same individual used as controls. Clones B, C and D have LOH at D9S1121 and clone D has LOH at D9S1118.

Table 1 Cohort characteristics

Total subjects	268	
Sex		
Male	212	(79%)
Female	56	(21%)
Age (years)		
<40	16	(6%)
40–49	49	(18%)
50–59	68	(25%)
60–69	65	(24%)
70–79	60	(22%)
≥80	10	(4%)
Segment length		
<3 cm	85	(32%)
3–6 cm	107	(40%)
7–10 cm	55	(21%)
>10 cm	21	(8%)

tetraploidy. The combination of factors that best predicted progression included the number of clones, genetic divergence, TP53 LOH and ploidy lesions (**Table 3**, multivariate Cox model with best Akaike Information Criterion (AIC)). Segment length, number of samples and *CDKN2A* LOH do not provide independent predictive value for cancer outcome. The Shannon index and the number of clones are strongly correlated (Pearson's $R = 0.92$), and one can substitute for the other in a multivariate model without significantly changing the predictive value of the model (likelihood ratio test, $P > 0.05$). However, number of clones was a slightly better predictor of progression than the Shannon index, and given the relative ease of measuring number of clones in a neoplasm, it may be a more useful measure of clonal diversity than the Shannon index.

We compared clonal diversity of Barrett's esophagus epithelium with 17p (*TP53*) LOH to Barrett's esophagus without *TP53* LOH. *TP53* LOH epithelium had a higher Shannon index (mean 0.56, range 0–1.65; Wilcoxon rank sum test, $P < 0.01$) and more clones per sample (mean 0.64, range 0.20–1.00; Wilcoxon rank sum test, $P < 0.001$) than epithelium without *TP53* LOH (mean 0.36, range 0–1.63; mean 0.42, range 0.08–1.00, respectively). Samples from *TP53* LOH epithelium did not have greater divergence, as measured by LOH at other loci (mean 0.042, range 0–0.500), than samples without *TP53* LOH (mean 0.032, range 0–0.429; Wilcoxon rank sum test, $P = 0.31$).

Increased Shannon diversity but lack of increased divergence in *TP53* LOH epithelium suggests that *TP53* LOH increases generation of viable genetic variants that may derive from a recent common ancestor. This hypothesis predicts that *TP53* LOH epithelium would typically be clonal. In 40 of 45 (89%) patients, *TP53* LOH epithelium showed loss of identical alleles in loci surrounding *TP53* in multiple samples, consistent with a clonal origin and with previous studies[10]. Four of the five remaining individuals show evidence of two independent *TP53* LOH clones.

If *TP53* LOH is simply a manifestation of genetic instability, then LOH in other loci should also show clonal diversity differences between tissue with and without LOH in those loci. We repeated the analysis twice and found no significant differences in the Shannon diversity between tissues with and without 17q or 9q LOH. Likewise, clonal diversity is not simply a manifestation of time since loss of *TP53* (**Supplementary Note**).

The association of clonal diversity with progression to esophageal adenocarcinoma provides an evolutionary mechanism for neoplastic progression at the clonal level but does not address cellular and molecular mechanisms generating diversity. To explore cellular mechanisms generating diversity, we used FISH probes to the centromere and *TP53* locus on chromosome 17 in 47 samples from Barrett's esophagus epithelium with or without *TP53* LOH. We calculated the Shannon index based on diversity of numbers of centromere and *TP53* spots in cells from a biopsy. Samples without 17p (*TP53*) LOH showed less cellular diversity (mean Shannon index = 0.65, s.d. = 0.46) than biopsies with *TP53* LOH (mean Shannon index = 0.93, s.d. = 0.43, two-tailed *t*-test $P < 0.05$). These data confirm that *TP53* LOH is associated with increased diversity at the cellular level.

Because Barrett's esophagus is a chronic inflammatory, hyperproliferative epithelium, we hypothesized that cellular diversity may be

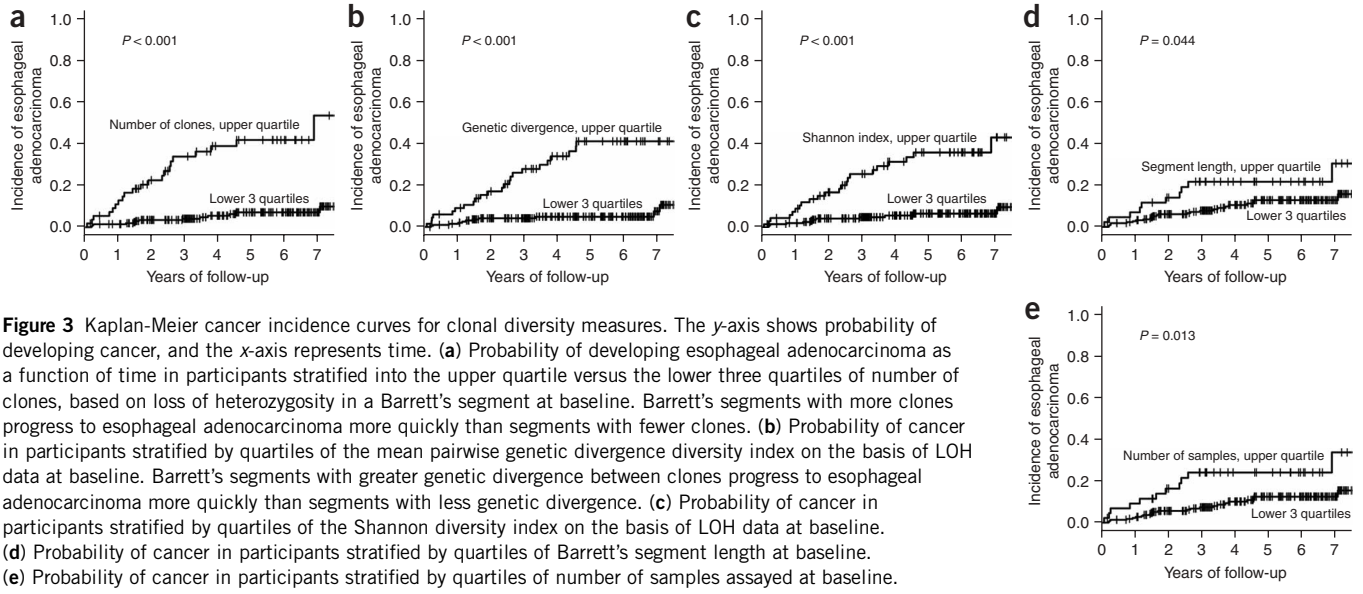

Figure 3 Kaplan-Meier cancer incidence curves for clonal diversity measures. The *y*-axis shows probability of developing cancer, and the *x*-axis represents time. (**a**) Probability of developing esophageal adenocarcinoma as a function of time in participants stratified into the upper quartile versus the lower three quartiles of number of clones, based on loss of heterozygosity in a Barrett's segment at baseline. Barrett's segments with more clones progress to esophageal adenocarcinoma more quickly than segments with fewer clones. (**b**) Probability of cancer in participants stratified by quartiles of the mean pairwise genetic divergence diversity index on the basis of LOH data at baseline. Barrett's segments with greater genetic divergence between clones progress to esophageal adenocarcinoma more quickly than segments with less genetic divergence. (**c**) Probability of cancer in participants stratified by quartiles of the Shannon diversity index on the basis of LOH data at baseline. (**d**) Probability of cancer in participants stratified by quartiles of Barrett's segment length at baseline. (**e**) Probability of cancer in participants stratified by quartiles of number of samples assayed at baseline. Because two samples are purified from a biopsy at 2-cm intervals of the Barrett's segment, the number of samples is closely correlated with the segment length.

Table 2 Diversity measures and the univariate relative risk for progression to esophageal adenocarcinoma

				Controlling for *TP53* and ploidy abnormalities	
Predictor	Range	RR (95% c.i.)	*P* value	RR (95% c.i.)	*P* value
Segment length (per cm)	1–19	1.16 (1.08–1.24)	<0.001	1.10 (1.03–1.19)	0.006
Number of clones (per clone)	1–9	1.68 (1.47–1.91)	<0.001	1.25 (1.05–1.49)	0.010
Number of LOH clones (per clone)	1–9	1.99 (1.71–2.32)	<0.001	1.40 (1.13–1.73)	0.002
Clones per sample (per 0.1)	0.1–1.0	1.18 (1.04–1.34)	0.012	0.93 (0.79–1.09)	0.38
LOH clones per sample (per 0.1)	0.08–1.0	1.18 (1.04–1.34)	0.008	0.94 (0.80–1.10)	0.45
Shannon diversity index (per 1.0)	0.0–2.10	8.46 (4.59–15.6)	<0.001	2.60 (1.25–5.38)	0.010
Shannon LOH diversity index (per 1.0)	0.0–2.05	11.0 (5.80–21.0)	<0.001	3.10 (1.37–7.01)	0.006
Mean pairwise divergence (per 0.1)	0.00–0.54	1.96 (1.54–2.50)	<0.001	1.38 (1.03–1.85)	0.031
Mean pairwise divergence by LOH (per 0.1)	0.00–0.54	2.15 (1.67–2.77)	<0.001	1.45 (1.08–1.95)	0.014

Each diversity index has been calculated with all the available genetic data (LOH and shifts in 19 microsatellites, sequence mutations in *CDKN2A* and *TP53* and ploidy differences $> 0.2N$) as well as calculated from the LOH and ploidy data alone (Cox regression analysis). Each sample consisted of the amplified DNA from a flow-sorted fraction of a biopsy. Clones per sample and mean pairwise divergence are constrained to the unit interval, so the relative risks have been calculated for increments of 0.1. Because *TP53* LOH and DNA content abnormalities (aneuploidy and tetraploidy) are associated with genomic instability, estimates of the relative risks of the diversity indices have been adjusted by including *TP53*, tetraploidy and aneuploidy as potential confounders in the rightmost column (Cox regression analyses).

generated by genomic instability owing to telomere shortening and subsequent bridge-breakage-fusion cycles. Telomere length was negatively correlated with Shannon index on the basis of patterns of FISH spots for the *TP53* locus and chromosome 17 centromere (Pearson's correlation, $R = -0.64$, 95% c.i.: –0.41 to –0.79, $P < 0.001$). In addition, we investigated the frequency of anaphase bridges in two primary Barrett's esophagus epithelial cell cultures known to contain the same genetic abnormalities present in the Barrett's segment from which they were derived[27]. The primary cultures had 0.14%–0.17% 4*N* cells with visible anaphase bridges, compared with no visible bridges in either culture after transduction with telomerase (stratified exact test, $P < 0.05$ for both cultures, data not shown).

Our results indicate that cellular and clonal diversity are fundamental to neoplastic progression and provide links between molecular mechanisms that maintain genome integrity and the evolution of neoplastic cell lineages. To our knowledge, this is the first demonstration that clonal diversity predicts progression to cancer and, thus, that the accumulation of viable, clonal genetic variants is a greater risk factor for progressing to cancer than a recent homogenizing clonal expansion (with a Shannon index and genetic divergence of 0). Evolutionarily neutral loci[10] provide additional prognostic information beyond the risk that can be attributed to lesions associated with genetic instability such as *TP53* LOH, tetraploidy and aneuploidy.

Clonal diversity is related to, but distinct from, genetic instability and cellular diversity (**Supplementary Note**). Genetic instability will lead to clonal diversity only if the genetic variants are viable and can expand into detectable clones. We have shown that clonal diversity predicts progression to cancer, although whether cellular diversity predicts progression remains an open question.

There are many potential sources of genetic instability, and thus genetic diversity, including genetic and epigenetic lesions as well as environmental exposures. One advantage of measuring clonal diversity is that it unifies all sources of genetic instability. Our results indicate that cellular diversity increases as molecular mechanisms to maintain genome integrity fail. Longer telomeres seem to be protective against generation of cellular diversity, and telomere shortening is associated with increased cellular Shannon diversity, as assessed by FISH, as well as anaphase bridges, providing a mechanistic link between short telomeres and generation of variants. Activation of telomerase may have a role in the generation of viable genetic variants (**Supplementary Note**). Genetic instability produces cellular diversity, which provides a fertile field from which natural selection can operate. However, until clonally expanded, these genetic variants, as seen in FISH abnormalities, represent evolutionary dead ends.

Our observation that epithelium with *TP53* LOH has high Shannon indices but not high divergence suggests that before *TP53* LOH, evolutionary dynamics are dominated by clonal expansion driven by loss of *CDKN2A* on chromosome 9p (refs. 10,13). Normal *TP53* function may prevent most genetic variants with chromosomal aberrations from expanding into clones. After *TP53* LOH, evolutionary dynamics are dominated by generation and expansion of new genetic variants, which seem to occur at very high frequencies as assessed by FISH in *TP53* LOH clones. Natural selection favoring some variants in clonally diverse *TP53* LOH epithelium provides an evolutionary mechanism to explain why larger clones with *TP53* LOH pose a greater risk of progression than smaller ones[28].

The generation of cellular genetic variants and clonal diversity on which natural selection acts may be a fundamental evolutionary mechanism of neoplastic progression with profound clinical implications. Our diversity measures are based on one-fourth the number of biopsies normally used for pathological grading[29]. If confirmed in other Barrett's esophagus cohorts and other neoplasms, then neoplasms may be considered as evolving ecosystems in which evolutionary and ecological measures of diversity are widely applicable for

Table 3 Best multivariate Cox model determined by stepwise selection

Predictors	RR (95% c.i.)	*P* value
Number of LOH clones (per clone)	1.43 (1.16–1.72)	<0.001
Number of LOH clones per sample (per 1.0)	0.25 (0.04–1.47)	0.130
Mean pairwise LOH divergence (per 0.1)	1.62 (1.15–2.27)	0.005
17p (*TP53*) LOH	3.59 (1.36–9.45)	0.010
Ploidy abnormality	4.79 (1.73–13.24)	0.003

A backwards and forwards stepwise Cox multivariate analysis based on the Akaike Information Criterion selected the model with the best fit. Mean pairwise divergence and number of clones are the best diversity measures, contributing independent information for the calculation of risk of progression to cancer. The length of the Barrett's segment and number of samples did not provide predictive value independent of the diversity measures. Replacing the number of clones with the Shannon index does not significantly reduce the fit of the model, owing to the high correlation (Pearson's $R = 0.92$) between the two diversity measures. LOH in *CDKN2A* was also excluded from the best-fit model by stepwise selection.

assessment of risk of progression to cancer because they quantify genetic heterogeneity within viable, evolving clones; integrate all mechanisms generating genomic instability and are easily generalizable to other premalignant neoplasms, as well as being readily scaled from single cells (as in FISH studies) to entire neoplasms. Clinically, assessment of clonal diversity may be a unified method to identify high-risk patients for early detection as well as warning of possible variants that may be resistant to cancer prevention interventions.

METHODS

Research participants and endoscopies. Endoscopic biopsies were collected with a standardized protocol[6] at 1–2 cm intervals of the Barrett's segment from 268 affected individuals with at least one follow-up endoscopy, between January 5, 1995 and August 14, 2003, for a total of 1,179 patient years (median 4.5, range 0.1–8.4 years). The Seattle Barrett's Esophagus Study has been approved annually by the Human Subjects Division of the University of Washington and/or the Institutional Review Board of the Fred Hutchinson Cancer Research Center since 1983. Informed consent was obtained from all participants.

Molecular data. Each biopsy was purified and separated by flow sorting into diploid proliferating cells (Ki67-positive) and either aneuploid cells, if present, or cells with 4*N* DNA content[12]. We divided 94.3% of the biopsies into two fractions, 0.7% into three fractions and 4.9% into one fraction. Tetraploidy and aneuploidy were defined as described previously[7,10]. DNA was extracted and amplified[12] for 1,467 sorted fractions as well as three constitutive controls per participant. Each sorted fraction was assayed for LOH and microsatellite shifts (new alleles) in 11 microsatellite loci spanning chromosome 17 in 256 participants and eight loci spanning chromosome 9 in 259 participants[12]. Exons 5–9 of the *TP53* gene were sequenced[12] for 839 flow-purified fractions from 236 participants. We analyzed mutations in exon 2 of the *CDKN2A* gene[13] in 1,195 flow-purified fractions from 239 participants.

FISH analysis. Forty-seven flow-purified samples of Barrett's esophagus from 19 adenocarcinoma esophagectomies were analyzed for 17p LOH and FISH. For comparison of FISH and telomeres, 22 Barrett's biopsies and 20 gastric controls were taken from 17 endoscopies and seven esophagectomies. We sorted epithelial cells from half of each biopsy for FISH by flow cytometry as described above. Cells sorted onto plain glass slides in 5 mM $CaCl_2$ were allowed to dry overnight and were subsequently fixed using 3:1 methanol:acetic acid. FISH was performed as described previously[30]. An average of 86 nuclei (range 17–133) were counted per probe and sample pair. Probes used were TP53 (17p13.1) and CEP 17 (centromere) and were directly conjugated as FITC/Spectrum Orange centromere/arm dual labels (Vysis).

Telomere length measurement. Telomere length was measured as described previously[30] on half of each of the same biopsies that were used for FISH analysis.

Anaphase bridges. We quantified anaphase bridges as previously described[30] using two primary, mortal epithelial cell cultures from Barrett's esophagus and the same two cell cultures after transduction with the reverse transcriptase component of human telomerase (hTERT)[27]. To enrich for bridged cells, the cells were stained with DAPI, and cells with G2/4N DNA content were sorted onto glass slides. For each slide, an average of 1,750 cells were examined by fluorescence microscopy for anaphase bridges. Bridged cell counts were expressed relative to the presorted cell number.

Statistical analysis. Follow-up time was calculated between baseline and the last endoscopy before August 14, 2003 or the first endoscopy with esophageal adenocarcinoma.

We have adapted three indices of diversity from ecology and evolutionary biology: number of clones, the Shannon diversity index and the mean pairwise genetic divergence. We distinguished clones either by ploidy differences $>0.2N$ and differences in LOH in the 19 microsatellites or by the entire set of molecular data described above. We did not distinguish clones by spatial location. In the comparison of epithelium with and without LOH (in 17p (*TP53*), 17q or 9q), loci on the chromosome arm used to distinguish the two types of tissue were potential confounders and therefore were not used to distinguish clones for that analysis. Clones per sample were calculated by dividing the number of clones by the number of samples (flow-sorted fractions, usually two per biopsy) taken from a Barrett's segment. **Figure 1** shows four examples from our cohort.

The Shannon index of diversity[23] measures both the number and abundance of clones and is defined by

$$H = -\sum_i p_i \ln(p_i)$$

where p_i is the abundance of clone i, which is the proportion of flow-sorted cells (excluding nonproliferating diploid cells as previously described[10]) that share genotype i. Uninformative loci and missing data did not distinguish clones.

The divergence between two samples in a neoplasm was the number of molecular markers differing between the two samples divided by the number of informative assays. If loci with LOH were adjacent or separated by uninformative loci, they were considered only one loss for the purpose of computing the divergence between two samples. The mean pairwise divergence (D) is defined as

$$D = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{X_{i,j}}{Y_{i,j}}$$

where n is the number of samples in a patient, $X_{i,j}$ is the number of noncontiguous loci that show LOH in either sample i or j but not in both and $Y_{i,j}$ is the number of loci that are heterozygous in the normal tissue and for which genotyping produced data for both samples i and j.

Univariate Cox proportional hazards regression models were used to test the association of each diversity measure with cancer outcome. We also tested each diversity measure in a multivariate Cox model that included the potential confounders of 17p LOH, aneuploidy and tetraploidy. We performed a forward and backward stepwise analysis to identify the multivariate Cox model with the best AIC. We performed Kaplan-Meier analyses by comparing the upper quartiles of the diversity indices against the lower three quartiles.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank R. Klausner, S. Self, S. Moolgavkar and H. Tang for their helpful suggestions. This work was supported by grants from the US National Institutes of Health (P01 CA91955, K01 CA89267-02 and K07 CA89147-03) and from funds provided by the Commonwealth Universal Research Enhancement Program of the Pennsylvania Department of Health.

AUTHORS' CONTRIBUTIONS

C.C.M. conceived of the experimental analysis based on the prospective cohort study data collected by B.J.R., P.L.B., P.S.R., C.A.S., P.C.G., T.G.P., V.J.W. and P.S.R. C.C.M., P.C.G. and X.L. analyzed the clonal data. J.C.F., R.-A.R., T.G.P. and P.S.R. contributed the telomere, telomerase and FISH data. C.C.M. wrote the paper with input and revisions from all the other authors.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Gonzalez-Garcia, I., Sole, R.V. & Costa, J. Metapopulation dynamics and spatial heterogeneity in cancer. *Proc. Natl. Acad. Sci. USA* **99**, 13085–13089 (2002).
2. Rajagopalan, H., Nowak, M.A., Vogelstein, B. & Lengauer, C. The significance of unstable chromosomes in colorectal cancer. *Nat. Rev. Cancer* **3**, 695–701 (2003).
3. Nowell, P.C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).
4. Sieber, O.M., Heininmann, K. & Tomlinson, I.P. Genomic instability - the engine of tumorigenesis? *Nat. Rev. Cancer* **3**, 701–708 (2003).

5. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
6. Reid, B.J. *et al.* Predictors of progression in Barrett's esophagus II: baseline 17p (p53) loss of heterozygosity identifies a patient subset at increased risk for neoplastic progression. *Am. J. Gastroenterol.* **96**, 2839–2848 (2001).
7. Rabinovitch, P.S., Longton, G., Blount, P.L., Levine, D.S. & Reid, B.J. Predictors of progression in Barrett's esophagus III: baseline flow cytometric variables. *Am. J. Gastroenterol.* **96**, 3071–3083 (2001).
8. Breivik, J. The evolutionary origin of genetic instability in cancer development. *Semin. Cancer Biol.* **15**, 51–60 (2005).
9. Tsao, J.L. *et al.* Genetic reconstruction of individual colorectal tumor histories. *Proc. Natl. Acad. Sci. USA* **97**, 1236–1241 (2000).
10. Maley, C.C. *et al.* Selectively advantageous mutations and hitchhikers in neoplasms: p16 lesions are selected in Barrett's esophagus. *Cancer Res.* **64**, 3414–3427 (2004).
11. Boland, C.R. *et al.* A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* **58**, 5248–5257 (1998).
12. Barrett, M.T. *et al.* Evolution of neoplastic cell lineages in Barrett oesophagus. *Nat. Genet.* **22**, 106–109 (1999).
13. Wong, D.J. *et al.* p16 INK4a lesions are common, early abnormalities that undergo clonal expansion in Barrett's metaplastic epithelium. *Cancer Res.* **61**, 8284–8289 (2001).
14. Haggitt, R.C. Barrett's esophagus, dysplasia, and adenocarcinoma. *Hum. Pathol.* **25**, 982–993 (1994).
15. Rudolph, R.E. *et al.* The effect of segment length on the risk of neoplastic progression in patients with Barrett's esophagus. *Ann. Intern. Med.* **132**, 612–620 (2000).
16. Doak, S.H. *et al.* Chromosome 4 hyperploidy represents an early genetic aberration in premalignant Barrett's oesophagus. *Gut* **52**, 623–628 (2003).
17. Conio, M. *et al.* Long-term endoscopic surveillance of patients with Barrett's esophagus. Incidence of dysplasia and adenocarcinoma: a prospective study. *Am. J. Gastroenterol.* **98**, 1931–1939 (2003).
18. Macdonald, C.E., Wicks, A.C. & Playford, R.J. Final results from 10 year cohort of patients undergoing surveillance for Barrett's oesophagus: observational study. *Br. Med. J.* **321**, 1252–1255 (2000).
19. Shackney, S.E. & Shackney, T.V. Common patterns of genetic evolution in human solid tumors. *Cytometry* **29**, 1–27 (1997).
20. Rocco, J.W. & Sidransky, D. p16(MTS-1/CDKN2/INK4a) in cancer progression. *Exp. Cell Res.* **264**, 42–55 (2001).
21. Sherr, C.J. Tumor surveillance via the ARF-p53 pathway. *Genes Dev.* **12**, 2984–2991 (1998).
22. Sampliner, R.E. Updated guidelines for the diagnosis, surveillance, and therapy of Barrett's esophagus. *Am. J. Gastroenterol.* **97**, 1888–1895 (2002).
23. Magurran, A.E. *Measuring Biological Diversity* (Blackwell, Malden, Massachusetts, 2004).
24. Clarke, K.R. & Warwick, R.M. A taxonomic distinctness index and its statistical properties. *J. Appl. Ecol.* **35**, 523–531 (1998).
25. Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York, 1987).
26. Rabinovitch, P.S., Reid, B.J., Haggitt, R.C., Norwood, T.H. & Rubin, C.E. Progression to cancer in Barrett's esophagus is associated with genomic instability. *Lab. Invest.* **60**, 65–71 (1989).
27. Palanca-Wessels, M.C. *et al.* Extended lifespan of Barrett's esophagus epithelium transduced with the human telomerase catalytic subunit: a useful *in vitro* model. *Carcinogenesis* **24**, 1183–1190 (2003).
28. Maley, C.C. *et al.* The combination of genetic instability and clonal expansion predicts progression to esophageal adenocarcinoma. *Cancer Res.* **64**, 7629–7633 (2004).
29. Levine, D.S., Blount, P.L., Rudolph, R.E. & Reid, B.J. Safety of a systematic endoscopic biopsy protocol in patients with Barrett's esophagus. *Am. J. Gastroenterol.* **95**, 1152–1157 (2000).
30. O'Sullivan, J.N. *et al.* Chromosomal instability in ulcerative colitis is related to telomere shortening. *Nat. Genet.* **32**, 280–284 (2002).