Evolution and progression of Barrett's oesophagus to oesophageal cancer

Sarah Killcoyne[1,2] *and Rebecca C. Fitzgerald*[1] ✉

Abstract | Cancer cells are shaped through an evolutionary process of DNA mutation, cell selection and population expansion. Early steps in this process are driven by a set of mutated driver genes and structural alterations to the genome through copy number gains or losses. Oesophageal adenocarcinoma (EAC) and the pre-invasive tissue, Barrett's oesophagus (BE), provide an ideal example in which to observe and study this evolution. BE displays early genomic instability, specifically in copy number changes that may later be observed in EAC. Furthermore, these early changes result in patterns of progression (that is, 'born bad', gradual or catastrophic) that may help to describe the evolution of EAC. As only a small proportion of patients with BE will go on to develop cancer, a better understanding of these patterns and the resulting genomic changes should improve early detection in EAC and may provide clues for the evolution of cancer more broadly.

Tumour genomes are the result of mutational and evolutionary pressures on somatic cell populations. These can be caused by erroneous endogenous processes, including DNA replication repair errors, and exogenous carcinogenic insults. Recent pan-cancer analyses have suggested that early oncogenic mutations may occur years or even decades before invasive disease is detected[1].

Alterations to the genomic structure generate mutational diversity in cellular populations and are known to occur in organisms, although major structural alterations to the genome are often embryonically lethal. In cancers, however, this genomic instability provides an additional mechanism for tumorigenesis and clonal evolution within cell populations[2,3]. Somatic copy number alterations (CNAs), ranging from small focal gains or losses affecting single driver genes to whole chromosome gains and losses, often occur alongside single point mutations, resulting in a complex spectrum of genomic events leading to tumorigenesis. Evidence from bulk tumour sequencing projects suggests that the spectrum of mutational events from point mutations to structural variation differs widely between cancer types. Oesophageal adenocarcinoma (EAC) displays one of the highest single-base mutational burdens at 8–10 mutations per megabase (mut Mb^{-1})[4,5] alongside high rates of complex structural variations, deletions and tandem duplications[2,6,7].

Histologically normal ageing tissues[8–10] and blood[11] display many of these early driver single-base mutations in what has been described as a patchwork of clonal populations. Allelic imbalances in chromosome arms have also been detected in histologically normal oesophagus alongside one of the highest rates of point mutations[12]. These studies make clear not only that oesophageal tissues accumulate mutations, chromosomal changes and various clonal cell populations as we age but that they do so in multiple clonal populations of cells. The majority of these mutations have no functional consequence, making these 'passenger' mutations unlikely to drive cancer development[13]. Instead, for these cells to develop into cancer the acquired mutations must be positively selected and lead to clonal expansion[14]. This early genomic instability prior to invasive stages of tumorigenesis may help drive the development of disease[15] by providing phenotypic diversity for selection to act upon.

Precancerous tissues display histological and molecular changes that both distinguish them from normal tissues and are associated with an increased risk of cancer. Barrett's oesophagus (BE) is a clear example of a precursor tissue histologically, clonally and genetically. It presents histologically as a mosaic of intestinal and gastric metaplasia[16] in which columnar cells replace squamous epithelia in the normal oesophagus (FIG. 1). This is presumed to occur to compensate for the effects of gastrointestinal reflux entering the oesophagus[17], but unfortunately this adaptation increases the risk that a patient may develop EAC. EAC is a distinct subtype of oesophageal cancer that has seen a rapid increase in incidence in many western countries[18], with a corresponding decrease in the rate of oesophageal squamous cell carcinoma, which continues to be the dominant subtype worldwide[19].

The lowest-risk BE tissues are non-dysplastic Barrett's oesophagus (NDBE), and patients diagnosed with this

[1]*Medical Research Council Cancer Unit, Hutchison/Medical Research Council Research Centre, University of Cambridge, Cambridge, UK.*

[2]*European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK.*

✉*e-mail: RCF29@MRC-CU.cam.ac.uk*

https://doi.org/10.1038/s41568-021-00400-x

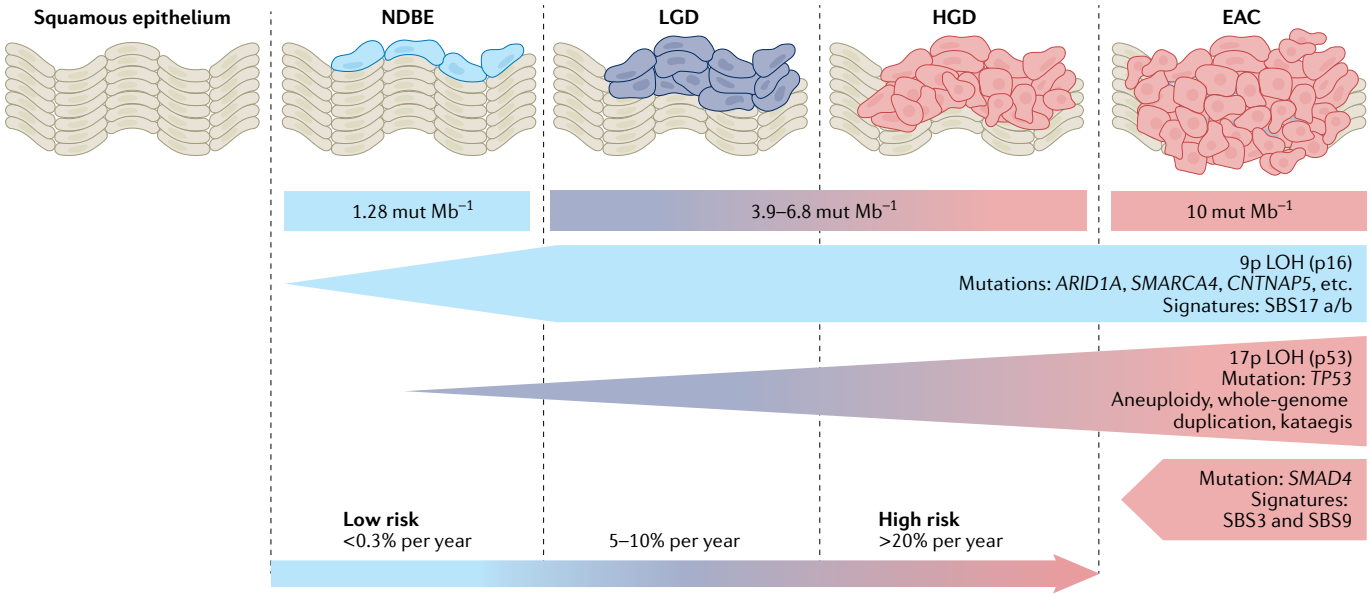


Fig. 1 | **Histopathological progression for BE.** Progression from squamous epithelia to non-dysplastic Barrett's oesophagus (NDBE), low-grade dysplasia (LGD), high-grade dysplasia (HGD) and oesophageal adenocarcinoma (EAC). Most Barrett's oesophagus (BE) will not progress to HGD or EAC but will display specific molecular hallmarks, from an increased mutation rate and SBS17a/b mutational signature to mutations in canonical cancer driver genes (such as *ARID1A*, *SMARCA4* and *CNTNAP5*) and 9p loss of heterozygosity (LOH). Patients destined to progress may display NDBE, LGD and/or HGD histology prior to EAC progression. Molecular hallmarks identified in samples from patients who later progress, independent of histology, will include the molecular changes seen across all grades of disease as well as genomic instability, high clonality and *TP53* mutations. EAC may include any or all of these molecular hallmarks, additional mutational signatures (for example SBS3 and SBS9) and additional mutations (for example in *SMAD4*). Blue indicates a lower risk of progression, purple indicates an increased risk and red indicates high risk. Specific risks for progression to EAC cited at the bottom are from Hvid-Jensen et al.[20]. Width of the shapes describing molecular alterations indicates how commonly each occurs at each stage. mut Mb^{-1}, mutations per megabase. SBS, single-base substitution.

grade have a risk of progression to EAC of approximately 0.3% per year based on epidemiological data[20]. As low-grade dysplasia (LGD) or high-grade dysplasia (HGD) develops, the risk of progressing to EAC increases[21], and although this is often presented as a linear progression sequence, the time it takes to progress between steps or across patients can vary dramatically[22]. Additionally, although BE is the only known precursor lesion for EAC, at least one half of patients diagnosed with EAC do not have detectable BE[23], and the majority of patients with EAC (more than 90%) lack a prior BE diagnosis[24]. It is currently debated whether all EAC arises from BE, as it is possible that the metaplastic precursor has been overgrown by tumour or, alternatively, may have arisen through another pathway (BOX 1).

This Review focuses on BE, as patient surveillance has provided clinicians and researchers with valuable longitudinal and spatial genomic evidence to better understand the early evolution of cancer. The recent completion of pan-cancer analyses by the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA)[1,6,7] has provided both new tools for investigating the patterns of evolution within somatic genomes and new understanding of the timescales in which mutation events occur within a single patient (relative to their age) and, more generally, between different cancer types (for example, *APC* and *KRAS* in colon cancer, and *EGFR* (which encodes epidermal growth factor receptor) and *TP53* (which encodes p53) in glioblastoma). These timescales for progression provide good evidence that precancerous tissues such as BE contain molecular signatures of cancer risk. In this Review we discuss the current understanding of the genomic and evolutionary landscape of EAC and its precursor lesion BE, with a specific focus on how this evolutionary view of BE can be used to improve early diagnostics.

Molecular hallmarks of BE and EAC

Large-scale genomic profiling studies of EAC performed as part of the ICGC[6] and TCGA[25] reveal a cancer that has a high mutational burden, placing EAC among melanoma and lung cancer as one of the most mutated cancer types[4,5]. Despite the high mutational load, researchers have identified few recurrent point mutations in driver genes aside from the tumour suppressor *TP53* (>70% frequency), with other canonical driver genes *SMAD4* (which encodes a DNA-binding transcription factor and is part of the TGFβ pathway), *ARID1A* (which encodes a SWI/SNF chromatin remodelling complex protein) and *CDKN2A* (which encodes the cyclin-dependent kinase inhibitor p16 that regulates p53) recurrently displaying point mutations in less than 20% of cases[26,27]. Aside from point mutations, patients with EAC share hallmarks of high genomic instability defined in other tissue types[28], including frequent CNAs, aneuploidy, kataegis and whole-genome duplication. These diverse molecular hallmarks provide some insight into the early evolution of the disease.

Genetic landscape. Numerous studies have provided multiregion data on BE. Maley et al.[29] used flow cytometry and fluorescence in situ hybridization (FISH)

Kataegis
Localized hypermutation often several hundred base pairs in length.

Mutational signature
A combination of mutations (specifically single base-pair substitutions) that generates a specific pattern, or signature, relating to specific mutational processes.

Minor-in groove
DNA facing in to the histone core (minor-in).

on multiple biopsy samples in patients with BE to assess different cellular clones, whereas Martinez et al.[30] used whole oesophagus brushings as well as separate biopsy samples to assess genetic and clonal diversity. Our recent work[31] also utilized separately sequenced biopsy samples to assess a genetic risk, and Ross-Innes et al.[5] sequenced paired EAC and adjacent BE tissues in patients.

Sequencing BE adjacent to EAC showed that the BE tissues, regardless of dysplasia, have a mutational burden (6.76 mut Mb^{-1}) that is only slightly lower than the median burden for EAC and higher than a majority of cancers in pan-cancer studies[32]. Sequencing tissues from patients without EAC showed that dysplastic BE samples have a similarly high mutational rate of 3.9 mut Mb^{-1}, whereas non-dysplastic BE may be lower at 1.28 mut Mb^{-1} (REF.[33]). It should be noted, however, that BE tissues are a patchwork of clones[5,30,34] similar to normal epithelial tissues. This clonal patchwork makes it likely that mutation rates, even in non-dysplastic BE, are underestimated. As dysplasia and, eventually, EAC develops, one or more of these clones will expand, increasing our ability to detect it and better estimate the mutation rate. The relatively high mutation rates throughout the histological progression, as well as the multiple clones throughout the tissue, suggest a rich environment for clonal selection, ongoing mutation and eventual tumorigenesis.

Mutational signature analyses in EAC have consistently identified a spectrum of single-base substitutions (SBSs) characterized by T > G or T > C substitutions in a CTT context[5,35–38], typically labelled SBS17a/b (REF.[7]). Although this signature has been linked to the use of 5-fluorouracil chemotherapy and its derivatives[39], the aetiology of this signature in BE and EAC in the absence of chemotherapy exposure is still unclear. SBS17a/b has also been identified as a dominant pattern in BE across both dysplastic and non-dysplastic tissues[5,33], indicating that the underlying mutational process occurs early and is potentially active throughout the tissue. As BE is thought to arise owing to chronic acid reflux injury to the oesophagus, it has been speculated that these signatures are related to ongoing oxidative damage[40]. Although this has not been confirmed in BE or EAC specifically, mutations related to oxidative damage and SBS17a/b appear to be more frequent than expected at the minor-in groove of nucleosomes. It is speculated that this could result in an increased mutation rate owing to poor accessibility of DNA damage repair machinery and some protection from selection pressures[41,42]. Overall, the dominant SBS17a/b signatures in BE may help to explain the high mutational burden of EAC in the absence of clear exogenous mutagenic exposures (for example, tobacco and ultraviolet light)[7,32]. However, as gastric cancers, which also display a high proportion of SBS17a/b-related mutations, have a lower mutation rate (estimated as around 5 mut Mb^{-1} (REF.[43])), other processes are likely to be involved.

Despite the high mutation rate and consistent mutational signatures identified throughout BE and EAC, there appear to be only a few early driver events identified in tumours. In the pan-cancer analysis by Martincorena et al.[44] EAC clearly showed patterns of mutations in cancer genes that were under positive selection, indicating that those genes would be driving the disease. This analysis identified relatively few recurrent driver events in EAC, although a subsequent analysis that focused entirely on EAC confirmed that, similar to other cancer types, EAC displays an average of four driver events per tumour[26]. As these events, by definition, should occur early to allow for clonal expansion, progression may also rely on other drivers not yet identified, such as epigenetic alterations or changes within the immune landscape of BE (BOX 2).

Using an EAC-derived panel of 26 driver genes most commonly identified from tumour samples, Weaver et al.[37] tried to identify the early mutations in BE that related to later EAC development by sequencing samples from different histological stages: EAC ($n = 112$), HGD ($n = 43$) and NDBE with no evidence of progression in multiple years of follow-up ($n = 66$). They found that mutations in a number of these genes such as *SMARCA4* (which, similar to *ARID1A*, encodes a SWI/SNF chromatin remodelling complex protein), *ARID1A* and *CDKN2A* could be identified at every histological stage from NDBE to EAC and that there was no significant difference detected in the frequency of mutations between the different stages. Only *SMAD4* and *TP53* were found to differ in frequency between dysplasia and NDBE. Whereas Weaver et al. found that mutations to *SMAD4* appeared to be exclusive to a small proportion of EAC cases (13%), *TP53* mutations were found in patients who progress to cancer by both Stachler et al.[45] and Ross-Innes et al.[46], including in some cases of

Box 1 | BE cell of origin

The concept of clonal expansion relies on an originating cell that can expand into a clonal population retaining early mutations, and later develop mutations that may expand into new subclonal populations. In Barrett's oesophagus (BE) there have been several hypotheses regarding a cell of origin that could give rise to intestinal metaplasia from a predominantly squamous epithelial background[107,108]: transdifferentiation of squamous epithelial cells, reprogramming them into columnar cells; or expansion of other progenitor cells within the oesophagus or gastric cardia. In either case, the cell of origin for BE would need to undergo a phenotypic change to acquire intestinal characteristics.

In the transdifferentiation model, alterations to various signalling pathways and gene expression patterns in squamous epithelium are presumed to be driven by the acidic conditions of reflux. In mouse models, overexpression of an intestinal transcription factor *Cdx2* in squamous cells resulted in a transitional cell type with glandular features[109]. In vitro studies using bone morphogenetic protein 4 (BMP4) induction also resulted in a phenotypic shift to a columnar type, most likely through *Cdx2* activation[110], although both models appeared to be more gastric than intestinal.

Alternatively, stem-like cells contained within the oesophagus in submucosal glands may be reprogrammed by reflux injury to differentiate into intestinal epithelia within the oesophagus through the interleukin-1β (IL-1β)–IL-6 signalling cascade and Notch signalling[111,112]. Other progenitors have also been suggested to arise from the gastro-oesophageal junction and may suggest that BE and oesophageal adenocarcinoma (EAC) are gastric in origin[113,114].

Although the various models of the origin of BE have significant differences, fundamentally the evolutionary mechanisms follow the same pattern. A phenotypically altered cell, regardless of its tissue origin, is selected for owing to the acidic environment created by reflux within the oesophagus. This cell has an advantage over the squamous epithelia surrounding it, and can expand clonally. It is also possible, given the evidence of multiple clones within BE and the existence of mutations between adjacent BE and EAC that are not shared[30,34], that multiple cells of origin could arise within the oesophagus.

Box 2 | **Inflammation as an early driver**

The tumour microenvironment in oesophageal adenocarcinoma (EAC) is thought to be shaped by the immune inflammatory response to chronic reflux within the oesophagus[115]. Increased recruitment of inflammatory immune cells and secretion of pro-inflammatory and anti-inflammatory cytokines may occur in both EAC and Barrett's oesophagus (BE)[116]. Whereas EAC displays mixed T helper 1 (T_H1) cell-type and T_H2 cytokine profiles, BE is primarily characterized by T_H2 cytokines[117,118], which could drive early carcinogenesis.

This raises the question of the sequence of changes in the immune microenvironment of BE as it progresses to EAC. Expression profiling of specific immune cells, cytokines and chemokines across grades of BE from non-dysplastic Barrett's oesophagus (NDBE) through dysplasia to EAC was undertaken recently by Lagisetty et al.[119]. Their work showed that the genes encoding chemokine receptors such as CXCR1 and CXCR2 were upregulated in EAC and high-grade dysplasia (HGD) samples as compared with NDBE and low-grade dysplasia (LGD). Genes encoding the acute inflammatory markers interleukin-6 (IL-6) and CXCL8 were also significantly upregulated (threefold to fourfold) in HGD and EACs, whereas genes encoding cytokines involved in immune activation (CXCR3, CXCL9 and CXCL10) were not. They also showed a stepwise increase in expression of the immune checkpoint molecule PDL1 from NDBE to EAC where PDL1 was also associated with activation of other immune checkpoint pathways (including PD1, PDL2, CTLA4, ICOS and TIGIT). In addition, whereas T_H1, T_H2 and pro-B immune cell populations were increased (measured by a method called xCell from transcriptional profiling) in HGD versus NDBE samples, other cell populations (including $CD4^+$ and $CD8^+$ T cells) then decrease between HGD and EAC. Another recent study by Wagener-Ryczek et al.[120] observed only a slight increase in PDL1 expression as compared with healthy tissue. Overall, these studies appear to show early inflammatory responses in BE and a progressive change in the immune microenvironment driving proliferation as dysplasia and EAC develop.

Chemoprevention strategies in EAC focus on decreasing the inflammatory response directly through treatment with anti-inflammatory drugs (such as NSAIDs) and acid-suppression agents (such as proton-pump inhibitors). Use of NSAIDs has been linked with a lower burden of genomic alterations and decreased risk of EAC[121–123]. However, the benefit of using NSAIDs in patients with BE is less clear, as the large (2,557 patients) prospective AspECT trial found that the NSAID aspirin appeared to have no protective effect, whereas high-dose proton-pump inhibitors showed a significant difference in the time to outcome (EAC or HGD)[124].

More longitudinal studies are needed in BE to understand the role of the immune microenvironment in cancer initiation and progression, as well as the benefit to patients in preventing or decreasing the inflammatory response to reflux injury.

NDBE prior to progression to dysplasia. Investigating the specific mutations in *TP53* through paired EAC and EAC-adjacent BE samples ($n = 23$ pairs) found that BE often exhibited private mutations that were not shared by the EAC sample, and a similar pattern was also found for CNAs[46].

These findings are consistent with those described by the Big Bang model of tumour growth described in colorectal cancer (CRC)[47], which suggests that very early private mutations occur within a single clone that expands with subsequent mutations developing in subclones within the cancer. BE tissues prior to EAC development display multiple clonal populations[30,34,48], as do both primary and metastatic EAC[49,50]. These patterns of early mutations in a BE clone followed by expansion are consistent with observations in CRC where a similar histological progression is observed[51]. Single-cell and bulk genomic profiling of precancerous adenomatous polyps and CRC found that clonal mutations identified in the precancerous tissues were shared by the CRC. This analysis also showed that the precancerous tissues were found to be derived from a single founder clone that later accumulated additional mutations, generating genomic diversity[52].

Chromosomal structural instability. Sequencing studies across hundreds of patients with EAC have consistently shown a disease with a high burden of structural aberrations and CNAs. Loci that include canonical cancer genes such as *ERBB2*, *EGFR*, *GATA4*, *GATA6* and *RUNX1* have been found to be recurrently amplified or lost in EAC[26,27,53]. Genomic catastrophes such as kataegis and chromothripsis are common (frequency ≥30%), with evidence of breakage–fusion–bridge processes, focal amplifications and extrachromosomal DNA driving high CNA rates[27,36,54].

Generally, it has been assumed that large-scale genomic catastrophes resulting in sudden phenotypic changes (often termed 'hopeful monsters') are a common event in cancer but that most clones would die out as they would be selectively disadvantaged ('maladaptive monsters'). This means that few such catastrophes would lead to cancer, or be detectable. Recent investigations have uncovered CNAs occurring in histologically normal or healthy tissues[8,12,55,56] and catastrophic genomic changes resulting in aneuploidy in pre-malignant tissues[51,57,58], indicating that these 'hopeful monsters' may arise prior to even the pre-malignant stage.

In BE, such catastrophes have been well documented in dysplastic tissues, but most of these appear to be preceded by two specific events. Loss of heterozygosity (LOH) mutations in 17p and 9p, affecting *TP53* and *CDKN2A*, respectively, are common in EAC, occurring in up to 95% of cases[59,60]. Early investigations of BE tissues using flow cytometry showed that both 9p and 17p LOH occurred with increasing frequency between diploid and aneuploid cell populations[61]. Spatial flow cytometry cell sorting and targeted sequencing of BE tissues suggested that 9p LOH is an early event and that it is found across different cell populations in BE tissues[62,63] in both patients who progress and those who do not. Stachler et al.[45] reported that 37% (9/24) of patients who progress had *TP53* mutations (truncating, missense or LOH) in their NDBE segments, compared with only 5% (4/73) of patients with NDBE who do not progress. These early alterations to *CDKN2A* and *TP53* within BE are often associated with, or followed by, aneuploidy or whole-genome doubling[38,61].

Early inactivation of these tumour suppressors provides a potential mechanism for subsequent genomic instability, followed by oncogenic amplification leading to the development of EAC. Whole-genome single-nucleotide polymorphism (SNP) array profiling of 248 patients in the Seattle Barrett's Esophagus Study provided evidence that somatic CNAs were occurring across all patients with BE in the earliest biopsy samples profiled[63]. Critically, however, they showed that patients who progressed to cancer displayed more frequent and larger CNAs and that this increased over time. Similarly, an exome sequencing study of BE showed an increase in the rate of focal amplifications and deletions along the histological progression to EAC from NDBE[38].

Whole-genome duplication and general aneuploidy have long been recognized as evidence of progression to EAC. Levine et al. correlated histopathological features of dysplasia from electron microscopy with flow cytometry measurements of tetraploid fractions[64], and

Chromothripsis
Hundreds of clustered breaks occurring in a single catastrophic event affecting a limited number of chromosomes.

Breakage–fusion–bridge processes
Mechanisms of genomic instability initiated by telomeric end fusions following double-stranded breaks, which can result in repetitive cycles of fusions and breaks.

Extrachromosomal DNA
(Often extrachromosomal circular DNA). DNA found separate to the chromosomes and often contributing to higher copy numbers and altered gene expression in cancer.

Chromoplexy
Chains of rearrangements that result from the repair of double-stranded breakages.

prospective studies of patients in surveillance for BE have shown that patients with high aneuploidy and/or tetraploid cell fractions have up to 20 times greater risk of developing EAC from NDBE[58,61,65–67]. Newell et al. analysed NDBE, HGD and EAC whole-genome sequencing samples and observed evidence of catastrophic alterations including breakage–fusion–bridge processes, oscillating copy numbers, clustered structural alterations and localized regions of hypermutation (known as kataegis) exclusively in HGD and EAC samples, indicating that genomic catastrophes are more likely as a patient progresses[33].

To date, these studies show that NDBE tissues can contain some CNAs including LOH and general genomic instability. However, catastrophic structural events such as whole-genome duplications, kataegis and high rates of aneuploidy, often following the loss of *TP53* (REF.[58]), are found only in dysplastic tissues, indicating a clear molecular pattern of cancer evolution through genomic instability in precancerous tissue. The question arises of how to reconcile the variable mutational patterns observed in this disease.

Evolutionary models of progression

Evidence from metastatic EAC suggests that tumorigenesis proceeds from an early clonal expansion with related subclonal populations similar to what is proposed by the Big Bang model of CRC[47], although with an important difference. Subclones in EAC show evidence of selection pressures and spatially discrete populations[50]. The many early mutations and mutational processes operating in BE provide ample ground for this selection to occur[62], resulting in a diversity of clonal populations.

The genetic evolution from BE to EAC may be explained by the canonical methods through which populations of organisms evolve: gradualism and punctuated equilibrium. Longitudinal surveillance of patients with BE with tissue biopsies have enabled researchers to describe genomic patterns of these methods within individual patients.

Gradualism and punctuated equilibria. In describing the evolution of cancers generally, the question of gradualism versus punctuated equilibria is often raised[68]. Gradualism describes the stepwise accumulation of mutations over time with multiple clonal expansions, followed by phenotypic changes (for example, neoplastic progression). Applied to BE, this would suggest that any patient with BE is in the process of progressing to cancer through a process of stepwise mutation accumulation, and more importantly that the timing should be predictable as mutation rates are relatively stable[7] prior to a 'punctuated' or more abruptly discernible phenotypic change.

Although the histological model of BE progression (FIG. 1) displays what may be considered a stepwise progression towards EAC, the timing of these steps is highly variable between patients who do progress[22]. The evidence for gradual accumulation of mutations is similarly variable between patients over time. One confounder in this model, however, is that we can only estimate how long a patient has had BE, owing to the unquantifiable delay between the onset of BE and its diagnosis, and these estimates vary widely[69].

Furthermore, most patients with BE will never develop cancer[20], and their genomic profiles tend to display a low number of mutations that are stable between biopsy samples over time[31,63]. This observation would suggest an alternative evolutionary model: punctuated equilibrium. Under this model, pre-malignant clones may gradually accumulate mutations over time until a single clone acquires the necessary mutations to enable clonal expansion, resulting in a sudden phenotypic change (such as cancer)[70]. This pattern of progression has been described in the tumorigenesis of breast cancer[71] and CpG island methylator phenotype (CIMP)-positive CRCs[56,72–74]. Bursts of mutation from genomic catastrophes such as chromothripsis and chromoplexy may result in 'hopeful monster' clones that can be viewed as examples of punctuated equilibrium observed in cancer genomes[70,75,76]. However, it should be noted that how cancer evolution fits within the wider context of evolutionary theory is still being debated.

These evolutionary patterns can be seen in both 'born bad' BE as well as in those patients who appear to progress in a single catastrophic event[31]. Importantly, punctuated equilibria and gradualism are not mutually exclusive evolutionary pathways in BE or across the histological and molecular models described. Multiple clones can coexist over indeterminate periods of time, gradually accumulating mutations, until a mutational event triggers clonal expansion and a punctuated phenotype change. This is consistent with the observation of heterogeneity in both BE and EAC[5,38], and a pattern of genomic alterations consistent with gradual accumulation has been described in BE prior to tumorigenesis. Li et al.[63] demonstrated this using whole-genome CNA patterns at two surveillance times in 248 patients. The overall amount of CNAs detected increased between the two time points in most of the 79 patients who eventually progressed to EAC[63]. Our work confirmed that a subset of patients who develop cancer show distinct CNA patterns consistent with a gradual accumulation of mutations[31,77]. This pattern is also supported by *TP53* driver mutations detected in dysplastic tissues, with subsequent mutation of *SMAD4* in early EAC[37,38]. These mutations are consistent with a gradual accumulation of mutations over a long period of time, with an event (for example, *TP53* losses) driving a clonal expansion to tumorigenesis (FIG. 2).

Without costly and technically difficult sequencing to characterize a large percentage of the oesophagus in patients with BE, the specific clones that display early evidence of neoplastic potential, through either specific accumulations of mutations or an early catastrophic event, would be difficult or impossible to detect. Currently, researchers are only able to infer the early evolutionary patterns from the histological evidence, in the gradual detection of dysplasia in patients followed for many years or the sudden transition of patients from non-dysplasia to cancer after a period of apparent phenotypic stability, and genomically in the heterogeneity and overall high numbers of clones within the BE tissue[31,34,63]. However, given the longitudinal nature

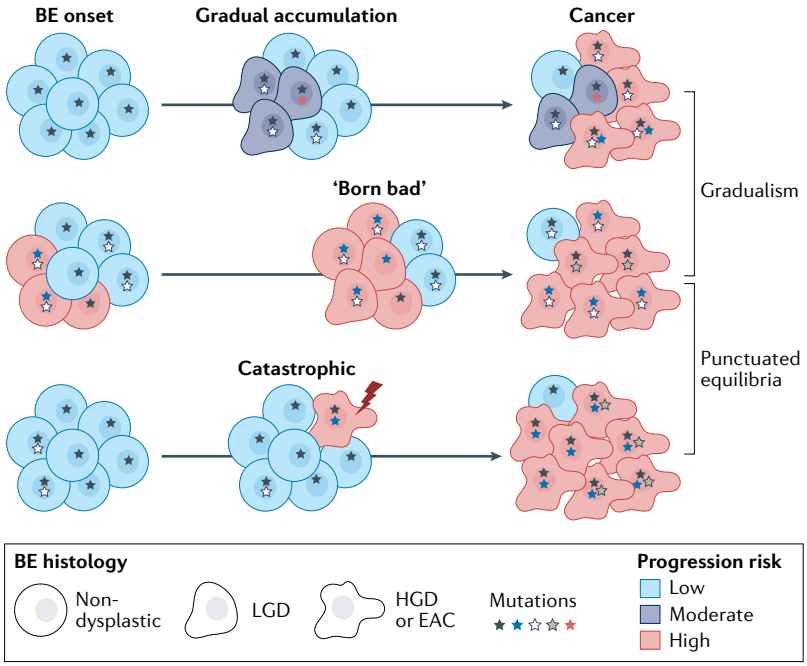


Fig. 2 | **Histological and molecular BE progression.** Three patterns of Barrett's oesophagus (BE) progression based on molecular and histopathological profiling have been observed and can be explained with the evolutionary models of gradualism and punctuated equilibria. In the 'gradual accumulation' model, BE clones acquire mutations over time in non-dysplastic tissues and later in tissues with low-grade dysplasia (LGD) before progressing to high-grade dysplasia (HGD) and early cancer; patient risk may depend on when they acquire specific mutations. The 'born bad' model proposes that BE was primed to progress to cancer even in non-dysplastic tissues and patients are at a high risk of progression even at that earliest point, whereas 'catastrophic' progression occurs in patients who display stable low-risk profiles over time. In this model, a catastrophic event suddenly alters the histology. Both evolutionary models (that is, gradualism and punctuated equilibria) can help to explain the patterns observed. A phenotypic change from molecularly stable non-dysplastic tissues to HGD or cancer is triggered in each case following clonal expansion. Histology is represented by the shape of the cell and overall risk of progressing by the colour. Representations of mutations indicating clonal relationships are shown as stars inside the cells. EAC, oesophageal adenocarcinoma.

of clinical surveillance sampling, these inferences will become more robust as more genomic and phenotypic data are generated over time and space.

Exploiting evolution in the clinic

Detecting the patterns of gradual, catastrophic or 'born bad' BE offers an alternative to the current histopathological-based approach[78,79] for quantifying patient risk in the clinic. As discussed earlier, patients with non-dysplastic BE have a very low risk of progressing to EAC, and in fact the majority of patients never will[21,80]. Current surveillance practices require patients with BE to undergo endoscopy every 1–3 years, even for patients who are not dysplastic who should be at the lowest risk of progression. Accurately stratifying high-risk and low-risk patients would enable clinicians to target surveillance and even perform earlier intervention for the highest risk patients[81,82], and decrease unnecessary procedures in the lowest risk patients.

Although the clinical benefits to improved surveillance are currently limited to the minority of patients who are diagnosed with BE, understanding the earliest molecular drivers of cancer progression may, in turn, enable the development of wider population screening approaches that utilize less invasive diagnostics (for example, Cytosponge[83] and liquid biopsy[84,85]).

Progression biomarkers. Previous studies have focused on those mutations and CNAs that were identified in both BE and EAC (TABLE 1). In the evolution of the cancer, the reasonable assumption is that the common genomic changes between the precursor lesion and the tumour must be drivers of cancer development. However, many of the mutations and CNAs identified in EAC and subsequently characterized in BE are found throughout the histological spectrum of BE to EAC, but are not specific to patients who develop EAC. For example, Weaver et al. showed that one half of NDBE samples from patients who never progressed to EAC displayed mutations in a panel of 26 EAC driver genes (21/40 NDBE samples) including mutations to canonical cancer drivers such as *ARID1A*, *CDKN2A* and *SMARCA4* (REF.[37]). This makes these mutations poor biomarkers for early stratification despite being relatively common EAC mutations. Similarly, despite being identified in EAC and tumour-adjacent BE samples by Ross-Innes et al., *SMAD4* was found to be exclusively mutated in only a small proportion of patients with EAC when compared with patients with BE who had not yet progressed to cancer[5,37]. *TP53* mutation and 17p LOH[45,66] have provided the most specific biomarker for progression to date. *TP53* expression characterized using immunohistochemistry has been consistently correlated to HGD[86] and progression to EAC[66,87]; however, as a biomarker for early risk stratification it has poor sensitivity in NDBE (identifying 11–46% of NDBE progressors[31,88]) or even LGD (identifying 25–67% in LGD progressors[31,88,89]).

CNAs are also a common feature of both EAC and BE, and have been shown to increase over time in patients who progress from NDBE to HGD or cancer[31,63]. Related to general CNAs, higher rates of aneuploidy measured through flow cytometry have been associated with an increased risk of developing EAC[45,67,90,91]. The number of clones identified in a patient also provides some estimate of the risk of progression, as patients with higher numbers of clones are more likely to progress to cancer[34]. Both of these observations are consistent with evolutionary models of either gradual accumulation or single catastrophic events and have recently led to the development of models for predicting progression.

The best accuracy to date has been derived from using whole-genome sequencing methods to generate CNA profiles and build patient risk prediction models on longitudinal samples. Li et al. at the Fred Hutchison Cancer Research Center in Seattle first demonstrated this using whole-genome CNA profiles from single-nucleotide polymorphism arrays to characterize the Seattle Barrett's Esophagus Study cohort of 268 patients. They identified 29 copy number features that could stratify patients who progress from those who do not and reported an AUC ROC of 0.84–0.94 (REFS[63,92]). Their analyses also demonstrated that general CNA rates in BE increased significantly over time in patients who progress.

Cytosponge
A non-endoscopic device for sampling cells within the oesophagus consisting of a pill with a sponge on a string that can be swallowed by the patient.

AUC ROC
(Area under the receiver operating characteristic curve). A performance metric for classification at various thresholds by plotting the true positive rate against the false positive rate.

Elastic-net regression model
A regularized regression method that combines the penalties of LASSO and ridge methods.

Long-segment BE
Barrett's oesophagus (BE) replaces normal squamous epithelium along measurable lengths of the oesophagus from <1 cm to ≥3 cm extending from the junction of the stomach and the oesophagus.

Recent work by Douville et al.[93] used oesophageal brushings and a sequencing method called RealSeqS to characterize a set of genome-wide primers for CNAs. Using a combination of sample aneuploidy and specific chromosome arm-level CNAs (1q gain, 9p loss, 12p gain, 17p loss and 20q gain), they trained an algorithm to classify patient samples ($n = 79$) by histopathology status (for example, NDBE or HGD or EAC). This method showed quite high sensitivity for classifying EAC and HGD (96% and 68% in the validation), and suggested equally high specificity in the samples that did not progress[93]. None of the samples was analysed at the initial diagnostic endoscopy for future risk of progression; instead, this method assessed the likelihood of dysplasia or cancer in the given sample.

Our retrospective longitudinal cohort from the United Kingdom analysed 777 biopsy samples from 88 patients with follow-up of up to 15 years, resulting in the largest longitudinal cohort to date aimed at stratifying patients as early as possible[31]. To characterize CNAs in the biopsy samples we used shallow whole-genome sequencing (average depth 0.4×) as an affordable method to characterize the CNA profiles of each sample and an elastic-net regression model to identify the genomic regions that were predictive of progression to HGD or cancer. This algorithm resulted in a single-sample risk prediction, independent of histopathology grades, with an AUC ROC of 0.89 with high sensitivity (0.82) and specificity (0.83). It also enabled us to classify samples as 'low', 'moderate' or 'high' risk and demonstrate that the 'high'-risk samples were consistent in patients who progressed more than 8 years before a HGD or EAC diagnosis. Critically, our data also provided further evidence for the described evolutionary patterns of progression (which we class as 'born bad', gradual accumulation and catastrophic).

To date, none of these methods has been developed into a tool to assist clinical decision-making. Each has provided the necessary retrospective evidence that molecular characterization of CNAs can offer accurate risk stratification of patients who are likely to develop HGD or cancer. Prospective analysis of patients in surveillance with BE will be necessary to demonstrate the utility and accuracy of CNA biomarkers as well as further analysis on the inclusion of known demographic risk factors. A better understanding of biomarkers in BE and EAC can help to improve our ability to determine who is really at risk[94].

Clinical risk prediction for EAC. Early detection for cancer has the potential to reduce cancer mortality, providing the primary motivation for patient screening and surveillance. Key to the aim of early detection is identifying the population most at risk of developing disease. In EAC this has meant endoscopic surveillance of patients with BE to identify dysplasia on biopsy samples as the only currently known method for early detection. However, as our understanding of the significance of somatic mutations in cancer has increased by large-scale genome sequencing efforts through TCGA and ICGC pan-cancer analyses[6], identifying molecular risk factors offers new opportunities to improve early detection.

Current clinical practice for EAC risk prediction in patients with BE relies almost entirely on the histopathological diagnosis of dysplasia[78,79]. A few large cohort studies have concluded that there are increased risks for patients who are male, have ever smoked, have long-segment BE or are older than 60 years[80,95,96]. However, these are the same risk factors that have been found to indicate that a patient will develop BE, along with chronic acid reflux[97,98].

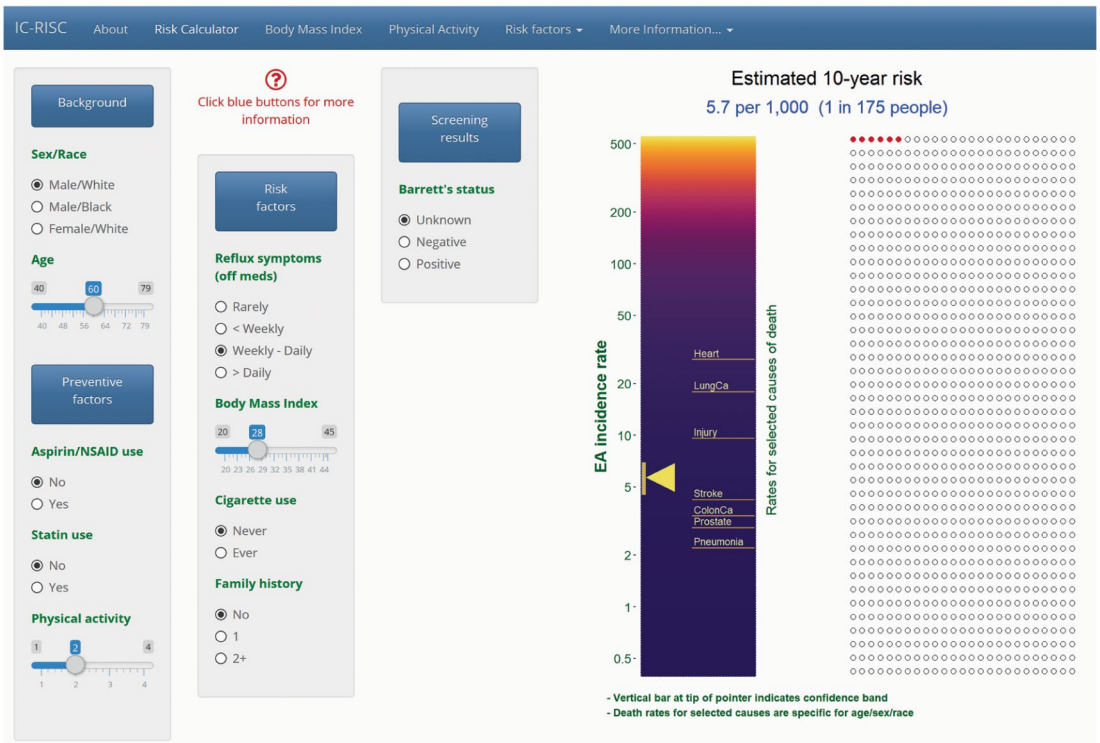
In the largest of these cohort studies from Northern Ireland, 8,522 patients with follow-up of up to 16 years were analysed for risk of progression. Men were at a greater risk of progression (0.28% per year) than women (0.13% per year) with an adjusted hazard ratio of 2.11 (95% CI = 1.41–3.16, $P < 0.001$), as were patients diagnosed above age 60 years (0.33% per year) versus patients

Table 1 | **Clinical risk prediction methods**

Method	Sample	Usage	Refs
Clinical and demographic			
Histopathology (LGD/HGD)	FFPE biopsy	Standard practice, H&E slides on standard surveillance biopsy	78,79
Age, gender, smoking, dysplasia	FFPE biopsy	Points-based system	80,95,96,102
Demographic risk factors, dysplasia	FFPE biopsy	Web application	103
Progression biomarkers			
Clonal diversity, DNA FISH probes	Endoscopic brushings	Complex, laboratory based	30
Clonal diversity, FISH, flow cytometry	Purified fresh-frozen biopsy	Complex, laboratory based	34
Aneuploidy, flow cytometry	Fresh-frozen biopsy	Complex, laboratory based	67,90
CNAs	Purified fresh-frozen biopsy	Moderate, laboratory based	63,92
CNAs	FFPE biopsy	Moderate, laboratory based	31
Aneuploidy, arm-level CNAs	Endoscopic brushings	Moderate, laboratory based	93
p53 immunohistochemistry or mutation	FFPE biopsy	Simple, laboratory based	45,104
Methylation, age	Fresh-frozen biopsy	Complex, laboratory based	105,106

CNA, copy number alteration; FFPE, formalin-fixed paraffin-embedded; FISH, fluorescence in situ hybridization; HGD, high-grade dysplasia; H&E, haematoxylin and eosin; LGD, low-grade dysplasia.

a IC-RISC (Interactive and Contextual RISk Calculator)



b BE genomic progression risk mock-up

aged under 50 years (0.12% per year) with an overall hazard ratio for age of 1.02 (95% CI = 1.01–1.03, $P < 0.05$)[80]. In a prospective cohort study of 411 patients from the Seattle Barrett's Esophagus Study, age was also found to be a significant risk for progression (HR = 1.03, 95% CI 1.00–1.06, $P = 0.02$), as was tobacco smoking, with patients who ever reported smoking at higher risk than never smokers (adjusted HR = 1.57, 95% CI 0.78–3.14), whereas alcohol and BMI had no effect on risk of progressing to EAC[96].

Both the Northern Ireland and Seattle studies identified dysplasia, either at the index biopsy or at any subsequent biopsy, as increasing the risk of progression. Patients were therefore excluded from some analyses if an index biopsy sample already showed evidence of LGD or HGD. In an effort to develop a clinically useful risk scoring system, another study, across five centres in the United States and one centre in the Netherlands, analysed 2,697 patients with BE who were under surveillance. They evaluated the influence of gender, BMI, segment

◀ Fig. 3 | **Interactive risk prediction. a** | Screenshot of the Interactive and Contextual RISk Calculator (IC-RISC) interactive application to estimate an individual's risk of developing oesophageal adenocarcinoma (EAC) using known clinical risk factors (for example, gender, ethnicity and age), Barrett's oesophagus (BE) histopathology (for example, non-dysplastic Barrett's oesophagus (NDBE), low-grade dysplasia (LGD) and high-grade dysplasia (HGD)) and known chemoprevention factors presented by Vaughan et al.[103]. **b** | Mock-up for an interactive support tool incorporating copy number alteration (CNA) risks presented by Killcoyne et al.[31] utilizing the BE progression risk tool provided therein (GitHub - gerstung-lab/BarrettsProgressionRisk) alongside known clinical risks presented by IC-RISC and BE histopathology. Part **a** is reprinted from REF.[103], CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Hiatal hernia
The upper part of the stomach bulges through the opening of the diaphragm (hiatus) into the oesophagus.

length, age, tobacco smoking, presence of hiatal hernia and baseline LGD diagnosis on risk of progression to EAC. Again, they identified the significant predictors of progression to be male gender, age, smoking, BE segment length and baseline LGD. Based on this analysis each of these features was assigned a points value to provide three risk classes: 'low' (≤10 points), 'moderate' (11–20 points) and 'high' (>20 points)[95]. Ultimately, however, risk stratification using clinical and demographic factors has not provided any additional information that alters how patients are stratified in clinical practice. In clinical practice, this means a diagnosis of dysplasia (LGD or HGD) and all men with long-segment BE (≥3 cm[78,79,99]) will be considered 'high' risk (TABLE 1).

A good clinical decision support tool for BE will ideally need to balance the clinical factors that are important to a patient's risk, such as their age, gender and history of smoking, alongside the molecular evidence of the evolutionary history and CNA patterns (FIG. 3). It also needs to consider the clinical applicability in terms of ease of use in routine practice.

One of the best recent examples of a clinical decision support tool that incorporates somatic genetic mutations, clinical information and treatment information was developed in a cohort of 1,540 patients with acute myeloid leukaemia. The analysis used 231 different variables including fusion genes, CNAs, point mutations and clinical information (for example, age, gender and treatments) to provide personalized risk predictions for therapeutic decisions[100]. Although not yet appropriate for clinical use, the tool demonstrates the utility of utilizing somatic genomic information alongside known clinical risk and can be a template for a BE risk tool (acute myeloid leukaemia multistage predictions)[101].

Whereas improving risk stratification in BE surveillance would impact only the relatively few patients who will develop EAC currently, identifying molecular biomarkers that accurately separate high-risk and low-risk patients will enable the development of less invasive, targeted surveillance strategies. It is currently unclear how the clinical and molecular information may influence the overall risk jointly, but with the example of algorithms such as the acute myeloid leukaemia multistage prediction tool[101], and the availability of clinical, longitudinal and genomic information for large patient cohorts, providing a simple to use tool for accurate risk stratification in BE is now within reach.

Conclusion and perspectives

The high mutation rate, early structural instability and clonal diversity offer evidence of an early and ongoing evolutionary process in BE. Longitudinal genomic data sets suggest that this process can result in three different patterns over time: a gradual accumulation of mutations that increase the risk of cancer each year; a tissue that is primed, or 'born bad', for tumorigenesis at its origin; or a sudden, catastrophic event that drives a previously stable lesion to cancer. This raises numerous questions about what the earliest drivers of genomic instability may be, especially in the cases where *TP53* is not mutated or lost, and whether these early events may even help to understand prognostic differences in more advanced EAC. Further work is also needed to explain the variable timescales in the progression from BE to EAC, and how this progression might help to understand early tumorigenesis across other cancer types.

It will also be important to continue investigating the origin of BE itself, whether this is related to submucosal glands in the oesophagus or arising from gastric tissues, as well as to better understand how the early inflammatory microenvironment may drive early BE and/or EAC evolution. This work is still in its infancy with a plethora of new tools now available to shed light on this topic. Finally, although there are still many other questions regarding the molecular evolution of BE, one of the critical questions for understanding EAC in patients will be to explain why one half of all patients with EAC display no histopathological evidence of BE when they present de novo and whether or how this may relate to the progression patterns described.

Ultimately, the long evolutionary history and repeated patient surveillance can be used to help improve personalized risk prediction in the management of patients with BE. By understanding the evolutionary dynamics across a BE segment through a global view of the genomic structural instability and the resulting clonal cell populations, we can also begin to improve our understanding of EAC with the aim to detect cancers earlier.

Published online 20 September 2021

1. Gerstung, M. et al. The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020). **This work, coming out of the largest pan-cancer analysis to date, establishes the concept of early and late mutations in the evolution of tumours, suggesting that early gene mutations might be detectable and offer the potential of earlier treatment.**
2. Li, Y. et al. Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112–121 (2020).
3. Zack, T. I. et al. Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
4. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).
5. Ross-Innes, C. S. et al. Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nat. Genet.* **47**, 1038–1046 (2015).
6. ICGC/TCGA. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
7. Alexandrov, L. B., Kim, J., Haradhvala, N. J. & Huang, M. N. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
8. Martincorena, I. et al. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).
9. Hoang, M. L. et al. Genome-wide quantification of rare somatic mutations in normal human tissues using massively parallel sequencing. *Proc. Natl Acad. Sci. USA* **113**, 9846–9851 (2016).
10. Brunner, S. F. et al. Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature* **574**, 538–542 (2019).
11. Lee-Six, H. et al. Population dynamics of normal human blood inferred from somatic mutations. *Nature* **561**, 473–478 (2018).

12. Yizhak, K. et al. RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science* **364**, eaaw0726 (2019).
13. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724 (2009).
14. Martincorena, I. & Campbell, P. J. Somatic mutation in cancer and normal cells. *Science* **349**, 961–968 (2016).
15. Jakubek, Y. A. et al. Large-scale analysis of acquired chromosomal alterations in non-tumor samples from patients with cancer. *Nat. Biotechnol.* **38**, 90–96 (2020).
16. Reid, B. J., Li, X., Galipeau, P. C. & Vaughan, T. L. Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. *Nat. Rev. Cancer* **10**, 87–101 (2010).
17. Spechler, S. J. Carcinogenesis at the gastroesophageal junction: free radicals at the frontier. *Gastroenterology* **122**, 1518–1520 (2002).
18. Ferlay, J. et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **136**, E359–E386 (2015).
19. Thrift, A. P. Global burden and epidemiology of Barrett oesophagus and oesophageal cancer. *Nat. Rev. Gastroenterol. Hepatol.* **18**, 432–443 (2021).
20. Hvid-Jensen, F. et al. Incidence of adenocarcinoma among patients with Barrett's esophagus. *N. Engl. J. Med.* **365**, 1375–1383 (2011).
This work is one of the largest population-based studies to show that the risk of progression from non-dysplastic BE to EAC is very low (≤0.3% per year).
21. Smyth, E. C. et al. Oesophageal cancer. *Nat. Rev. Dis. Prim.* **3**, 17048 (2017).
22. Wani, S., Rubenstein, J. H., Vieth, M. & Bergman, J. Diagnosis and management of low-grade dysplasia in Barrett's esophagus: expert review from the clinical practice updates committee of the American Gastroenterological Association. *Gastroenterology* **151**, 822–835 (2016).
23. Killcoyne, S. et al. Identification of prognostic phenotypes of esophageal adenocarcinoma in two independent cohorts. *Gastroenterology* **155**, 1720–1728 (2018).
24. Bhat, S. K. et al. Oesophageal adenocarcinoma and prior diagnosis of Barrett's oesophagus: a population-based study. *Gut* **64**, 20–25 (2015).
25. The Cancer Genome Atlas Research Network. Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169–175 (2017).
26. Frankell, A. M. et al. The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic. *Nat. Genet.* **51**, 506–516 (2019).
27. Secrier, M. et al. Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nat. Genet.* **48**, 1131–1141 (2016).
28. Nik-Zainal, S. et al. Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
29. Maley, C. C. et al. The combination of genetic instability and clonal expansion predicts progression to esophageal adenocarcinoma. *Cancer Res.* **64**, 7629–7633 (2004).
30. Martinez, P. et al. Dynamic clonal equilibrium and predetermined cancer risk in Barrett's oesophagus. *Nat. Commun.* **7**, 12158 (2016).
31. Killcoyne, S. et al. Genomic copy number predicts esophageal cancer years before transformation. *Nat. Med.* **26**, 1726–1732 (2020).
32. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
This paper together with Alexandrov et al. (*Cell Reports*, 2013 and *Nature*, 2020) establishes the concept of mutational signatures, helping to study mutational processes active within the somatic genome.
33. Newell, F. et al. Complex structural rearrangements are present in high-grade dysplastic Barrett's oesophagus samples. *BMC Med. Genomics* **12**, 31 (2019).
34. Maley, C. C. et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat. Genet.* **38**, 468–473 (2006).
Together with Maley et al. ('The combination …', *Cancer Research*, 2004), this paper provides evidence that early genetic instability, rather than individual biomarkers, can be used to predict the risk of progression in a single patient.
35. Dulak, A. M. et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat. Genet.* **45**, 478–486 (2013).
This paper from TCGA alongside Secrier et al. (2016) and Frankell et al. (2019), provides a comprehensive genomic characterization of EAC from the largest patient cohorts to date.
36. Nones, K. et al. Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. *Nat. Commun.* **5**, 5224 (2014).
37. Weaver, J. M. J. et al. Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. *Nat. Genet.* **46**, 837–843 (2014).
38. Stachler, M. D. et al. Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat. Genet.* **47**, 1047–1055 (2015).
39. Christensen, S. et al. 5-Fluorouracil treatment induces characteristic T>G mutations in human cancer. *Nat. Commun.* **10**, 1–11 (2019).
40. Tomkova, M., Tomek, J., Kriaucionis, S. & Schuster-Böckler, B. Mutational signature distribution varies with DNA replication timing and strand asymmetry. *Genome Biol.* **19**, 129 (2018).
41. Pich, O. et al. Somatic and germline mutation periodicity follow the orientation of the DNA minor groove around nucleosomes. *Cell* **175**, 1074–1087. e18 (2018).
42. Gonzalez-Perez, A., Sabarinathan, R. & Lopez-Bigas, N. Local determinants of the mutational landscape of the human genome. *Cell* **177**, 101–114 (2019).
43. Bass, A. J. et al. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).
44. Martincorena, I. et al. Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029–1041. e21 (2017).
45. Stachler, M. D. et al. Detection of mutations in Barrett's esophagus before progression to high-grade dysplasia or adenocarcinoma. *Gastroenterology* **155**, 156–167 (2018).
46. Ross-Innes, C. S. et al. Evaluation of a minimally invasive cell sampling device coupled with assessment of trefoil factor 3 expression for diagnosing barrett's esophagus: a multi-center case–control study. *PLoS Med.* **12**, e1001780 (2015).
47. Sottoriva, A. et al. A big bang model of human colorectal tumor growth. *Nat. Genet.* **47**, 209–216 (2015).
48. Martinez, P. et al. Evolution of Barrett's esophagus through space and time at single-crypt and whole-biopsy levels. *Nat. Commun.* **9**, 794 (2018).
49. Liu, Y. et al. Comparative molecular analysis of gastrointestinal adenocarcinomas. *Cancer Cell* **33**, 721–735.e8 (2018).
50. Noorani, A. et al. Genomic evidence supports a clonal diaspora model for metastases of esophageal adenocarcinoma. *Nat. Genet.* **52**, 74–83 (2020).
51. Saito, T. et al. A temporal shift of the evolutionary principle shaping intratumor heterogeneity in colorectal cancer. *Nat. Commun.* **9**, 1–11 (2018).
52. Wu, H. et al. Evolution and heterogeneity of non-hereditary colorectal cancer revealed by single-cell exome sequencing. *Oncogene* **36**, 2857–2867 (2017).
53. Dulak, A. M. et al. Gastrointestinal adenocarcinomas of the esophagus, stomach, and colon exhibit distinct patterns of genome instability and oncogenesis. *Cancer Res.* **72**, 4383–4394 (2012).
54. Hadi, K. et al. Distinct classes of complex structural variation uncovered across thousands of cancer genome graphs. *Cell* **183**, 197–210.e32 (2020).
55. Jakubek, Y. et al. Genomic landscape established by allelic imbalance in the cancerization field of a normal appearing airway. *Cancer Res.* **76**, 3676–3683 (2016).
56. Conconi, D. et al. Unexpected frequency of genomic alterations in histologically normal colonic tissue from colon cancer patients. *Tumor Biol.* **37**, 13831–13842 (2016).
57. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338–345 (2013).
58. Galipeau, P. C. et al. 17p (p53) allelic losses, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's esophagus. *Proc. Natl Acad. Sci. USA* **93**, 7081–7084 (1996).
59. Barrett, M., Galipeau, P., Sanchez, C., Emond, M. & Reid, B. Determination of the frequency of loss of heterozygosity in esophageal adenocarcinoma by cell sorting, whole genome amplification and microsatellite polymorphisms. *Oncogene* **12**, 1873–1878 (1996).
60. Barrett, M. T. et al. Allelic loss of 9p21 and mutation of the CDKN2/p16 gene develop as early lesions during neoplastic progression in Barrett's esophagus. *Oncogene* **13**, 1867–1873 (1996).
61. Galipeau, P. C., Prevo, L. J., Sanchez, C. A., Longton, G. M. & Reid, B. J. Clonal expansion and loss of heterozygosity at chromosomes 9p and 17p in premalignant esophageal (Barrett's) tissue. *JNCI* **91**, 2087–2095 (1999).
62. 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).
63. Li, X. et al. Temporal and spatial evolution of somatic chromosomal alterations: a case–cohort study of Barrett's esophagus. *Cancer Prev. Res.* **7**, 114–127 (2014).
64. Levine, D. S., Reid, B. J., Haggitt, R. C., Rubin, C. E. & Rabinovitch, P. S. Correlation of ultrastructural aberrations with dysplasia and flow cytometric abnormalities in Barrett's epithelium. *Gastroenterology* **96**, 355–367 (1989).
65. Reid, B. et al. Flow-cytometric and histological progression to malignancy in Barrett's esophagus: prospective endoscopic surveillance of a cohort. *Gastroenterology* **102**, 1212–1219 (1992).
66. 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).
Together with Galipeau et al. (1996), Galipeau et al. (1999) and Reid et al. (2001), this paper provides evidence that early copy number changes occur in BE and confer increased risk of cancer to patients with BE.
67. 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).
68. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
69. Curtius, K. et al. A molecular clock infers heterogeneous tissue age among patients with Barrett's esophagus. *PLoS Comput. Biol.* **12**, e1004919 (2016).
70. Stephens, P. J. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
71. Gao, R. et al. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat. Genet.* **48**, 1119–1130 (2016).
72. Bonnington, S. N. & Rutter, M. D. Surveillance of colonic polyps: are we getting it right? *World J. Gastroenterol.* **22**, 1925–1934 (2016).
73. Cheng, Y.-W. et al. CpG island methylator phenotype associates with low-degree chromosomal abnormalities in colorectal cancer. *Clin. Cancer Res.* **14**, 6005–6013 (2008).
74. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
75. Baca, S. C. et al. Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677 (2013).
76. Ly, P. & Cleveland, D. W. Rebuilding chromosomes after catastrophe: emerging mechanisms of chromothripsis. *Trends Cell Biol.* **27**, 917–930 (2017).
77. Killcoyne, S. & Fitzgerald, R. C. Practical early cancer detection: distinguishing stable from unstable genomes in pre-cancerous tissues. *Br. J. Cancer* **124**, 683–685 (2020).
78. Shaheen, N. J., Falk, G. W., Iyer, P. G. & Gerson, L. B. ACG clinical guideline: diagnosis and management of Barrett's esophagus. *Am. J. Gastroenterol.* **111**, 30–50 (2016).
79. Fitzgerald, R. C. et al. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. *Gut* **63**, 7–42 (2014).
80. Bhat, S. et al. Risk of malignant progression in Barrett's Esophagus patients: results from a large population-based study. *J. Natl. Cancer Inst.* **103**, 1049–1057 (2011).
81. Shaheen, N. J. et al. Radiofrequency ablation in Barrett's esophagus with dysplasia. *N. Engl. J. Med.* **360**, 2277–2288 (2009).
82. Phoa, K. N. et al. Radiofrequency ablation vs endoscopic surveillance for patients with Barrett esophagus and low-grade dysplasia: a randomized clinical trial. *JAMA* **311**, 1209–1217 (2014).
83. Fitzgerald, R. C. et al. Cytosponge-trefoil factor 3 versus usual care to identify Barrett's oesophagus

in a primary care setting: a multicentre, pragmatic, randomised controlled trial. *Lancet* **396**, 333–344 (2020).
84. Davidson, M. et al. Detecting and tracking circulating tumour DNA copy number profiles during first line chemotherapy in oesophagogastric adenocarcinoma. *Cancers* **11**, 736 (2019).
85. Babayan, A. & Pantel, K. Advances in liquid biopsy approaches for early detection and monitoring of cancer. *Genome Med.* **10**, 21 (2018).
86. van der Wel, M. J. et al. Improved diagnostic stratification of digitised Barrett's oesophagus biopsies by p53 immunohistochemical staining. *Histopathology* **72**, 1015–1023 (2018).
87. Hamelin, R. et al. TP53 gene mutations and p53 protein immunoreactivity in malignant and premalignant Barrett's esophagus. *Gastroenterology* **107**, 1012–1018 (1994).
88. Bang, Y. J. et al. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet* **376**, 687–697 (2010).
89. Davelaar, A. L. et al. Aberrant TP53 detected by combining immunohistochemistry and DNA-FISH improves Barrett's esophagus progression prediction: a prospective follow-up study. *Genes Chromosom. Cancer* **54**, 82–90 (2015).
90. Reid, B. J., Levine, D. S., Longton, G., Blount, P. L. & Rabinovitch, P. S. Predictors of progression to cancer in Barrett's esophagus: baseline histology and flow cytometry identify low- and high-risk patient subsets. *Am. J. Gastroenterol.* **95**, 1669–1676 (2000).
91. Hadjinicolaou, A. V. et al. Aneuploidy in targeted endoscopic biopsies outperforms other tissue biomarkers in the prediction of histologic progression of Barrett's oesophagus: a multi-centre prospective cohort study. *Ebiomedicine* **56**, 102765 (2020).
92. Li, X. et al. Assessment of esophageal adenocarcinoma risk using somatic chromosome alterations in longitudinal samples in Barrett's esophagus. *Cancer Prev. Res.* **8**, 845–856 (2015).
93. Douville, C. et al. Massively parallel sequencing of esophageal brushings enables an aneuploidy-based classification of patients with Barrett's esophagus. *Gastroenterology* **160**, 2043–2054 (2021).
94. Vaughan, T. L. & Fitzgerald, R. C. Precision prevention of oesophageal adenocarcinoma. *Nat. Rev. Gastroenterol. Hepatol.* **12**, 243–248 (2015).
95. Parasa, S. et al. Development and validation of a model to determine risk of progression of Barrett's esophagus to neoplasia. *Gastroenterology* **154**, 1282–1289.e2 (2018).
96. Hardikar, S. et al. The role of tobacco, alcohol, and obesity in neoplastic progression to esophageal adenocarcinoma: a prospective study of Barrett's esophagus. *PLoS ONE* **8**, e52192 (2013).
97. Zagari, R. M. et al. Gastro-oesophageal reflux symptoms, oesophagitis and Barrett's oesophagus in the general population: the Loiano-Monghidoro study. *Gut* **57**, 1354–1359 (2008).
98. Ronkainen, J. et al. Prevalence of Barrett's esophagus in the general population: an endoscopic study. *Gastroenterology* **129**, 1825–1831 (2005).
99. Hamade, N. et al. Lower annual rate of progression of short-segment vs long-segment Barrett's esophagus to esophageal adenocarcinoma. *Clin. Gastroenterol. Hepatol.* **17**, 864–868 (2019).
100. Papaemmanuil, E. et al. Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med.* **374**, 2209–2221 (2016).
101. Gerstung, M. et al. Precision oncology for acute myeloid leukemia using a knowledge bank approach. *Nat. Genet.* **49**, 332–340 (2017).
102. Cook, M. B. et al. Cigarette smoking increases risk of Barrett's esophagus: an analysis of the Barrett's and esophageal adenocarcinoma consortium. *Gastroenterology* **142**, 744–753 (2012).
103. Vaughan, T. L., Onstad, L. & Dai, J. Y. Interactive decision support for esophageal adenocarcinoma screening and surveillance. *BMC Gastroenterol.* **19**, 109 (2019).
104. Keswani, R. N., Noffsinger, A., Waxman, I. & Bissonnette, M. Clinical use of p53 in Barrett's esophagus. *Cancer Epidemiol. Biomarkers Prev.* **15**, 1243–1249 (2006).
105. Jin, Z. et al. A multicenter, double-blinded validation study of methylation biomarkers for progression prediction in Barrett's esophagus. *Cancer Res.* **69**, 4112–4115 (2009).
106. Sato, F. et al. Three-tiered risk stratification model to predict progression in Barrett's esophagus using epigenetic and clinical features. *PLoS ONE* **3**, e1890 (2008).
107. Souza, R. F. Reflux esophagitis and its role in the pathogenesis of Barrett's metaplasia. *J. Gastroenterol.* **52**, 767–776 (2017).
108. Wang, D. H. The esophageal squamous epithelial cell — still a reasonable candidate for the Barrett's esophagus cell of origin? *Cell. Mol. Gastroenterol. Hepatol.* **4**, 157–160 (2017).
109. Kong, J., Crissey, M. A., Funakoshi, S., Kreindler, J. L. & Lynch, J. P. Ectopic Cdx2 expression in murine esophagus models an intermediate stage in the emergence of Barrett's esophagus. *PLoS ONE* **6**, 18280 (2011).
110. Clemons, N. J. et al. Sox9 drives columnar differentiation of esophageal squamous epithelium: a possible role in the pathogenesis of Barrett's esophagus. *Am. J. Physiol. Gastrointest. Liver Physiol.* **303**, 1335–1346 (2012).
111. Owen, R. P. et al. Single cell RNA-seq reveals profound transcriptional similarity between Barrett's oesophagus and oesophageal submucosal glands. *Nat. Commun.* **9**, 4261 (2018).
112. Leedham, S. J. et al. Individual crypt genetic heterogeneity and the origin of metaplastic glandular epithelium in human Barrett's oesophagus. *Gut* **57**, 1041–1048 (2008).
113. Jiang, M. et al. Transitional basal cells at the squamous-columnar junction generate Barrett's oesophagus. *Nature* **550**, 529–533 (2017).
114. Wang, X. et al. Residual embryonic cells as precursors of a Barrett's-like metaplasia. *Cell* **145**, 1023–1035 (2011).
115. McQuaid, K. R., Laine, L., Fennerty, M. B., Souza, R. & Spechler, S. J. Systematic review: The role of bile acids in the pathogenesis of gastro-oesophageal reflux disease and related neoplasia. *Aliment. Pharmacol. Ther.* **34**, 146–165 (2011).
116. Gokon, Y. et al. Immune microenvironment in Barrett's esophagus adjacent to esophageal adenocarcinoma: possible influence of adjacent mucosa on cancer development and progression. *Virchows Arch.* **477**, 825–834 (2020).
117. Fitzgerald, R. C. et al. Diversity in the oesophageal phenotypic response to gastro-oesophageal reflux: Immunological determinants. *Gut* **50**, 451–459 (2002).
118. Kavanagh, M. E. et al. Impact of the inflammatory microenvironment on T-cell phenotype in the progression from reflux oesophagitis to Barrett oesophagus and oesophageal adenocarcinoma. *Cancer Lett.* **370**, 117–124 (2016).
119. Lagisetty, K. H. et al. Immune determinants of Barrett's progression to esophageal adenocarcinoma. *JCI Insight* **6**, e143888 (2021).
This paper provides an overview of the changes within the immune environment between BE and EAC.
120. Wagener-Ryczek, S. et al. Immune profile and immunosurveillance in treatment-naive and neoadjuvantly treated esophageal adenocarcinoma. *Cancer Immunol. Immunother.* **69**, 523–533 (2020).
121. Galipeau, P. C. et al. NSAIDs modulate CDKN2A, TP53, and DNA content risk for progression to esophageal adenocarcinoma. *PLoS Med.* **4**, e67 (2007).
122. Galipeau, P. C. et al. NSAID use and somatic exomic mutations in Barrett's esophagus. *Genome Med.* **10**, 17 (2018).
123. Liao, L. M. et al. Nonsteroidal anti-inflammatory drug use reduces risk of adenocarcinomas of the esophagus and esophagogastric junction in a pooled analysis. *Gastroenterology* **142**, 442–452.e5; quiz e22–e23 (2012).
124. Jankowski, J. A. Z. et al. Esomeprazole and aspirin in Barrett's oesophagus (AspECT): a randomised factorial trial. *Lancet* **392**, 400–408 (2018).
This paper presents AspECT, the largest clinical trial to date investigating the real effect of aspirin and proton-pump inhibitors to delay or inhibit EAC in patients with BE.

Acknowledgements
The laboratory of R.C.F. is funded by a Programme Grant from the Medical Research Council (MRC) (RG84369), and Cancer Research UK provided funding for the Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS)/International Cancer Genome Consortium (ICGC) oesophageal cancer programme (RG66287).

Author contributions
Both authors contributed equally to this article.

Competing interests
R.C.F. is named on patents for Cytosponge and related assays that have been licensed by the Medical Research Council (MRC) to Covidien GI Solutions (now Medtronic). R.C.F is a co-founder and shareholder of Cyted Ltd, an Early Detection company. S.K. declares no competing interests.

Peer review information
Nature Reviews Cancer thanks M. Stachler, T. Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

RELATED LINKS
Acute myeloid leukaemia multistage predictions: https://cancer.sanger.ac.uk/aml-multistage/
GitHub — gerstung-lab/BarrettsProgressionRisk: https://github.com/gerstung-lab/BarrettsProgressionRisk
Interactive and Contextual RISk Calculator (IC-RISC): https://ic-risc.fredhutch.org/