 doi:10.20944/preprints202207.0321.v1

Somatic variation in normal tissues: friend or foe of cancer early detection?

A. Acha-Sagredo[1,2†], P. Ganguli[1,2†], F. D. Ciccarelli[1,2*]

[1] Cancer Systems Biology Laboratory, The Francis Crick Institute, London NW1 1AT, United Kingdom

[2] School of Cancer and Pharmaceutical Sciences, King's College London, London SE11UL, United Kingdom

[†] = shared first authorship

*** Corresponding Author**
Prof. Francesca D. Ciccarelli
Address: Cancer Systems Biology Laboratory, The Francis Crick Institute, London NW1 1AT, United Kingdom
Telephone: +44 (0)20 3796 3460
Email: francesca.ciccarelli@crick.ac.uk

Author ORCID identifiers:

A. Acha-Sagredo: 0000-0001-8265-8604

P. Ganguli: 0000-0002-9670-3928

F. D. Ciccarelli: 0000-0002-9325-0900

Twitter: @fcicca, @ameachasag, @pganguli17, @TheCrick, @KCLcancer_pharm

 doi:10.20944/preprints202207.0321.v1

ABSTRACT

Seemingly normal tissues progressively become populated by mutant clones over time. Most of these clones bear mutations in well-known cancer genes but only rarely do they transform into cancer. This poses questions on what triggers cancer initiation and what implications somatic variation has for cancer early detection. We analysed recent mutational screens of healthy and cancer-free diseased tissues to compare somatic drivers and the causes of somatic variation across tissues. We then reviewed the mechanisms of clonal expansion and their relationships with age and diseases other than cancer. We finally discussed the relevance of somatic variation for cancer initiation and how it can help or hinder cancer detection and prevention. The extent of somatic variation is highly variable across tissues and depends on intrinsic features, such as tissue architecture and turnover, as well as the exposure to endogenous and exogenous insults. Most somatic mutations driving clone expansion are tissue-specific and inactivate tumour suppressor genes involved in chromatin modification and cell growth signalling. Some of these genes are more frequently mutated in normal tissues than cancer, indicating a context-dependent cancer promoting or protective role. Mutant clones can persist over a long time or disappear rapidly, suggesting that their fitness depends on the dynamic equilibrium with the environment. The disruption of this equilibrium is likely responsible for their transformation into malignant clones and knowing what triggers this process is key for cancer prevention and early detection. Somatic variation should be considered in liquid biopsy, where it may contribute cancer-independent mutations, and in the identification of cancer drivers, since not all mutated genes favouring clonal expansion also drive tumourigenesis. Somatic variation and the factors governing homeostasis of normal tissues should be taken into account when devising strategies for cancer prevention and early detection.

 doi:10.20944/preprints202207.0321.v1

Key words: Somatic evolution, driver gene, clone selection, healthy tissues, cancer initiation, cancer early detection

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

INTRODUCTION

Cancer has long been referred to as a disease of the genome because of the pivotal role played by genetic alterations in driving its initiation and progression[1]. Only recently, however, cancer mutational screens have revealed the extent of cancer genomic modifications that often accumulate over several years[2, 3]. These studies have also expanded our knowledge on the genetic basis of cancer. The analysis of thousands of cancer exomes and genomes has led to the identification of more than 3,000 mutated genes with a proven or predicted driver role in cancer[4, 5].

Currently, the cancer driver activity of a gene is either assessed experimentally or predicted with computational approaches that measure the evolutionary forces acting on it or the effect and properties of its alterations[6, 7]. With only few notable exceptions, the vast majority of known or predicted cancer drivers promote cancer only in specific tissues[4]. Moreover, the majority of cancer genomes bear mutations in more than one driver, supporting early theoretical work on the need of multiple hits to initiate tumourigenesis[8].

In addition to identifying the driver events, cancer mutational screens have been used to infer the mutational processes active in cancer cells and formulate models of cancer evolution. Phylogenetic trees based on alteration clonality[9] enable reconstruction of the evolutionary paths of individual cancer samples from the seeding cell to the time of sequencing. These can then be used to interpret and predict future evolutionary trajectories, including response to therapy[10]. Knowing the genome sequence of fully-fledged tumours, however, does not inform on events predating cancer transformation. In fact, it tell very little about the early phases of tumour formation, namely the events and conditions that promote transformation of normal cells into cancer cells.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

One of the main challenges for detecting pre-cancer mutations is that, prior to the clonal expansion associated with cancer, they hit only a small fraction of cells. These mutations are therefore diluted within the tissue and their frequency is usually below the detection power of conventional sequencing methods. Until the advent of high throughput sequencing technology, only a few somatic alterations occurring in apparently normal tissues were documented. Among these were the inactivation of cytochrome c oxidase and *TP53* in colon and skin detected through immunostaining or conventional Sanger sequencing[11-14]. However, the extent of somatic variation occurring in the human genome has started to be fully appreciated only recently[15, 16]. High-throughput sequencing coupled with bioinformatic data analysis have finally enabled quantification of low frequency alterations occurring in phenotypically normal tissues.

In this review, we summarise the results of mutational screens in non-cancer tissues, focusing on what they have revealed about the origin of somatic mutations and their impact on tissue homeostasis and disease. We then discuss the relevance of somatic variation for cancer initiation and how it can help or hinder strategies to improve cancer detection and prevention.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

The mutational landscapes of histologically normal tissues

Recent advances in DNA sequencing technologies and computational approaches for data analysis have enabled detection of somatic mutations occurring in only few cells within adult tissues. DNA extracted from macro-dissected tissue slides (**Figure 1A**), microscopically identifiable clonal structures (**Figure 1B**), clones expanded *ex-vivo* (**Figure 1C**), or single cell populations (**Figure 1D**) can be sequenced at high depth to identify rare alterations. The resulting repertoire of somatic mutations can then be used to quantify the selective pressure driving clone expansion, identify the underneath mutagenic processes, and rebuild tissue somatic evolution in time and space (**Figure 1E**).

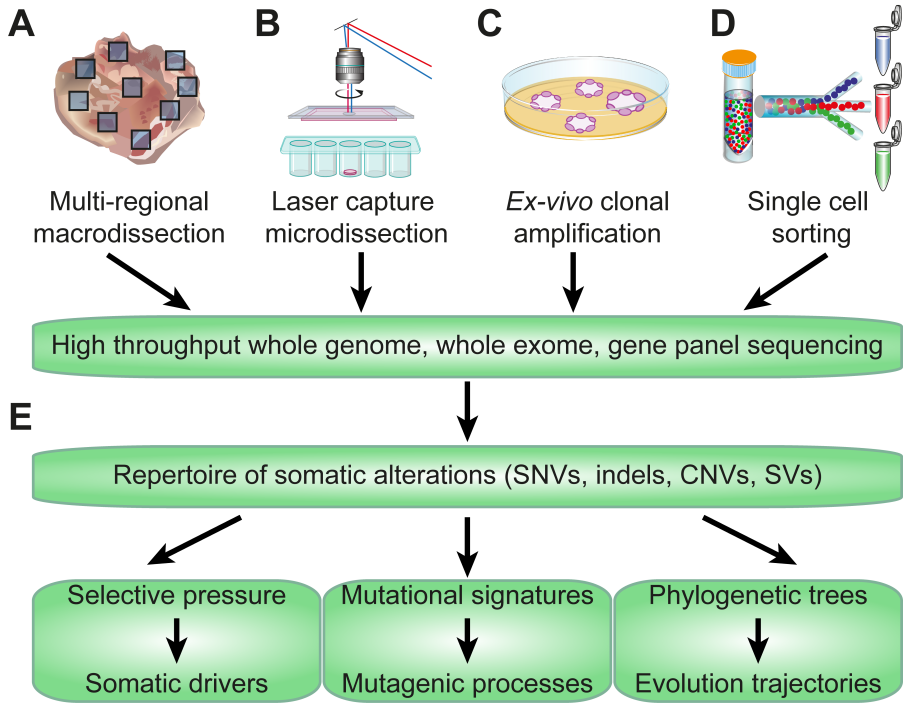

Figure 1 – Approaches to detect and analyse somatic mutations in normal tissues. DNA extracted from **(A)** macro or **(B)** micro dissected tissues, with or without subsequent targeted bulk resequencing, **(C)** *ex-vivo* clonal expansion of isolated cells and **(D)** single cell sorting is sequenced using next-generation sequencing approaches that allow high-throughput detection of somatic mutations that can then be used to identify the drivers of clone expansion, the mutational processes causing them and to trace tissue evolution **(E)**. SNVs: single nucleotide variants, CNVs: copy number variants, SVs: structural variants

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

During life, the homeostasis of most tissues is preserved through the asymmetric divisions of adult stem cells, which enable the maintenance of a stem cell pool while sustaining tissue renewal through the progressive differentiation of progenitor cells **(Figure 2A)**. The acquisition of somatic alterations in the genome of stem or progenitor cells may result in their increased fitness that fuels the clonal expansion of their progenies, which eventually populate part of the tissue **(Figure 2B)**.

Somatic variation has shown recurrent features across all tissues sequenced so far. For example, the mutational load as well as the number and size of mutant clones increase with age, in the presence of inflammatory conditions and upon exposure to mutagens **(Figure 2C)**. Moreover, somatic clones only rarely acquire copy number alterations, structural rearrangements or chromosomal abnormalities.

Despite these commonalities, the number and size of clones vary substantially across tissues suggesting that their proliferative potential does not depend uniquely on the intrinsic advantages contributed by mutations. The architecture of the tissue and the frequency of its turnover **(Figure 2D)** also likely play major roles in determining the clone fate. Hematopoietic stem cells produce thousands of mature blood cells every day and mutant clones can in principle expand freely in the bloodstream. Accordingly, age-dependent clonal haematopoiesis, *i.e.* the expansion of mutant haematopoietic cells sharing a common origin, is highly diffuse in the ageing general population[17-19].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

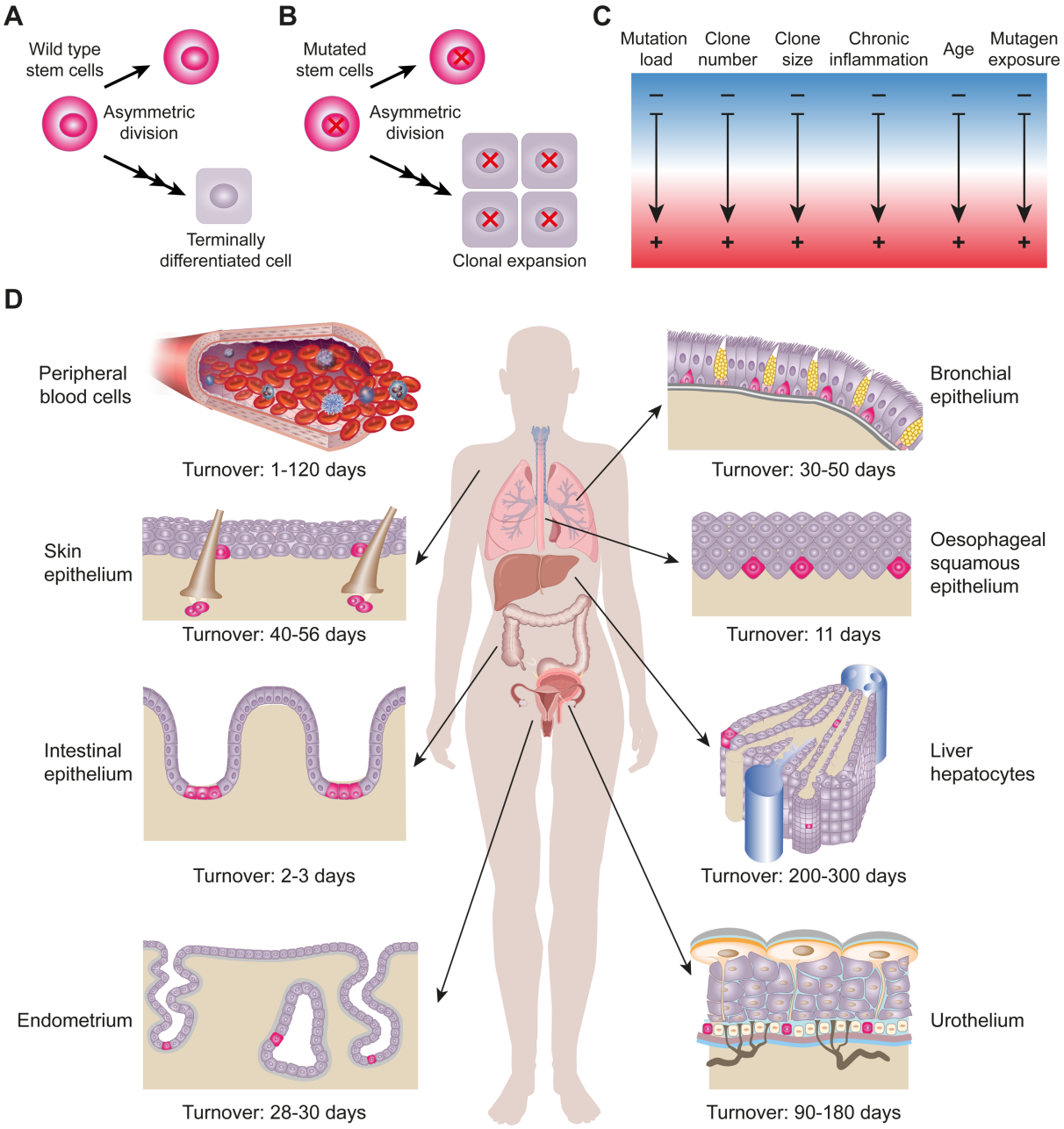

Figure 2 – Somatic evolution of normal tissues. (A) Tissue homeostasis is maintained through asymmetric division of wild type stem cells. **(B)** Somatic mutations conferring fitness advantages result in clonal expansion of the mutant progenies. **(C)** Recurrent features of somatic evolution across normal tissues. **(D)** Schematic representation of the structure and turnover of histologically normal human adult tissues. Turnover data were taken from[20-24].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

Unlike the blood, solid tissues pose spatial barriers to clone expansion. For example, the intestinal epithelium is organised into well-defined clonal structures known as crypts that undergo continual renewal during life. Despite the high tissue turnover, clone expansion beyond the single crypt (a phenomenon known as 'crypt fission') rarely occurs in healthy gut[25-29]. Normal liver also usually hosts relatively few mutant clones[30-32], possibly due to the low turnover and the lobular structure of the tissue. The mutational landscape of both gut and liver changes drastically in the presence of inflammatory disorders such as inflammatory bowel disease or cirrhosis, which positively correlate with the number of mutant clones[30, 32].

An increased number of clones is also observed in endometriotic endometrium[33-35], confirming that chronic inflammation remodels adult tissues through continuous cycles of destruction and repair that favour clone outgrowth. Unlike normal colon and liver, mutant clones almost completely replace non-inflamed endometrium by menopause[35-37]. This is likely facilitated by the 'rhizome' structure of the endometriotic epithelium, in which vertical glands acquire additional mutations during every menstrual cycle[38].

The epithelia of skin and oesophagus also progressively become a patchwork of mutant clones during life[39-44]. In both tissues, the stem/progenitor cell compartments are localised above the basement membrane of the epithelium **(Figure 2D)**, which poses a weaker barrier to the propagation of mutant clones than intestinal crypts or hepatic lobes. As expected due to the higher exposure to external mutagens, skin accumulates around ten-fold more mutations than oesophagus[42]. Interestingly, recent observations suggest that the mutagenic effect of some exogenous insults, and the consequent expansion of mutant clones, may be reversible. For example, despite the mutation burden being generally higher in smokers or ex-smokers than in never

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

smokers, high variability has been observed across and within individuals. In particular, some clones show comparably low mutational burden in current, former, and never smokers[45], indicating that their stem cells are less susceptible to smoking mutagens. Lowly mutant clones are fourfold more frequent in ex-smokers than current smokers and can repopulate the bronchial epithelium once the exposure to smoking ends. This suggests that the fitness advantage of somatic mutations is context-dependent and varies with circumstances.

Extensive inter- and intra-individual variation in the mutational spectrum has also been observed in the urothelium of bladder and ureter, which, despite the relatively low turnover, become substantially populated by mutant clones over time[46, 47].

Genes and mutational processes driving somatic clone expansion

Genes acquiring somatic mutations that increase cell fitness and drive clonal expansion (somatic drivers) are identified using similar approaches to those used for cancer drivers, preferentially detecting frequently mutated genes[4]. So far, these approaches have identified 147 somatic drivers across nine tissues **(Supplementary Table S1)**. Almost 90% of these genes are well-known (canonical) or predicted (candidate) cancer drivers and tumour suppressors outnumber oncogenes **(Figure 3A)**. This is in line with the prevalence of somatic point mutations and small indels that are more likely to inactivate tumour suppressors.

Functionally, somatic drivers that are also cancer drivers are typically signalling genes mediating cell growth or chromatin modifiers **(Figure 3B)**. Given their role in cell differentiation[48], it is tempting to speculate that mutations in chromatin modifiers promote cell dedifferentiation and self-renewal that, in turn, favour clone expansion. The few somatic drivers that are not cancer drivers do not show any significant

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

functional enrichment, indicating no convergence towards the disruption of any particular biological process.

Unlike cancer, where the higher the size of the analysed cohort the more drivers become detectable[4], the number of somatic drivers does not increase with sample or donor size **(Figure 3C)**. For example, clone expansion in blood is driven by a similar number of genes as in intestine or diseased endometrium, despite 20-fold more blood samples having been sequenced. This suggests that the early phases of somatic clone expansion tend to be promoted by the same genes driving cancer, but the extent of inter-individual heterogeneity of the somatic driver repertoire is more limited. This also confirms that clone expansion depends on the features of the tissue as well as its exposure to mutagens, in addition to the intrinsic advantages of the mutant cells.

The tissue-specificity of the somatic driver landscape is further supported by the low recurrence of drivers across tissues, with only 13 genes driving clonal expansion in three or more tissues **(Figure 3D)**. An extreme case is again blood that shares only *TP53* with other tissues, indicating that clonal haematopoiesis is promoted by a small and tissue-specific set of somatic drivers. There are clear differences even across solid tissues. For example, multiple mutational screens of skin and endometrium have reported alterations in the same drivers (*NOTCH1, FAT1,* and *PIK3CA*, *KRAS*, respectively, **Figure 3D**), due to parallel or convergent evolution. In the former case, clones carry distinct, inactivating mutations (*NOTCH1* or *FAT1*), while in the latter they converge towards the same activating mutation (*KRAS* or *PIK3CA*). This does not occur in other tissues, where different screens identified different drivers.

Intriguingly, a few well-known cancer drivers, notably *KRAS* in endometrium, the *NOTCH* genes in skin and oesophagus and the *ERBB* genes in colon, are more frequently altered in normal tissues than in the corresponding cancers

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

(Supplementary Table S1). This suggests that some cancer drivers may have either a cancer promoting or a cancer protective role depending on the context and time of their alteration.

The patterns of mutations occurring in the genome of mutant clones, known as mutational signatures, are indicative of the processes responsible for somatic mutagenesis. Signatures related to endogenous mutational processes are prevalent in all tissues sequenced so far **(Figure 3E)**. Most of these mutations are likely acquired in the early stages of embryonic development[49, 50] and continue to accumulate throughout life. The pervasiveness of endogenous signatures indicates that the main source of mutational variation in somatic tissues is ageing. Signatures induced by reactive oxygen species, APOBEC and tobacco smoking are also relatively frequent. Other external mutagens, such as UV light, aristolochic acid or colibactin are instead specific to skin, urothelium and intestine, respectively. This is consistent with their cancer promoting role in these organs confirming that, at least in these cases, normal clone expansion and cancer initiation have the same mutagenic origins.

Together with the mutational signatures found in cancer, normal tissues show several novel signatures that have never been described before **(Supplementary Table S1)**. These may be hidden by the prevalence of stronger mutational processes that take over during cancer evolution or may indicate a different origin of somatic mutations that do not eventually evolve into cancer. None of these novel mutational signatures have a known aetiology, which prevents from discriminating between these two scenarios.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

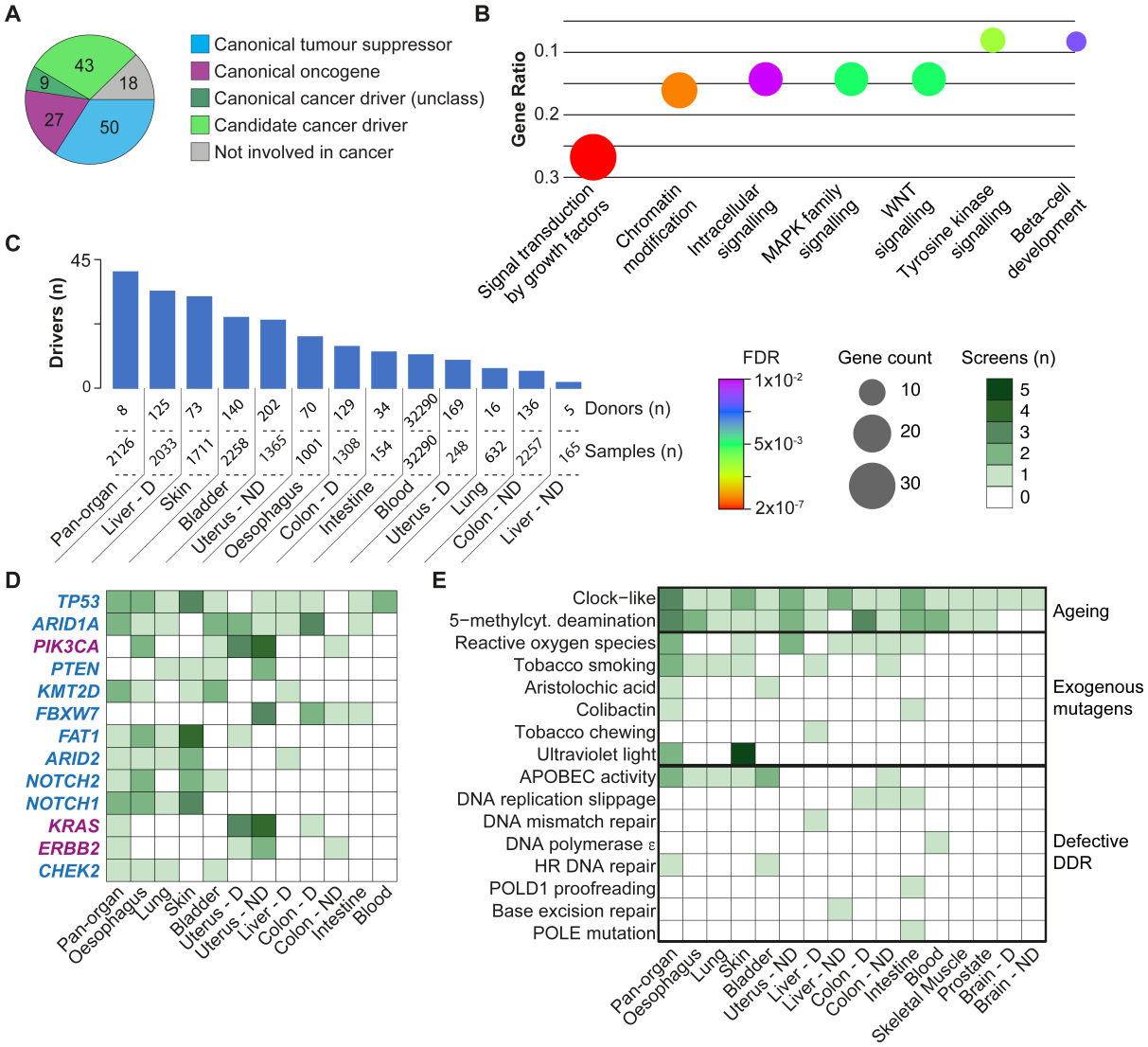

Figure 3 – Somatic driver repertoire. (A) Breakdown of somatic drivers that are also cancer drivers or that have not been associated with cancer. Canonical and candidate cancer drivers were derived from the NCG database (http://www.network-cancer-genes.org/)[4]. **(B)** Gene set enrichment analysis of somatic drivers in level 2 Reactome pathways v.72[51] as compared to the other cancer drivers from the NCG database. Enrichment was calculated using one-sided Fisher's exact test corrected with Benjamini-Hochberg for multiple testing. Gene ratio represents the proportion of somatic drivers over the total. Circle size indicates the gene count per pathway. **(C)** Number of unique somatic drivers per tissue. **(D)** Somatic drivers recurring in three or more tissues. **(E)** Aetiologies of somatic mutations as derived from the signatures reported in **Supplementary Table S1**. The aetiologies were assigned using COSMIC v.2 and v3.2[5] and grouped based on similarities. D: diseased (non-cancer) tissue, ND: non-diseased tissue, FDR: false discovery rate, HR: homologous recombination, DDR: DNA damage response, 5-methylcyt.: 5-methylcytosine; POL: polymerase.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

Origins and consequences of somatic mutations in ageing and disease

Somatic mutations are acquired from early development throughout adult life, with clones growing in number and size over time **(Figure 4A)**.

Interestingly, mutation rate is higher during foetal development than in post-natal cells[52-55], especially in the first three embryonic divisions[49, 53, 56]. This is likely due to the absence of transcription associated DNA repair[57, 58] and a higher tolerance towards DNA damage due to the lack of apoptosis[58-60] during very early development. Fixation of embryonic mutations often occurs by neutral drift rather than selection and mutant cells can eventually populate large portions of one or more tissues, as in the case of the same mutations found in brain and spleen[54].

Somatic mutations that promote clone expansion during embryonic development or adult life hit dividing cells that most likely are stem or progenitor cells. However, mutations may occur also in post-mitotic tissues and affect slowly- or non-dividing cells, such as visceral smooth muscle and neurons[61, 62]. For example, the post-mitotic expansion of CAG repeats in neurons is known to cause Huntington's disease[63]. Recent technical innovations, including single-cell[64] and single-molecule[62] DNA sequencing have shown that post-mitotic neurons accumulate mutations at a similar rate than mitotically active cells. This surprising result indicates that, together with errors generated during cell divisions, mutations can continuously arise from non-mitotic insults. Although the signatures of post-mitotic mutations do not point towards any specific aetiology, their linear accumulation over time suggests that they are the result of a dynamic equilibrium between DNA damage and repair throughout life[62].

Do somatic mutations result always in disease conditions? While a clear link exists between mutation accumulation and cancer, as extensively discussed below, still relatively little is known about their role in other diseases. Embryonic mutations

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

that disrupt Mendelian genes may result in similar but less severe syndromes than germline mutations. Examples include overgrowth syndromes where somatic mutations confer growth advantages to mutant cells located in specific areas of the body[65, 66] and almost 10% of mutations causing autism spectrum disorder[67]. Clonal haematopoiesis is a known risk factor in cardiovascular disease due to a combination of increased inflammation and mutation-specific effects[68], while somatic mutations in immune cells may favour the onset of immune disorders[69]. Despite these examples, however, the widespread diffusion of somatic mutations in the normal population and the phenotypically normal appearance of mutated tissues suggest that most mutations, even when favouring clone expansion, are not pathogenic.

In addition to disease, the accumulation of mutations has long been associated with ageing. Mutations are thought to favour the progressive decline of cell functions[70-72], although the molecular basis of this remains largely elusive. It has been proposed that somatic mutations could reduce the efficiency of gene regulatory networks and increase cell-to-cell transcriptional heterogeneity[72, 73]. However, the high somatic mutation rate due to germline *POLE/POLD1* defects does not lead to any appreciable sign of accelerated ageing at least in gut and endothelium[29], suggesting that a more complex relationship likely exists between mutation and ageing.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

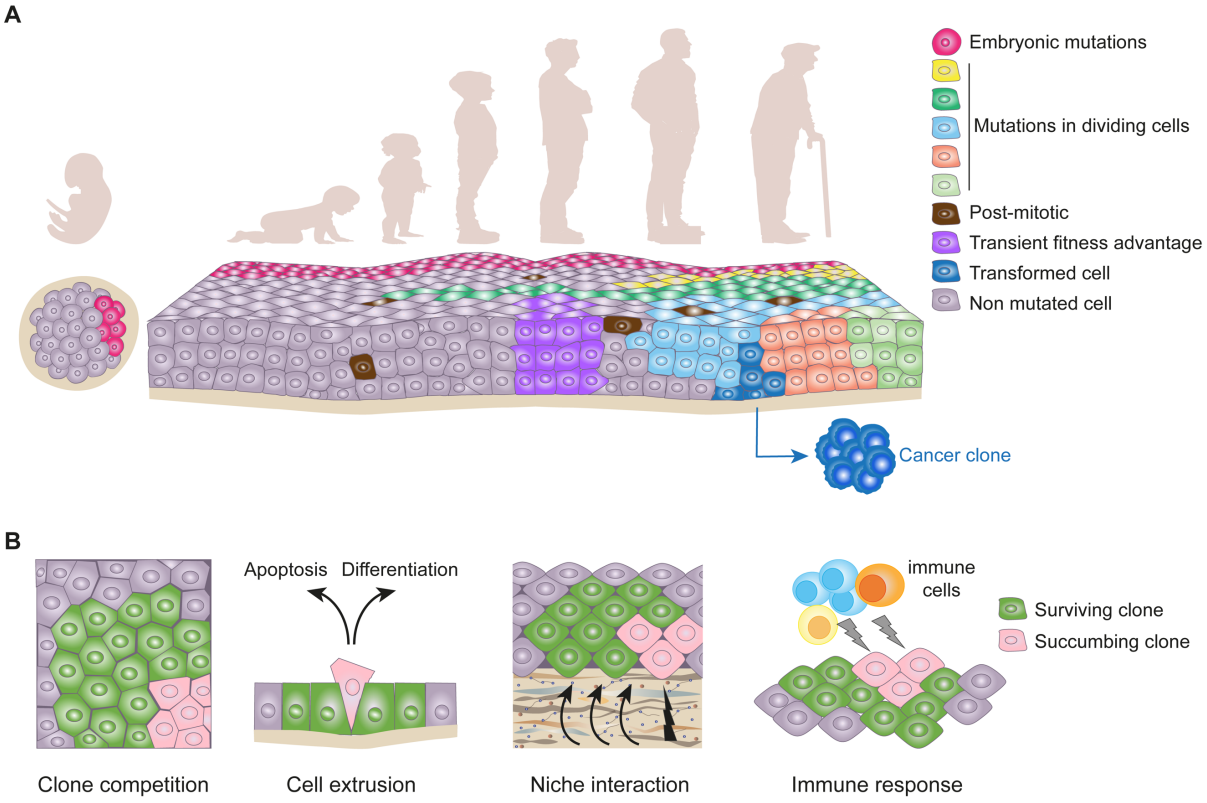

Figure 4 – Origin, evolution and fates of somatic clones. (A) Somatic mutations accumulate throughout life resulting in somatic mosaicism. Some mutations are acquired during embryo development and fixed by neutral drift. Other mutations arise in post-mitotic or actively dividing cells due to the exposure to endogenous or exogenous insults. Mutations that confer fitness advantages initiate clone expansion of actively dividing cells. In some cases, these advantages are transient and the clone disappears when the insult is removed. Some mutant cells may acquire transforming potential and start the tumorigenic process. **(B)** Clone selection and expansion are driven by several intrinsic and extrinsic factors such as competition between mutant cells with different fitness, cell extrusion leading to apoptosis or differentiation, active responses of the stromal niche and selective pressure of the immune system.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

Somatic mutations and cancer transformation

The pervasiveness of mutant clones in phenotypically normal tissues poses the questions of how these clones form, grow and survive and under what circumstances they transform into cancer.

In the early phases of clone formation, competition between mutant cells with different fitness is a key factor for their survival[74, 75]. Lineage tracing of mutagen-driven clone formation in the mouse oesophagus has shown that *NOTCH1* mutant clones have higher fitness and outcompete *NOTCH1* wild type clones, causing their extrusion from the basal epithelium[74]. *NOTCH1* mutant clones become progressively selected for in the normal oesophagus and this could explain why *NOTCH1* mutations are more frequent than in oesophageal cancer[39, 40]. As alluded to earlier, an intriguing speculation is that mutations in cancer drivers may have different roles and consequences depending on the context and time of alteration. In the case of *NOTCH1*, early mutations may create a decoy fitness peak that reduces the chances of malignant transformation[76, 77]. This variable role is further supported by the effect of conditional heterozygous deletion of somatic drivers in liver, including the two tumour suppressors *ARID1A* and *KMT2D*. Their deletion promotes liver regeneration and reduce damage susceptibility in the presence of injury[31]. Therefore, as seen by the reduction in the number of mutant clones in the lung of ex-smokers[45], the selective advantage of *NOTCH1*, *ARID1A* and *KMT2D* may be transient and context dependent.

It is likely that additional mechanisms also contribute to clone selection **(Figure 4B)**. Mutant cells can be extruded from the epithelium through the activation of cytoskeletal proteins in neighbouring cells, leading to apoptosis[78] or differentiation[79]. Moreover, cell-extrinsic factors, whose contribution has been investigated only

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

marginally, are also likely to support or hinder clone expansion. Active responses of the stromal niche surrounding the mutant cells, including a mesenchymal activation or a change in the composition of the extracellular matrix, may influence the expansion of certain clones and favour the clearance of others. For instance, increased mechanical stiffening of the extracellular matrix is thought to attenuate the extrusion of mutant cells from the epithelium[80].

Finally, the role of the immune system during clone expansion remains largely unknown. The immune system acts as an additional bottleneck during cancer evolution by exerting a selective pressure on cancer cells and shaping their immunogenicity[81]. Since it is now clear that mutant clones in normal tissues only rarely evolve into tumours, it is tempting to speculate that immunosurveillance starts well before cancer transformation. It may be that only non-immunogenic clones survive, while the others are eliminated by a concerted innate and adaptive immune response. Surviving clones may reach a dynamic equilibrium with the immune system that keeps their size at bay or may evolve immune evasion mechanisms to survive and continue to grow. It should be noted, however, that mutations in immune evasion genes are not under selection in normal skin[82].

When and how do mutant cells transform into cancer cells? The most striking differences between somatic and cancer clones are the number of mutated drivers and the extent of chromosomal instability. Somatic clones have at most two drivers and usually lack copy number alterations. In contrast, multiple drivers are needed for tumourigenesis[8] and chromosomal instability is often a hallmark of pre-cancer to cancer transition[83]. Therefore, a prerequisite for transformation may be the acquisition of multiple hits that may favour the onset of chromosomal instability.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

The order by which driver alterations are acquired is likely to be another required factor to promote transformation. Individuals with clonal haematopoiesis have higher risk to develop acute myeloid leukaemia if they bear *TP53* mutations compared to mutations in other genes[84]. Similarly, progressive mutations in *APC*, *KRAS* and *TP53* are paradigmatic of the adenoma to carcinoma transition in colon but are not observed in the normal colonic epithelium. Finally, the overall genotype of the mutant cell as well as the phenotype of the surrounding niche, including the interplay with the immune system, may decide the fate of the clone towards transformation.

Implications of somatic variation for cancer prevention and early detection

The accumulation of cancer driver mutations long before the appearance of cancer represents both opportunities and challenges for cancer prevention and early detection.

A better understanding of the endogenous and exogenous factors that trigger transformation of mutated but still normal cells into cancer cells has the potential to open avenues to improve or develop prevention strategies. For example, it could improve the sensitivity and specificity of cancer risk prediction algorithms, thus allowing clinicians to restrict cancer surveillance only to individuals at highest risk[85]. A deeper knowledge on the determinants of transformation could also point towards preventive therapies aimed at actively interrupting or at least delaying the carcinogenic process. Long-term use of aspirin has been associated with reduced risk of gastrointestinal cancers[86]. Although the molecular mechanism is not fully understood, the anti-inflammatory action of aspirin is probably a major component of its cancer-prevention effect. Similarly, the inhibition of the proinflammatory cytokine interleukin 1 has been proposed as a potential cancer preventive strategy[87].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

The pervasiveness of somatic clones also poses some challenges for cancer detection and monitoring. A prime example is liquid biopsy, a non-invasive approach increasingly used for tumour early detection and for monitoring response to therapy[88]. Liquid biopsy is based on the identification of circulating DNA fragments bearing driver mutations, which are usually thought to derive from dead cancer cells or extracellular vesicles. Circulating DNA from mutated but normal cells can act as a confounding factor particularly in old patients who are likely to bear a high number of mutant clones and experience age-induced cell death. For example, *TP53* mutations were detected in the circulating DNA of 49% lung cancer patients but also in 11% non-cancer controls[89]. Circulating fragments of mutated DNA in healthy individuals can derive from clones originally resident in solid tissues or, more often, from mutant blood cells. Clonal haematopoiesis is a known source of noise in liquid biopsy[90, 91], but it can be efficiently accounted for through the parallel sequencing of matched leukocyte DNA[92].

Circulating DNA of mutant cells from solid tissues is more difficult to distinguish. In this case, focusing on cancer methylation patterns in addition to mutations, as in the case of the GRAIL test[93], could improve the test specificity. Overall, however, the performance of liquid biopsy in detecting early stage cancer is poorer than for advanced disease[91].

Additional issues concern the definition of cancer driver genes and how they can be efficiently distinguished from mutated genes that increase cell fitness but do not drive tumourigenesis. Identifying cancer drivers based uniquely on their recurrence across samples will lead to false positives when the mutations were inherited from the normal progenitors and play no role in tumourigenesis[16]. In this case, deep sequencing of normal tissues surrounding the tumour and a patient-level rather than a cohort-level approach to driver identification may help.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 July 2022 doi:10.20944/preprints202207.0321.v1

FUTURE PERSPECTIVES

The ability to precisely quantify the extent of genomic variation occurring in seemingly normal tissues is radically changing our understanding of somatic evolution. The idea of a stable genome inherited from germline cells and maintained strictly unaltered throughout adult life does not hold true. Rather, the genome of somatic cells undergoes continuous modifications, some of which confer fitness advantages that can initiate clonal expansion. This results in dynamic tissue remodelling that starts during embryo development, where it is mostly driven by neutral drift, and continues as we age, where the fittest clones undergo positive selection.

The long-term fate of somatic clones depends on the interplay between the intrinsic features of the host tissue and the extrinsic features of the surrounding ecosystem, which are likely to change over time. Ending the exposure to damage and stress may reduce the fitness of previously selected clones, causing their shrinkage and clearance. Alternatively, clones may persist for a long time in equilibrium with the surrounding ecosystem. The disruption of this equilibrium may result in gaining transforming capacity.

Currently, very little is known on what regulates the homeostatic equilibrium within tissues, and this limits our understanding of the initial phases of cancer initiation and the efficacy of early clinical intervention. Further studies are needed to define the functional activity of driver genes in different contexts, including the role of epigenetic alterations and mutations in non-coding regions during somatic evolution. Moreover, a detailed knowledge of the functional composition of the niche surrounding mutant clones will reveal key extrinsic factors supporting their survival. Finally, new model systems are needed to follow the fate of mutant clones exposed to changing

conditions. Addressing these fundamental questions will advance novel cancer detection and prevention programs.

ACKNOWLEDGEMENTS

We thank Joe Brock for his help with the figures and Kalum Clayton for critical reading of the manuscript.

FUNDING

This work was supported by Cancer Research UK [C43634/A25487 to F.D.C.] and [EDDPJT-Nov21\100010 to F.D.C], the Cancer Research UK City of London Centre [C7893/A26233 to F.D.C], innovation programme under the Marie Skłodowska-Curie grant agreement No CONTRA-766030 and the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001002), the UK Medical Research Council (FC001002), and the Wellcome Trust (FC001002). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

DISCLOSURE

The authors have declared no conflicts of interest.

REFERENCES

1 Nowell PC. The clonal evolution of tumor cell populations. Science 1976; 194 (4260): 23-28.
2 Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990; 61 (5): 759-767.
3 Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature 2009; 458 (7239): 719-724.
4 Dressler L, Bortolomeazzi M, Keddar MR et al. Comparative assessment of genes driving cancer and somatic evolution in non-cancer tissues: an update of the Network of Cancer Genes (NCG) resource. Genome Biol 2022; 23 (1): 35.
5 Sondka Z, Bamford S, Cole CG et al. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer 2018; 18 (11): 696-705.
6 Porta-Pardo E, Kamburov A, Tamborero D et al. Comparison of algorithms for the detection of cancer drivers at subgene resolution. Nat Methods 2017; 14 (8): 782-788.
7 Bailey MH, Tokheim C, Porta-Pardo E et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell 2018; 174 (4): 1034-1035.
8 Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer 1954; 8 (1): 1-12.
9 Schwartz R, Schaffer AA. The evolution of tumour phylogenetics: principles and practice. Nat Rev Genet 2017; 18 (4): 213-229.
10 Bailey C, Black JRM, Reading JL et al. Tracking Cancer Evolution through the Disease Course. Cancer Discov 2021; 11 (4): 916-932.
11 Taylor RW, Barron MJ, Borthwick GM et al. Mitochondrial DNA mutations in human colonic crypt stem cells. J Clin Invest 2003; 112 (9): 1351-1360.
12 Fellous TG, McDonald SA, Burkert J et al. A methodological approach to tracing cell lineage in human epithelial tissues. Stem Cells 2009; 27 (6): 1410-1420.
13 Jonason AS, Kunala S, Price GJ et al. Frequent clones of p53-mutated keratinocytes in normal human skin. Proc Natl Acad Sci U S A 1996; 93 (24): 14025-14029.
14 Tian D, Feng Z, Hanley NM et al. Multifocal accumulation of p53 protein in esophageal carcinoma: evidence for field cancerization. Int J Cancer 1998; 78 (5): 568-575.
15 Kakiuchi N, Ogawa S. Clonal expansion in non-cancer tissues. Nat Rev Cancer 2021; 21 (4): 239-256.
16 Wijewardhane N, Dressler L, Ciccarelli FD. Normal Somatic Mutations in Cancer Transformation. Cancer Cell 2021; 39 (2): 125-129.
17 Genovese G, Kahler AK, Handsaker RE et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med 2014; 371 (26): 2477-2487.
18 Jaiswal S, Fontanillas P, Flannick J et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med 2014; 371 (26): 2488-2498.
19 Xie M, Lu C, Wang J et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med 2014; 20 (12): 1472-1478.
20 Bowden DH. Cell turnover in the lung. Am Rev Respir Dis 1983; 128 (2 Pt 2): S46-48.
21 Pan Q, Nicholson AM, Barr H et al. Identification of lineage-uncommitted, long-lived, label-retaining cells in healthy human esophagus and stomach, and in metaplastic esophagus. Gastroenterology 2013; 144 (4): 761-770.
22 Cousins FL, Pandoy R, Jin S et al. The Elusive Endometrial Epithelial Stem/Progenitor Cells. Front Cell Dev Biol 2021; 9: 640319.
23 Khandelwal P, Abraham SN, Apodaca G. Cell biology and physiology of the uroepithelium. Am J Physiol Renal Physiol 2009; 297 (6): F1477-1501.
24 Milo R, Jorgensen P, Moran U et al. BioNumbers--the database of key numbers in molecular and cell biology. Nucleic Acids Res 2010; 38 (Database issue): D750-753.
25 Blokzijl F, de Ligt J, Jager M et al. Tissue-specific mutation accumulation in human adult stem cells during life. Nature 2016; 538 (7624): 260-264.

26 Lee-Six H, Olafsson S, Ellis P et al. The landscape of somatic mutation in normal colorectal epithelial cells. Nature 2019; 574 (7779): 532-537.
27 Kakiuchi N, Yoshida K, Uchino M et al. Frequent mutations that converge on the NFKBIZ pathway in ulcerative colitis. Nature 2020; 577 (7789): 260-265.
28 Nanki K, Fujii M, Shimokawa M et al. Somatic inflammatory gene mutations in human ulcerative colitis epithelium. Nature 2020; 577 (7789): 254-259.
29 Robinson PS, Coorens THH, Palles C et al. Increased somatic mutation burdens in normal human cells due to defective DNA polymerases. Nat Genet 2021; 53 (10): 1434-1442.
30 Brunner SF, Roberts ND, Wylie LA et al. Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. Nature 2019; 574 (7779): 538-542.
31 Zhu M, Lu T, Jia Y et al. Somatic Mutations Increase Hepatic Clonal Fitness and Regeneration in Chronic Liver Disease. Cell 2019; 177 (3): 608-621 e612.
32 Ng SW, Rouhani FJ, Brunner SF et al. Convergent somatic mutations in metabolism genes in chronic liver disease. Nature 2021; 598 (7881): 473-478.
33 Anglesio MS, Papadopoulos N, Ayhan A et al. Cancer-Associated Mutations in Endometriosis without Cancer. N Engl J Med 2017; 376 (19): 1835-1848.
34 Lac V, Verhoef L, Aguirre-Hernandez R et al. Iatrogenic endometriosis harbors somatic cancer-driver mutations. Hum Reprod 2019; 34 (1): 69-78.
35 Suda K, Nakaoka H, Yoshihara K et al. Clonal Expansion and Diversification of Cancer-Associated Mutations in Endometriosis and Normal Endometrium. Cell Rep 2018; 24 (7): 1777-1789.
36 Lac V, Nazeran TM, Tessier-Cloutier B et al. Oncogenic mutations in histologically normal endometrium: the new normal? The Journal of pathology 2019; 249 (2): 173-181.
37 Moore L, Leongamornlert D, Coorens THH et al. The mutational landscape of normal human endometrial epithelium. Nature 2020; 580 (7805): 640-646.
38 Yamaguchi M, Nakaoka H, Suda K et al. Spatiotemporal dynamics of clonal selection and diversification in normal endometrial epithelium. Nat Commun 2022; 13 (1): 943.
39 Martincorena I, Fowler JC, Wabik A et al. Somatic mutant clones colonize the human esophagus with age. Science 2018; 362 (6417): 911-917.
40 Yokoyama A, Kakiuchi N, Yoshizato T et al. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. Nature 2019; 565 (7739): 312-317.
41 Fowler JC, King C, Bryant C et al. Selection of oncogenic mutant clones in normal human skin varies with body site. Cancer Discov 2021; 11 (2): 340-361.
42 Martincorena I, Roshan A, Gerstung M et al. High burden and pervasive positive selection of somatic mutations in normal human skin. Science 2015; 348 (6237): 880-886.
43 Saini N, Giacobone CK, Klimczak LJ et al. UV-exposure, endogenous DNA damage, and DNA replication errors shape the spectra of genome changes in human skin. PLoS Genet 2021; 17 (1): e1009302.
44 Tang J, Fewings E, Chang D et al. The genomic landscapes of individual melanocytes from human skin. Nature 2020; 586 (7830): 600-605.
45 Yoshida K, Gowers KHC, Lee-Six H et al. Tobacco smoking and somatic mutations in human bronchial epithelium. Nature 2020; 578 (7794): 266-272.
46 Lawson ARJ, Abascal F, Coorens THH et al. Extensive heterogeneity in somatic mutation and selection in the human bladder. Science 2020; 370 (6512): 75-82.
47 Li R, Du Y, Chen Z et al. Macroscopic somatic clonal expansion in morphologically normal human urothelium. Science 2020; 370 (6512): 82-89.
48 Chen T, Dent SY. Chromatin modifiers and remodellers: regulators of cellular differentiation. Nat Rev Genet 2014; 15 (2): 93-106.
49 Park S, Mali NM, Kim R et al. Clonal dynamics in early human embryogenesis inferred from somatic mutation. Nature 2021; 597 (7876): 393-397.
50 Coorens THH, Moore L, Robinson PS et al. Extensive phylogenies of human development inferred from somatic mutations. Nature 2021; 597 (7876): 387-392.

51 Jassal B, Matthews L, Viteri G et al. The reactome pathway knowledgebase. Nucleic Acids Res 2020; 48 (D1): D498-D503.
52 Hasaart KAL, Manders F, van der Hoorn ML et al. Mutation accumulation and developmental lineages in normal and Down syndrome human fetal haematopoiesis. Sci Rep 2020; 10 (1): 12991.
53 Spencer Chapman M, Ranzoni AM, Myers B et al. Lineage tracing of human development through somatic mutations. Nature 2021; 595 (7865): 85-90.
54 Bae T, Tomasini L, Mariani J et al. Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. Science 2018; 359 (6375): 550-555.
55 Kuijk E, Blokzijl F, Jager M et al. Early divergence of mutational processes in human fetal tissues. Sci Adv 2019; 5 (5): eaaw1271.
56 Ju YS, Martincorena I, Gerstung M et al. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. Nature 2017; 543 (7647): 714-718.
57 Schulz KN, Harrison MM. Mechanisms regulating zygotic genome activation. Nat Rev Genet 2019; 20 (4): 221-234.
58 Palmer N, Kaldis P. Regulation of the Embryonic Cell Cycle During Mammalian Preimplantation Development. Curr Top Dev Biol 2016; 120: 1-53.
59 Vazquez-Diez C, FitzHarris G. Causes and consequences of chromosome segregation error in preimplantation embryos. Reproduction 2018; 155 (1): R63-R76.
60 McCoy RC. Mosaicism in Preimplantation Human Embryos: When Chromosomal Abnormalities Are the Norm. Trends Genet 2017; 33 (7): 448-463.
61 Lodato MA, Rodin RE, Bohrson CL et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. Science 2018; 359 (6375): 555-559.
62 Abascal F, Harvey LMR, Mitchell E et al. Somatic mutation landscapes at single-molecule resolution. Nature 2021; 593 (7859): 405-410.
63 Gonitel R, Moffitt H, Sathasivam K et al. DNA instability in postmitotic neurons. Proceedings of the National Academy of Sciences 2008; 105 (9): 3467-3472.
64 Luquette LJ, Miller MB, Zhou Z et al. Ultraspecific somatic SNV and indel detection in single neurons using primary template-directed amplification. bioRxiv 2021: 2021.2004.2030.442032.
65 Erickson RP. Somatic gene mutation and human disease other than cancer. Mutat Res 2003; 543 (2): 125-136.
66 Erickson RP. Somatic gene mutation and human disease other than cancer: an update. Mutat Res 2010; 705 (2): 96-106.
67 Lim ET, Uddin M, De Rubeis S et al. Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. Nat Neurosci 2017; 20 (9): 1217-1224.
68 Tall AR, Fuster JJ. Clonal hematopoiesis in cardiovascular disease and therapeutic implications. Nature Cardiovascular Research 2022; 1 (2): 116-124.
69 Van Horebeek L, Dubois B, Goris A. Somatic Variants: New Kids on the Block in Human Immunogenetics. Trends Genet 2019; 35 (12): 935-947.
70 Szilard L. On the Nature of the Aging Process. Proc Natl Acad Sci U S A 1959; 45 (1): 30-45.
71 Lopez-Otin C, Blasco MA, Partridge L et al. The hallmarks of aging. Cell 2013; 153 (6): 1194-1217.
72 Vijg J, Dong X. Pathogenic Mechanisms of Somatic Mutation and Genome Mosaicism in Aging. Cell 2020; 182 (1): 12-23.
73 Levy O, Amit G, Vaknin D et al. Age-related loss of gene-to-gene transcriptional coordination among single cells. Nature Metabolism 2020; 2 (11): 1305-1315.
74 Colom B, Herms A, Hall MWJ et al. Mutant clones in normal epithelium outcompete and eliminate emerging tumours. Nature 2021; 598 (7881): 510-514.
75 Colom B, Alcolea MP, Piedrafita G et al. Spatial competition shapes the dynamic mutational landscape of normal esophageal epithelium. Nat Genet 2020; 52 (6): 604-614.

76 Laconi E, Marongiu F, DeGregori J. Cancer as a disease of old age: changing mutational and microenvironmental landscapes. Br J Cancer 2020; 122 (7): 943-952.
77 Higa KC, DeGregori J. Decoy fitness peaks, tumor suppression, and aging. Aging Cell 2019; 18 (3): e12938.
78 Tanimura N, Fujita Y. Epithelial defense against cancer (EDAC). Semin Cancer Biol 2020; 63: 44-48.
79 van Neerven SM, de Groot NE, Nijman LE et al. Apc-mutant cells act as supercompetitors in intestinal tumour initiation. Nature 2021; 594 (7863): 436-441.
80 Pothapragada SP, Gupta P, Mukherjee S et al. Matrix mechanics regulates epithelial defence against cancer by tuning dynamic localization of filamin. Nat Commun 2022; 13 (1): 218.
81 Dunn GP, Bruce AT, Ikeda H et al. Cancer immunoediting: from immunosurveillance to tumor escape. Nat Immunol 2002; 3 (11): 991-998.
82 Luijts T, Elliott K, Siaw JT et al. A clinically annotated post-mortem approach to study multi-organ somatic mutational clonality in histologically healthy tissues. bioRxiv 2022: 2022.2002.2027.482170.
83 Dou Y, Gold HD, Luquette LJ et al. Detecting Somatic Mutations in Normal Cells. Trends Genet 2018; 34 (7): 545-557.
84 Abelson S, Collord G, Ng SWK et al. Prediction of acute myeloid leukaemia risk in healthy individuals. Nature 2018; 559 (7714): 400-404.
85 Fitzgerald RC, Antoniou AC, Fruk L et al. The future of early cancer detection. Nat Med 2022; 28 (4): 666-677.
86 Rothwell PM, Fowkes FG, Belch JF et al. Effect of daily aspirin on long-term risk of death due to cancer: analysis of individual patient data from randomised trials. Lancet 2011; 377 (9759): 31-41.
87 Garlanda C, Mantovani A. Interleukin-1 in tumor progression, therapy, and prevention. Cancer Cell 2021; 39 (8): 1023-1027.
88 Cescon DW, Bratman SV, Chan SM et al. Circulating tumor DNA and liquid biopsy in oncology. Nat Cancer 2020; 1 (3): 276-290.
89 Fernandez-Cuesta L, Perdomo S, Avogbe PH et al. Identification of Circulating Tumor DNA for the Early Detection of Small-cell Lung Cancer. EBioMedicine 2016; 10: 117-123.
90 Hu Y, Ulrich BC, Supplee J et al. False-Positive Plasma Genotyping Due to Clonal Hematopoiesis. Clin Cancer Res 2018; 24 (18): 4437-4443.
91 Liu MC, Oxnard GR, Klein EA et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann Oncol 2020; 31 (6): 745-759.
92 Razavi P, Li BT, Brown DN et al. High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. Nat Med 2019; 25 (12): 1928-1937.
93 Klein EA, Richards D, Cohn A et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Ann Oncol 2021; 32 (9): 1167-1177.