Identification of a key role of widespread epigenetic drift in Barrett’s esophagus and esophageal adenocarcinoma

E. Georg Luebeck 1*, Kit Curtius 2*, William D. Hazelton 1, Sean Maden 3, Ming Yu 3, Prashanthi N. Thota 4, Deepa T. Patil 5, Amitabh Chak 6, Joseph E. Willis 6 and William M. Grady 3,7

Abstract

Background: Recent studies have identified age-related changes in DNA methylation patterns in normal and cancer tissues in a process that is called epigenetic drift. However, the evolving patterns, functional consequences, and dynamics of epigenetic drift during carcinogenesis remain largely unexplored. Here we analyze the evolution of epigenetic drift patterns during progression from normal squamous esophagus tissue to Barrett’s esophagus (BE) to esophageal adenocarcinoma (EAC) using 173 tissue samples from 100 (nonfamilial) BE patients, along with publically available datasets including The Cancer Genome Atlas (TCGA).

Results: Our analysis reveals extensive methylomic drift between normal squamous esophagus and BE tissues in nonprogressed BE patients, with differential drift affecting 4024 (24%) of 16,984 normally hypomethylated cytosine-guanine dinucleotides (CpGs) occurring in CpG islands. The majority (63%) of islands that include drift CpGs are associated with gene promoter regions. Island CpGs that drift have stronger pairwise correlations than static islands, reflecting collective drift consistent with processive DNA methylation maintenance. Individual BE tissues are extremely heterogeneous in their distribution of methylomic drift and encompass unimodal low-drift to bimodal high-drift patterns, reflective of differences in BE tissue age. Further analysis of longitudinally collected biopsy samples from 20 BE patients confirm the time-dependent evolution of these drift patterns. Drift patterns in EAC are similar to those in BE, but frequently exhibit enhanced bimodality and advanced mode drift. To better understand the observed drift patterns, we developed a multicellular stochastic model at the CpG island level. Importantly, we find that nonlinear feedback in the model between mean island methylation and CpG methylation rates is able to explain the widely heterogeneous collective drift patterns. Using matched gene expression and DNA methylation data in EAC from TCGA and other publically available data, we also find that advanced methylomic drift is correlated with significant transcriptional repression of ~200 genes in important regulatory and developmental pathways, including several checkpoint and tumor suppressor-like genes.

Conclusions: Taken together, our findings suggest that epigenetic drift evolution acts to significantly reduce the expression of developmental genes that may alter tissue characteristics and improve functional adaptation during BE to EAC progression.

Keywords: Barrett’s esophagus (BE), Esophageal adenocarcinoma (EAC), Tissue age, Epigenetic drift, DNA methylation, Neoplastic progression, Transcriptional repression in cancer, Endogenous retroviruses (ERVs)
Background
Research into the connection between aging and cancer is being fueled by advances in molecular profiling of age-related processes such as chronic inflammation, accumulation of somatic DNA/mtdNA mutations, and epigenetic changes in tissues in which cancers arise [1]. Two distinct, albeit not entirely independent, concepts have emerged recently that relate changes in DNA methylation to biological tissue age. The first is based on the discovery (of sets) of CpG dinucleotides (CpGs) in the genome that are subject to age-dependent, possibly complex changes in methylation levels that, in aggregate, correlate strongly with chronological age [2–4]. We refer to these types of CpGs as clock-CpGs. A second and simpler concept is based on the observation of gradual age-related changes in methylation levels at specific CpG sites or CpG-rich regions, a process commonly referred to as epigenetic or methylomic drift [5–11]. For example, some CpG islands show very low methylation levels early in life but are known to become gradually methylated over time as a result of sporadic de novo methylation events during DNA replication. We identify these as drift CpGs. It is worth pointing out that data supporting these concepts come mainly from cross-sectional studies that include individuals of different age. In contrast, individual-level (longitudinal) drift, unless studied directly in select individuals over time as we have collected for this study, is typically inferred from population drift.

In a recent study, we used a combination of cross-sectional and longitudinally collected biopsy samples to identify a set of highly correlated CpGs in premalignant Barrett’s esophagus (BE) tissue that undergo differential epigenetic drift relative to normal squamous (NS) tissue [12]. This study arrived at a set of 67 drift CpGs that show significant age-related methylation differences between NS and BE and was used as an epigenetic clock to estimate the time of BE onset. Unlike previous epigenetic clocks that were constructed to predict the age of an individual, the BE tissue-specific clock model was designed to infer unknown tissue ages. Because BE is essentially asymptomatic, it is usually not known how long a patient has lived with BE. However, the time a patient has lived with BE may be considered a risk factor since older BE tissue has had more time for cancer to evolve compared to younger BE tissue which is more likely to be free of neoplastic changes [13, 14]. Although the earlier study was the first to develop an epigenetic clock for BE tissue age, we did not evaluate the full scope of epigenetic drift occurring in BE following its formation, nor its dynamics or functional consequences at the genomic level.

The aims of this study are to characterize individual heterogeneity and genome-wide patterns of age-related epigenetic drift in tissue samples from BE and EAC patients, to develop a mechanistic understanding of methylation dynamics during tissue aging, including the degree of cooperativity and possible presence of nonlinear feedback within CpG islands during the temporal evolution of epigenetic drift, and to explore the impact of advanced drift on gene expression in EACs for which we have both gene expression and methylation data.

To accomplish the aims of this study, it was necessary to analyze methylation drift patterns from an extensive collection of NS, BE, and EAC biopsy tissue samples, including 173 tissue samples from 100 nonprogressed and progressed (nonfamilial) BE patients, along with methylation and gene expression data in 87 EAC from The Cancer Genome Atlas (TCGA) [15] and in 47 EAC + 4 BE samples previously analyzed by Krause et al. [16].

Results
Using data from the HM450 methylation arrays (see “Methods”), we aimed to better characterize the full extent of epigenetic drift occurring in BE and the temporal dynamics of epigenetic drift and to explore the impact of advanced drift on gene expression in EAC for which we have both gene expression and methylation data. Methylation levels are measured either in terms of a β value (methylation fraction) or M value (logit2(β)) as indicated in the text.

Genomic scope of drift
Out of 146,029 hypomethylated CpG probes in normal squamous (NS) tissue, we identified 18,013 (12%) probes that have significant positive correlation and 560 (0.4%) that have a significant negative correlation (q < 0.01) with the mean differential drift (relative to NS levels) of 67 previously validated drift CpGs in 64 BE samples from patients without a diagnosis of dysplasia or cancer (see “Methods”). In contrast, out of 133,857 CpG probes that were hypermethylated in NS tissue, only 795 (0.6%) probes correlated positively and 3402 (2.5%) probes correlated negatively with the mean methylation drift levels of our 67 probe reference clock (Fig. 1). Thus, significant differential drift in BE involves thousands of CpGs, occurring predominantly in hypomethylated regions that are associated with CpG islands, affecting 4024 (24%) of the 16,984 hypomethylated CpG islands in NS tissue. In contrast, we found only 7% of the identified drift CpG probes to be “open sea,” i.e., isolated in the genome [17], compared to about 10% in the hypomethylated normal background.

The majority (63%) of islands that include drift CpGs are associated with gene promoter regions, i.e., they involve a transcription start site (TSS200 or TSS1500), while only 11% of islands that undergo drift overlap with the gene body, compared to 73% (TSS-associated) and 10% (body-associated) CpG islands on the HM450 array, respectively. In contrast, the relative abundance of intergenic CpG islands is significantly higher among CpG islands that
undergo drift compared to the fraction of intergenic CpG islands found on the array (24 vs. 16%, see Fig. 1).

Out of 16,832 island-based (BE-specific) drift CpGs we identified, 1317 unique CpG islands with 5 or more drift CpGs per island (comprising a total of 11,425 drift CpGs). Figure 2 presents a karyograph of 4 autosomes indicating the genomic locations and mean β values for these islands across the 64 BE samples (Additional file 1: Figure S4 for all 22 autosomes). Figure 3 shows a heatmap of the island-level mean β values of the 1317 drift-associated CpG islands for the first 10 NS tissue samples, and all 64 BE samples used to identify CpG probes undergoing drift. For this map, both CpG islands and tissue samples were ordered by their respective mean values. As expected, all 10 normal control tissue samples show no island-level drift. In contrast, we see significant heterogeneity in mean methylation levels of these CpG islands ranging from < 20 to > 80% methylation across the 64 BE samples. With the notable exception of a group of samples that have undergone minimal drift, most BE samples show bimodal patterns of drift where some islands appear to linger at low levels and others show advanced drift. We later use the following categorization for the observed drift patterns in BE and EAC: unimodal low drift (group L), bimodal high drift with a major mode β > 50% (group H) and the remaining bimodal intermediate drift (group I).

Pairwise correlations between island-associated CpGs

CpG islands are considered functional genomic units that may exert transcriptional control by their collective state of methylation rather than through individual CpG sites. To demonstrate this collective behavior in an island-level DNA methylation, we evaluated the pairwise correlations between all island CpGs that are hypomethylated in NS tissue. In general, for static islands (that do not show significant drift), pairwise correlations are moderate (< 0.5) across the span of an island and exhibit anti-correlations near and beyond the island boundaries (Fig. 4). In contrast, island CpGs that drift have stronger pairwise correlations reflecting a collective response of these CpGs to drift consistent with progressive DNA methylation maintenance [18, 19].

Figure 4 also shows that the pairwise correlations decay with genomic distance and, for drift CpGs, extend further into the island shelves than static CpGs. Islands that show tissue age-related drift are also significantly larger (in terms of genomic length) than static islands. The mean

---

**Fig. 1** Proportion of CpGs and CpG islands that drift differentially in Barrett’s esophagus vs. normal squamous (NS) esophagus among over 146 k hypomethylated probes in NS tissue

**Fig. 2** Representative karyographs of four autosomes. Chromosomes 2 and 12 exhibit typical methylomic drift patterns while chromosomes 17 and 19 exhibit high-density methylomic drift. Top track: chromosome banding. Middle track: array-based CpG island positions. Bottom track: positions of CpG islands that undergo methylomic drift in 64 BE samples (mean levels color-coded)
sizes of the static vs. drift CpG islands were 0.9 vs 1.1 kb, respectively ($p = 2 \times 10^{-10}$, two-sided t test).

**Bimodal nature of epigenetic drift in BE and EAC**

To see whether methylomic drift is uniformly distributed within our BE and EAC samples, we examined β value distributions for a subset of island-associated drift CpGs with a minimum of 5 detected drift CpGs per island (11,425 drift CpGs in total) for 64 BE and 24 EAC samples from the BETRNet (see “Methods”). Consistent with the patterns seen in Fig. 3, these individual-level distributions show signatures that fall into the arbitrary three types: with unimodal distributions showing low or no drift (group L), distinctly bimodal distributions with intermediate drift (both modes $\beta < 0.5$, group I), or bimodal with a major mode near or above $\beta = 0.5$ and a minor mode at lower levels (group H). (See Fig. 5a, for an aggregated view of the samples in these groups). While the distributions are similar for BE and EAC, EAC show more advanced drift in the third group (bimodal high) which may be attributed to EAC patients being on average older than the BE patients (68 vs 62 years, respectively), or to the fact that EAC undergoes more frequent stem cell divisions thereby increasing replication-coupled de novo methylation, or to the possibility that BE arises earlier in patients with EAC compared to patients who have not progressed to dysplasia or EAC. We found similar unimodal/bimodal drift signatures in 87 EAC from TCGA and in a combined set of 19 BE and 47 EAC tissue samples provided by Krause et al. [16] (GEO accession number: GSE72874).

**Advanced drift is associated with low tumor stage**

Using tumor stage information from the TCGA, we found a statistically significant association ($p$ value = 0.024; Fisher’s exact test) of low tumor stage (AJCC stage I) vs advanced stage (AJCC stage III and higher) with the type of drift pattern (group H vs group L + I). Specifically, low-stage tumors are more prevalent in group H compared with group L + I among the 74 TCGA EAC for which tumor stage information was available (odds ratio 6.0 (1.1–63.3)).
In particular, the gene most significantly repressed in TCGA (Additional file 2: Table S1) were found repressed in the smaller study by Krause et al. [21], and most recently as a significantly silenced gene in a large clustering analysis of esophageal adenocarcinoma [22]. This striking asymmetry between gene expression changes and methylomic drift is consistent with parallel findings that CpG promoter hypermethylation in cancers often is correlated with gene-silencing [6]. A Gene Ontology (GO)-based over-representation analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID) shows a highly significant greater than threefold enrichment of sequence-specific DNA binding transcription factor activity (\( p = 2 \times 10^{-9} \), Additional file 2: Table S2). The most prominent group identified by this analysis is a family of repressive Krueppel-associated box (KRAB) domain zinc finger (ZNF) transcription factors (greater than sixfold, \( p = 1.3 \times 10^{-15} \), Additional file 2: Table S2). KRAB-mediated transcriptional repression involves the binding of the KRAB domain to co-repressors potentially resulting in heterochromatin formation and silencing of endogenous retroviruses [23, 24].

**Evidence for a threshold effect leading to bimodal drift**

The drift patterns shown in Fig. 5a for BE and EAC samples suggest nonlinear drift dynamics in BE tissues. Specifically, the presence of a persistent mode in the drift distribution at low levels (\( \beta < 0.2 \)) is indicative of a threshold below which drift is suppressed but advances rapidly once the mean level is surmounted. To validate that epigenetic drift occurs in our longitudinal samples (20 patients with 2 biopsies each separated by at least 3–4 years), we determined for each individual at two time points the number of drift CpGs that remained below \( n_{11} \), respectively, the number that remained above \( n_{22} \), and the number of drift CpGs that had crossed the threshold from low to high at \( \beta = 0.2 \) (\( n_{12} \)) and, vice versa, from high to low (\( n_{21} \)). The results, including the % fraction of drift CpGs advancing, \( n_{12}/(n_{12} + n_{11}) \), and the % fraction retarding between the two time points, \( n_{21}/(n_{21} + n_{22}) \), are listed in Additional file 2: Table S3 and shown as annual rates in Fig. 6. For 19/20 patients, we detect greater methylation flow from sub-threshold levels to higher levels at the second (later) biopsy compared to flow in the opposite direction.

We note that these findings are surprisingly consistent with the unimodal-to-bimodal epigenetic drift predictions made by Sontag et al. [8] who proposed a mathematical model that included a nonlinear relationship between de novo methylation and the ambient level of methylation present in a region of CpGs. To demonstrate that such a model results in unimodal-to-bimodal drift transitions over
time, we explicitly simulated sporadic de novo methylation on an island of 50–100 CpGs, independently in 1000 cells, mimicking crudely the cell population in the tissue samples. Each CpG was assumed to be in a binary state (0/1 of being (un)methylated), and the states of the CpGs initially (at time $t = 0$) were sampled from a binomial distribution with probability 0.06 which equals the mean methylation level in our NS tissue samples. The CpG states were then propagated stochastically with a rate (probability per time step) of becoming methylated that increases 100-fold from a background of $10^{-4}$ to $10^{-2}$ when the mean level of methylation on the island crosses a threshold of $\beta = 0.2$.

Without a mathematical exploration of this Markov model, but straightforward in silico experimentation with the baseline distribution of methylation rates (specifically, a gamma distribution with mean $10^{-4}$ and variance $4 \times 10^{-8}$) and threshold value, our simulations show that this simple model generates methylation density trajectories that typically bifurcate and strikingly resemble the observed drift signatures in our samples. Figure 5b shows a typical density trajectory for a region of 50 CpGs, an arrayed population of 1000 cells, every 100 time steps, for a total duration of 3000 time steps. Although our model differs in functional form from the model described in Sontag et al., it shares important features, including a suppression of de novo methylation at low levels and a nonlinear acceleration as the ambient (regional) level of methylation increases. In contrast, models that do not include this ambient methylation feedback on the local (site-specific) rate of methylation do not, in general, lead to bifurcations in the main (initial) mode of the evolving drift pattern, but still exhibit a weak bimodality as shown in Additional file 1: Figure S2. Additional file 1: Figure S3 further illustrates the stochastic nonlinear behavior of our model via simulated time course trajectories of the mean methylation levels for 10 CpG islands that share an identical drift rate distribution across their CpGs.

**Discussion**

Here we take a closer look at how differential epigenetic drift is organized in BE-associated genomes, and its scope and association with gene expression, motivating further investigation of its role in neoplastic progression in BE. To do so, we first surveyed the array-based DNA methylome for significant correlations with the mean drift measured by 67 drift CpGs previously identified by our group to estimate BE dwell time, i.e., the time a patient has lived with BE [12, 13]. Following this study, we targeted CpGs that are hypomethylated in NS tissue but are subject to differential drift in BE tissue caused by accelerated age-related de novo methylation. While NS tissue may not be the tissue of origin for BE, the similarity of methylation levels at drift-associated CpGs between NS and other normal tissues, such as fundus (see Additional file 1: Figure S1), justifies the use of NS as a normal reference tissue to identify differential drift in BE. Our previous study did not reveal the full extent of this differential drift due to highly restrictive pre-filtering.

Our genome-wide “drift survey” revealed that, at the island level, >24% of CpG islands undergo methylomic drift and are predominantly promoter-associated (i.e., overlap transcription start sites (TSS)). To investigate whether epigenetic drift occurs on isolated CpG sites or is a nonlocal phenomenon at the CpG island level, we evaluated correlations of methylation between pairs of CpGs (across BE samples) using all island-associated CpG probes (available on the HM450 platform) as a function of genonic distance between the probes (Fig. 4). Our results confirm the prevailing view that CpG islands essentially exert epigenetic control by their collective methylation state rather than through specific CpG sites [25, 26]. Importantly, we found evidence that drift does not evolve uniformly in BE and EAC but appears to be governed by a nonlinear, threshold-like stochastic methylation process which depends nonlocally on the methylation status of other island CpGs. Simulations using a stochastic model, which reflects these dynamics at the island level, show characteristic transitions from unimodal to bimodal drift similar to what we observe in our data. Although other models may provide similar fits to the observed drift distributions, this model has its origins in earlier work aimed at understanding the stable, somatic inheritance of methylation imprints [8] and predicts epigenetic drift as a series of sporadic de novo methylation events at the island level. Our nonlinear feedback model for methylomic drift suggests that the various drift distributions we see in our tissue samples may simply be attributed to tissue age itself (i.e., at what
Conclusions

Our results are consistent with the hypothesis that epigenetic drift heralds the onset of (epi)genomic instability via bifurcations (as seen in Fig. 5) that associate with the transcriptional repression of important regulatory genes [29–32]. Thus, under this hypothesis, epigenetic drift not only defines tissue aging (i.e., provides a molecular clock) but also “throttles” the expression and function of developmental genes forcing transitions in tissue characteristics that better cope with the erosive and damaging milieu in BE. Further studies of whether changes in methylyomic drift simply reflect transcriptional changes during neoplastic progression or induce such changes are therefore of critical importance to better understand mechanisms that drive age-related cancer evolution.

Methods

Tissue samples

Formalin fixed paraffin-embedded (FFPE) tissue slides and cores were obtained from Case Western Reserve University/University Hospitals of Cleveland (Cleveland, OH) and the Cleveland Clinic (CC) following protocols approved by the Institutional Review Board of each institution. For the cross-sectional analysis, we used HM450 methylation array data from 52 NS, 64 nondysplastic BE, and 24 EAC samples through the Barrett’s Esophagus Translational Research Network (BETRNet) [33]. For the longitudinal drift analysis, we utilized 33 additional tissue samples from two studies with 10 patients each (CC and CW). Each patient had two biopsies separated by at least 3–4 years (40 samples total). Of these, seven samples were included in the cross-sectional analysis. See Additional file 2: Table S4 for relevant clinical information on the patient samples used in this study.

Sample pre-processing

Tissue sample preparation and DNA extraction were performed as described previously [34]. The quality of DNA extracted from FFPE samples were determined with Illumina HD FFPE QC assay (Illumina, San Diego, USA) following the manufactures’ instructions. Two hundred fifty nanograms of DNA samples that passed the QC assay were bisulfite converted using the EZ DNA Methylation Kit (Zymo Research, Irvine, USA). DNA restoration was performed using the Illumina HD FFPE Restoration Kit (Illumina, San Diego, USA) according to the manufacturer’s instructions. Intermediate DNA purifications were performed using the Zymo DNA Clean and Concentrator-5 Kit (Zymo Research, Irvine, USA). The BETRNet DNA samples were run on Illumina HumanMethylation450 BeadChip (HM450) arrays following the manufacturer’s instructions.
(Illumina Inc.) at the Fred Hutch Genomics Core facility. Data were then accessed as raw two-color channel intensities in idat format and pre-processed. Arrays were normalized using two functions implemented in the minfi (v1.18.6) R module, including an initial background intensity correction identical to the correction implemented in Illumina’s Genome Studio software, followed by subset quantile within-array normalization (SWAN) to harmonize data across assay design types [35, 36]. Probes showing mean detection \( p \) value > 0.05 were filtered out. Furthermore, we checked for the presence of previously identified cross-reactive CpGs in our drift CpG sets. Our drift CpG sets are uniformly under-enriched for cross-reactive probes, and we found that the presence of cross-reactive probes did not affect the integrity of our findings.

**Methylation and gene expression datasets**

Matched methylation (HM450 platform) and gene expression (Illumina HumanHT-12V4.0 expression BeadChip platform) data collected for 4 BE and 47 EAC, and 17 normal esophagus tissue samples published by [16] were accessed via the Gene Expression Omnibus (GEO) online repository (Series GSE72874). Methylation and expression BeadChip array data were obtained as normalized and pre-processed intensity counts or \( \beta \) values and prepared as described in [16].

Additional validation data, including HM450 array methylation and Illumina HiSeq 2000 RNA Sequencing data from the Version 2 analysis pipeline, were obtained for samples provided by the Cancer Genome Atlas (TCGA) via the NCI Genomic Data Commons [37] and the Firehose resource hosted by the Broad Institute (http://gdac.broadinstitute.org/). HM450 array data were obtained as raw two-color channel intensity readings, which were subjected to the same pre-processing pipeline as the BETRNet cohort data, described above. RNA-seq expression sequencing data was obtained as level 3 RNAseq by expectation maximization (RSEM)-normalized and pre-processed intensity counts [38]. A nonparametric Mann-Whitney-Wilcoxon (MWW) \( q \) test was applied to gene-specific count data to detect differential gene expression between low methylation samples (\( \beta < 0.2 \)) and advanced methylation samples (\( \beta \geq 0.2 \)).

**Quantification of drift**

The methylation state of a CpG dinucleotide on a specific chromosome is essentially a binary variable; the cytosine is either methylated or unmethylated. However, DNA methylation arrays (such as the Illumina HM450 beadchip) provide only aggregate measurement across thousands of cellular epigenomes in a given tissue sample and therefore can only provide population fractions (i.e., \( \beta \) values) of methylated probes expressed as the ratio \( \beta = M/(M + U) \), with \( M \) and \( U \) representing the number of methylated and unmethylated probes in the sample, respectively.

Genome-wide differential epigenetic drift in BE was quantified by scanning over 146,000 hypomethylated CpG probes (\( \beta < 0.25 \) in NS tissue) on the HM450 platform for significant correlations with the mean methylation levels of 67 CpGs previously identified to drift differentially between BE and matched NS tissue samples from 30 BE patients [12]. Note, the differences in mean \( M \) values (defined as logit\( 2(\beta \) value)) of the 67 drift CpGs reflect patient-specific differences in individual BE tissue dwell times as described in [12]. Figure 7 illustrates the two-step method used to identify CpG probes that were significantly correlated with this BE tissue clock: (1) we computed the mean \( M \) value drift over the 67 BE clock probes for each of the 64 cross-sectional BE samples and (2) for each CpG in the hypomethylated test set (146,029 CpG probes), we obtained the Pearson correlation and \( p \) value using the cor.test R-function. Only CpG probes that were significantly (\( q < 0.01 \)) and positively (\( r > 0.5 \)) correlated with the BE tissue clock were retained and formed the set of 18,013 island and non-island-based drift CpGs used in this study.

**Statistical analysis and visualization**

All data pre-processing and the majority of statistical testing was performed in R programming language with base R graphics and analysis functions (v3.3.0). The minfi (v1.18.6) and GEOquery (v2.38.4) Bioconductor modules were used to access, pre-process, normalize, and analyze both methylation and gene expression array data, respectively [35, 39, 40].

**Data access**

Data prepared for this study are available online at the GEO website (Series Number: GSE104707). Scripts for study analyses and visualizations are available at https://github.com/gluebeck/Scope-of-methylomic-drift-in-BE.

**Fig. 7** See “Methods"
Additional files

Additional file 1: Figure S1. Boxplot of normal squamous (NS) methylation fine-structure (represented using M values) for five representative, consecutively positioned CpGs at the MGMT (O-6-methylguanine-DNA methyltransferase) gene which overlaps a CpG-rich island at chr10:131264948-131265710. Mean methylation fractions (n = 52) range from 1% (lowest) to 19% (highest) for the five promoter-associated CpGs shown. Superimposed are the M values of 12 normal tissue samples collected in fundus (red). Nearly identical methylation patterns were observed in normal colon samples (not shown). Figure S2. Simulated methylation densities (arbitrary time scale) using a linear drift model without ambient methylation feedback on the rate of site-specific methylation. Figure S3. Simulated trajectories of mean methylation levels for 10 islands with 50 CpGs each under the nonlinear (threshold) model described in the main text. As methylation levels approach the threshold of β = 0.2, rapid stochastic transitions occur followed by accelerated drift. Figure S4 Karyograph showing locations of methylomic drift across 64 BE samples for all 22 autosomes. Figure S5. The same as Fig. 5a, but for 87 EAC from TCGA. (DOCX 863 kb)

Additional file 2: Table S1. List of genes with differential expression among low- and high-drift samples in TCGA (n = 87) using a β value threshold of 0.2 to delineate the two groups. Two hundred genes were significantly underepressed in the advanced drift group, 15 genes (not shown) were significantly overexpressed (q < 0.01, Mann-Whitney-Wilcoxon two-sided test). Highlighted genes were also found to be independently and significantly underepressed in the combined set of 47 EAC and 4 BE samples for which both gene expression and DNA methylation data were available used by Krause et al. (Carcinogenesis 37(4), 2016). Table S2, DAVID enrichment analysis (by protein class) of the 200 repressed genes listed in Additional file 2: Table 1. Highlighted protein classes are significantly enriched. Table S3 CpG dinucleotide methylation transition rates for 20 patients with longitudinally collected BE biopsy samples separated by at least 3–4 years, including 10 patients from BETRNet/CW and 10 patients from BETRNet/CW. A threshold of β = 0.2 was used to classify CpG methylation as low (1) or high (2). The first two columns provide patient ages at biopsy, third is a patient label, columns 4–7 represent CpG fractions that begin and end at low methylation (n₀), transition from high to low (n₁), transition from low to high (n₂) and remain high (n₃). Conditional transition fractions are in columns 8–11, and annual increasing and decreasing methylation rates are in columns 12–13. Table S4. Patient ID (encoded), project (BETRNet/MEMO), tissue type (normal squamous (NS), Barrett’s esophagus (BE), esophageal adenocarcinoma (EAC)), sex, age at biopsy, and patient diagnosis (Dx) at the time of biopsy. (DOCX 723 kb)

Acknowledgements
We thank Prof Trevor Graham (Barts Cancer Institute, London, UK) and Dr. Hamid Bolouzi (Fred Hutch) for their helpful comments. This study was supported NIH grant UO1CA182940 (EGL, WDH, WMG, SKM, KC), NIH grants F50CA150964, U54CA163060 and P30CA43703 (WMG, JEW, AC); NIH grants U01CA152756, P30CA015704, and U01CA086402 (SKM, YM, WMG); the DeGirogio Family and Price Family Foundation (WMG, EGL); and Barts Charity, London (KC).

Authors’ contributions
PNT, DTP, AC, JEW, and WMG contributed the biospecimens and materials for this study, DNA processing and HM450 arrays were performed by MY and WMG, bioinformatic analyses were performed by SKM, WDH and EGL, and authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interest.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
1Program in Computational Biology, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. 2Centre for Tumour Biology, Barts Cancer Institute, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK. 3Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. 4Department of Gastroenterology, Digestive Disease & Surgery Institute, Cleveland Clinic, Cleveland, OH 44195, USA. 5Department of Pathology, Cleveland Clinic, Cleveland, OH 44195, USA. 6University Hospitals Case Medical Center, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA. 7Department of Medicine, University of Washington School of Medicine, Seattle, WA 98195, USA.

Received: 25 May 2017 Accepted: 24 September 2017
Published online: 16 October 2017

References
1. Campsi J, Aging, cellular senescence, and cancer. Annu Rev Physiol. 2013; 75:685–705. https://doi.org/10.1146/annurev-physiol-030212-183653. PubMed PMID: WOS:000316381400031

2. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell 2013;49(2):359–67. Epub 2012/11/28. doi: https://doi.org/10.1016/j.molcel.2012.10.016. PubMed PMID: 23177740.

3. Horvath S. DNA methylation age of human tissues and cell types. Genome biology. 2013;14(10). doi: Attn R115. Doi https://doi.org/10.1186/Gb-2013-14-10-R115. PubMed PMID: WOS:000323987500008.

4. Alisch RS, Barwick BG, Chopra P, Myrick LK, Satten GA, Conneely KN, et al. Age-associated DNA methylation in pediatric populations. Genome Res. 2012;22(4):623–32. https://doi.org/10.1101/gr.125187.111. PubMed PMID: WOS:000302203800004.

5. Ahuja N, Li Q, Mohan AL, Baylin SB, Issa JP. Aging and DNA methylation in colorectal mucose and cancer. Cancer Research. 1998;58(23):5489–94. PubMed PMID: WOS:000077343000042.

6. Toyota M, Sasaki Y, Satoh A, Ogi K, Kikuchi T, Suzuki H, et al. Epigenetic inactivation of CHFR in human tumors. Proc Natl Acad Sci USA. 2003; 100(13):7818–23. https://doi.org/10.1073/pnas.1337066100. PubMed PMID: WOS:000183845800074.

7. Issa JP, Ahuja N, Toyota M, Bronner MP, Brentnall TA, Accelerated age-related CpG island methylation in ulcerative colitis. Cancer Research 2001; 61(9):3573–7. Epub 2001/04/28. PubMed PMID: 11352821.

8. Sontag LB, Lorincz MC, Georg Luebeck E. Dynamics, stability and inheritance of somatic DNA methylation imprints. J Theor Biol. 2006;242(4):890–9. Epub 2006/06/30. doi: https://doi.org/10.1016/j.jtbi.2006.05.012. PubMed PMID: 16806276.

9. Shibata D. Mutation and epigenetic molecular clocks in cancer. Carcinogenesis. 2011;32(2):123–8. https://doi.org/10.1093/carcin/bqz239. PubMed PMID: WOS:000286676400001.

10. Issa JP. Aging and epigenetic drift: a vicious cycle. J Clin Invest 2014;124(1): 24–9. Epub 2014/01/03. doi: https://doi.org/10.1172/JCI69735. PubMed PMID: 24382386; PubMed Central PMCID: PMC3871228.

11. Teschendorff AE, West J, Beck S. Age-associated epigenetic drift: implications, and a case of epigenetic theft?. Hum Mol Genet. 2013;22(R1): R7-R15. Epub 2013/08/07. doi: https://doi.org/10.1093/hmg/ddt375. PubMed PMID: 23918660; PubMed Central PMCID: PMC3782071.

12. Curtius K, Wong CJ, Hazeltown WD, Kaz AM, Chak A, Willis JE, et al. A molecular clock infers heterogeneous tissue age among patients with Barrett’s esophagus. PLoS Comput Biol. 2016;12(5):e1004919. https://doi.org/10.1371/journal.pcbi.1004919. PubMed PMID: 27168458; PubMed Central PMCID: PMC4864310.

13. Hazeltown WD, Curtius K, Inadomi JM, Vaughan TL, Meza R, Rubenstein JH, et al. The role of gastroesophageal reflux and other factors during progression to esophageal adenocarcinoma. Cancer Epidemiol Biomarkers Prev. 2015. doi: https://doi.org/10.1158/1055-9966.EPI-15-0323-T. PubMed PMID: 25931440.

14. Kong CY, Kroep S, Curtius K, Hazeltown WD, Jeon J, Meza R, et al. Exploring the recent trend in esophageal adenocarcinoma incidence and mortality using comparative simulation modeling. Cancer Epidemiol Biomarkers Prev.
32. Quail T, Shrier A, Glass L. Predicting the onset of period-doubling bifurcations in noisy cardiac systems. Proc Natl Acad Sci U S A. 2015;112(30):9358–63. doi: https://doi.org/10.1073/pnas.1424320112. PubMed PMID: WOS:000358655600066

33. Abrams JA, Appelman HD, Beer DG, Berry LD, Chak A, Falk GW, et al. Barrett’s Esophagus Translational Research Network (BETRNet): the pivotal role of multi-institutional collaboration in esophageal adenocarcinoma research. Gastroenterology. 2014;146(7):1586-90. Epub 2014/04/29. doi: https://doi.org/10.1053/j.gastro.2014.04.014. PubMed PMID: 24768332.

34. Luo Y, Wong CJ, Kaz AM, Dzieciatkowski S, Carter KT, Morris SM, et al. Differences in DNA methylation signatures reveal multiple pathways of progression from adenoma to colorectal cancer. Gastroenterology. 2014. Epub 2015/06/06; https://doi.org/10.1053/j.gastro.2014.04.039. PubMed PMID: 249739120.

35. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363–9. https://doi.org/10.1093/bioinformatics/btu049. PubMed PMID: WOS:000336530000004

36. Makimovic J, Gordon L, Oshlack A. SWAN: subset-quantile within array normalization for Illumina Infinium HumanMethylation450 BeadChips. Genome biology. 2012;13:6. doi: https://doi.org/10.1186/Gb-2012-13-6-R44. PubMed PMID: 21709354.

37. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for cancer genomic data. N Engl J Med. 2016;375(12):1109–12. https://doi.org/10.1056/NEJMoa1607591. PubMed PMID: WOS:000385371000002

38. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics. 2011;12. doi: https://doi.org/10.1186/1471-2105-12-323. PubMed PMID: WOS:000294361700001.

39. R Core Team. R: a language and environment for statistical computing. Austria: R. Foundation for Statistical Computing; 2013. Available from: http://www.R-project.org.

40. Sean D, Meltzer PS, GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. Bioinformatics. 2007;23(14):1846–7. https://doi.org/10.1093/bioinformatics/btm254. PubMed PMID: WOS:000249248300022