

Perspective

Pathogenic Mechanisms of Somatic Mutation and Genome Mosaicism in Aging

Jan Vijg^{1,2,*} and Xiao Dong¹¹Department of Genetics, Albert Einstein College of Medicine, New York, NY 10461, USA²Center for Single-Cell Omics, School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China*Correspondence: jan.vijg@einsteinmed.org<https://doi.org/10.1016/j.cell.2020.06.024>

Age-related accumulation of postzygotic DNA mutations results in tissue genetic heterogeneity known as somatic mosaicism. Although implicated in aging as early as the 1950s, somatic mutations in normal tissue have been difficult to study because of their low allele fractions. With the recent emergence of cost-effective high-throughput sequencing down to the single-cell level, enormous progress has been made in our capability to quantitatively analyze somatic mutations in human tissue in relation to aging and disease. Here we first review how recent technological progress has opened up this field, providing the first broad sets of quantitative information on somatic mutations *in vivo* necessary to gain insight into their possible causal role in human aging and disease. We then propose three major mechanisms that can lead from accumulated *de novo* mutations across tissues to cell functional loss and human disease.

Genome mosaicism of the soma refers to genetic heterogeneity within tissues and organs because of postzygotic mutations; i.e., mutations in the genome of somatic cells not transmitted to the next generation. The concept of genome mosaicism stems from the early 20th century, when [Boveri \(1914\)](#), [Morgan and Bridges \(1919\)](#), and others speculated that cancer was a somatic mosaic caused by genetic alterations. Mutations were then generally understood to be sources of genetic variation that alter the physical and functional units of heredity, first termed “genes” by [Johannsen \(1911\)](#). Later, based on new insights into the mutagenic effects of radiation on mammals, [Failla \(1958\)](#) and [Szilard \(1959\)](#) proposed that cancer and aging were caused by accumulation of *de novo* somatic mutations. Almost two decades later, this proved to be correct for cancer, a disease caused by cycles of mutation, selection, and new mutations, which can eventually result in a metastatic tumor ([Nowell, 1976](#)). However, a causal role of postzygotic mutations and genome mosaicism in aging and age-related diseases other than cancer proved more difficult to establish and remains unclear. This is because postzygotic mutations are generally unique for each individual cell and can only be detected when they clonally amplify to a large allelic fraction. The latter is rare, which explains why it was not noticed until the era of large-scale genotyping began.

Somatic mutations arise because of errors in repair or replication of damaged DNA ([Figure 1](#)). In the germline, such errors generate genetic diversity by providing the substrate for natural selection, fueling evolution. The germline mutation rate is itself genetically determined and depends on various factors, including environmental constraints and population size. Mutation rates cannot decrease to zero because this would prevent any subsequent evolutionary change and would have fitness costs for the organism in terms of the advanced repair and replication systems required ([Baer et al., 2007](#); [Lynch et al., 2016](#)). In multicellular organisms, mutation rate of the soma may also vary,

and there is some evidence of increased somatic genome maintenance in long-lived species ([Vijg, 2007](#)). This would suggest that long-lived species have a lower somatic mutation rate than short-lived species, but this remains to be demonstrated. Of course, evolutionary theory does not predict prevention of somatic mutations for extended time periods after reproduction ([Williams, 1957](#)). In addition, somatic mutagenesis also fulfills certain essential physiological roles ([Box 1](#)).

Until very recently, almost all available information on *de novo* mutagenesis in normal cells and tissues was obtained indirectly using reporter genes (endogenous genes or transgenes that can be selected for mutations) ([Dollé et al., 1997](#); [Grist et al., 1992](#)). In virtually all of these studies, a significant and often tissue-specific increase in somatic mutations with age has been observed ([Dollé et al., 1997](#); [Ono et al., 2000](#)). However, although informative, such mutational reporter models have the major limitation that no reporter gene can be considered to represent an entire mammalian genome. Hence, they do not allow obtaining quantitative information about mutation loads across the genome.

Based on the dramatic progress in next-generation sequencing over the last few years, our knowledge regarding the frequency and type of somatic mutations in human tissues during aging has increased enormously. In not much more than half a decade, more information about somatic mutations in human and animal tissues and their accumulation as a function of age has been obtained than in the previous half century. Hence, the time has come to discuss strategies to translate the technological breakthroughs that are now leading to a treasure trove of datasets on *in vivo* somatic mutations into a mechanistic understanding of their possible causal relationship with age-related cell functional loss and chronic disease. Here we first summarize the status of the field on the hand of the technological advances that enabled this great progress and then discuss the

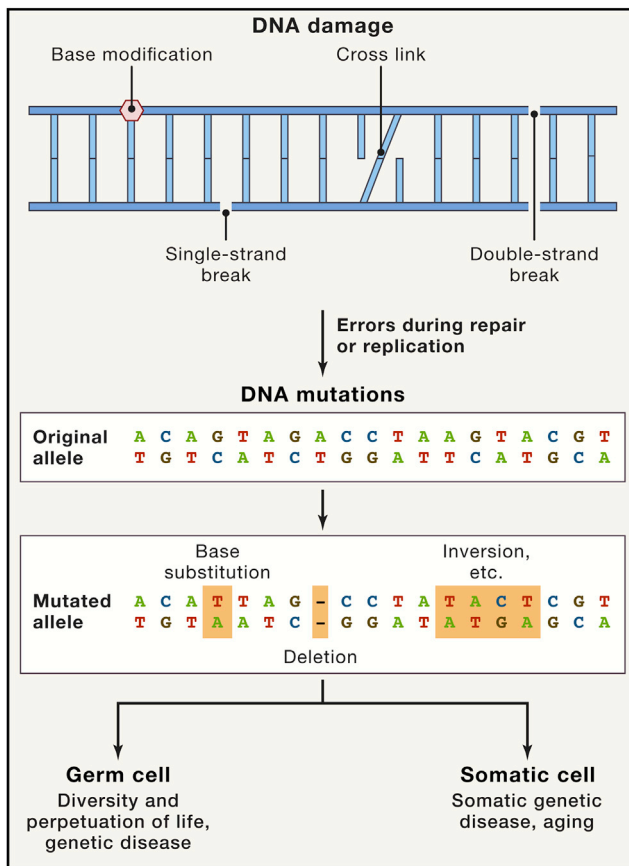


Figure 1. Causes and Consequences of Somatic DNA Mutations
Different types of DNA damage occur daily in every cell of an organism, including strand breaks, base modifications, and crosslinks. This damage is repaired rapidly through a large variety of pathways. However, errors that occur during DNA repair and replication result in *de novo* mutations, which range from single-nucleotide variants (SNVs) to chromosomal aberrations. In contrast to DNA damage, mutations are irreversible and, therefore, inevitably accumulate over time. When mutations occur in germ cells, they become substrates for evolution, giving rise to diverse life forms and genetic disease. When they occur in the soma, mutations can result in somatic genetic diseases and possibly aging.

possible mechanisms that can lead from random somatic mutations to age-related cellular degeneration and disease.

How Technological Advances Led to the First Accurate Account of Genome Mosaicism

Mutations arise in a stochastic fashion in the different cells of a tissue. In contrast to germline mutations, which are copied from the fertilized egg to all somatic cells, postzygotic mutations result in genetic heterogeneity within a tissue or cell population. The vast majority of postzygotic mutations are unique to each cell or recurrently present only in very small numbers of cells. This makes their detection a major challenge, which explains the very limited progress in our understanding of somatic mutations, their frequency, how they are generated, their possible tissue specificity, and their relationship to human aging and disease. The exception is cancer because tumors are essentially clones and, therefore, serve as surrogates for the single cells

from which they arose. In this section, we discuss the technological advances that made it possible to characterize the somatic mosaicism resulting from postzygotic mutagenesis in human tissues in relation to aging and disease, including diseases other than cancer.

Genome mosaicism and its increased occurrence in older organisms first became apparent with the development of cytogenetic analysis. Using simple Giemsa staining of metaphase spreads, later replaced by fluorescence *in situ* hybridization (FISH), chromosomal abnormalities were discovered in individual somatic cells and later found to increase with age in human blood cells (Jacobs et al., 1964) and regenerating mouse liver (Crowley and Curtis, 1963).

The most widely studied age-related chromosomal abnormality that gives rise to somatic genome mosaicism, reported early in the 1970s from studying metaphases from human blood lymphocytes (Jacobs et al., 1963) and bone marrow (Pierre and Hoagland, 1972), is mosaic loss of the Y chromosome (LOY) in males during aging, which has now been widely confirmed with more advanced technology. LOY is defined as a lower-than-expected abundance of DNA from the Y chromosome with a certain threshold of detection, for example, as 10% or more of affected cells (Dumanski et al., 2016). In a recent study of 205,011 men from the UK Biobank, LOY was found to affect from 2.5% of men at age 40 to 43.6% at age 70, which makes it the most common *de novo* somatic mutation over the human lifetime (Thompson et al., 2019). LOY frequency has been associated with a shorter lifespan, a higher risk of cancer, smoking, Alzheimer's disease, cardiovascular disease, diabetes, immune deficiencies, and other age-related diseases (Dumanski et al., 2016; Loftfield et al., 2018; Thompson et al., 2019). LOY has a genetic component, and in the aforementioned UK Biobank study, more than 150 autosomal genetic determinants of LOY were identified in the male cohort. LOY is most likely a general biomarker for genome instability in somatic cells. Indeed, the loci found to be genetically associated with LOY in males were themselves genetically associated, in a female cohort, with female cancers (breast, ovarian, and endometrial cancer) and age at natural menopause (Thompson et al., 2019). Of note, early menopause has been genetically associated with DNA damage response (DDR) genes (Day et al., 2015). Based on these results, it is tempting to speculate that the association of LOY with a diverse series of age-related pathologies points toward a causal role of somatic mutations in aging and age-related disease.

Although it is the most common postzygotic mutation, LOY is only one among many. With the emergence of large-scale genotyping technology about a decade ago, genome mosaicism was found to occur at the sub-chromosomal level across the genome in blood. For example, using SNP arrays, mosaic copy number variations (CNVs) were found in 0.5% of young individuals and 2%–3% of older people (Laurie et al., 2012). With the emergence of high-depth next-generation sequencing of specific genes or whole exomes, a general pattern of genome mosaicism was uncovered in blood, called clonal hematopoiesis. For example, in blood, genome mosaicism for specific mutations has been found at loci frequently mutated in hematologic cancers (Jaiswal et al., 2014). In all of these cases, the frequency of expanded variants increased with age and was found associated with hematologic

Box 1. Physiological Role of Somatic Mutations

Somatic mutagenesis fulfills some important physiological roles. For example, somatic hypermutation generates immunological diversity to produce high-affinity antibodies through accumulation of point mutations at a very high rate in the V-regions of the immunoglobulin heavy and light chains of B lymphocytes (Odegard and Schatz, 2006). Somatic hypermutation occurs during differentiation in the germinal centers of the lymphoid organs and is mediated by activation-induced cytidine deaminase (AID), which catalyzes targeted deamination of deoxycytidine residues in DNA, turning them into uracil. This leads to mutations at C:G and neighboring A:T base pairs (Di Noia and Neuberger, 2007; Zhang et al., 2019). This process allows vertebrates to produce a vast repertoire of antibody molecules to combat infection.

Somatic mutations might also have important roles in the brain (Paquola et al., 2017) and liver (Duncan et al., 2012). In the brain, most neuronal populations are postmitotic and cannot be renewed, with the exceptions of the hippocampus and subventricular zone (Gage, 2019). Somatic mutations generated early in development would therefore remain present until old age, generating individual physiological diversity in brain function.

The liver has a very high frequency of aneuploidy, possibly up to 50% of all cells (Duncan et al., 2012), and can regenerate, probably because it is heavily exposed to xenobiotics. This means that it can select specific aneuploid karyotypes to adapt to chronic liver injury. Recently, *de novo* mutations in specific genes have been found to be enriched for in the liver of individuals with cirrhosis, including in *Pkd1*, *Kmt2d*, and *Arid1a*, which were subsequently demonstrated to promote hepatic clonal fitness and regeneration in mice (Zhu et al., 2019).

cancers, increased mortality, increased risk of type 2 diabetes, atherosclerosis, coronary heart disease, and ischemic stroke (Jaiswal et al., 2014).

More recently still, ultra-deep sequencing of DNA or RNA from multiple human tissues showed that somatic mutations leading to genome mosaicism do not only occur in blood but are a widespread phenomenon. For example, in numerous small samples taken across a grid of normal esophageal epithelium from nine human organ transplant donors of different ages, ultra-deep (>800-fold) targeted sequencing of known cancer driver genes revealed mutant mosaics that reflect the clonal outgrowth of *de novo* mutations, with the number of such mutations increasing significantly with age (Martincorena et al., 2018). Similar results were obtained by Yokoyama et al. (2019). Finally, clonal expansion of somatic mutations has been detected by RNA sequence analysis of normal tissues in the Genotype-Tissue Expression (GTEx) dataset (collected from about 30 non-diseased tissue types across nearly 500 individuals; <https://www.gtexportal.org/home/>). The results indicate the expected age-related increase in mutations, the association of the number of mutations with the tissue-specific proliferation rate, and the enrichment of cancer driver mutations (Yizhak et al., 2019).

From these findings, it can be concluded that genome mosaicism in blood and in other tissues, ranging from LOY in males to large deletions across the genome and point mutations in genes, is common, with the frequency of expanded variants increasing with age. These findings also confirm the aforementioned age-related increase in mutation frequency observed using reporter systems in humans and mice. They also show that *de novo* mutations frequently expand into clonal lineages, allowing their detection in DNA from bulk tissue.

If aging and age-related diseases occur as a consequence of losing genome sequence integrity, it raises the question of how such a random process can lead to the specific pathophysiology of so many different diseases. This question cannot be easily addressed by studying mutant clones in bulk cell populations. Indeed, ultra-deep sequencing would only uncover the tip of an iceberg of accumulating mutations that occur independently

in every cell. To understand whether and how increased mutation loads have a functional effect and predispose to multiple diseases and intrinsic aging, we need to analyze single cells or clones grown out of a single cell.

A number of studies have now used *in vitro* clonal expansion of single cells followed by sequencing to avoid the selective view generated by sequencing bulk DNA for clonal mosaics. This approach has generated complete landscapes of somatic mutations from progenitor cells of the small intestine, liver, and colon (Blokzijl et al., 2016), from neuronal stem cells (Bae et al., 2018) and skeletal muscle satellite cells (Franco et al., 2018), and from bronchial epithelial cells (Yoshida et al., 2020), kidney tubular progenitor cells, fat progenitor cells, and epithelial progenitor cells (Franco et al., 2019) of human donors as a function of age. This approach has established the total mutation loads of human somatic cells as varying from several hundred base substitutions per genome to well over 3,000, depending on the cell type and donor age. These data have also provided an unselected, comprehensive view of the mutational spectra and signatures of these normal tissues, which, in turn, provides insights into the mutagen exposure of these tissues. Most importantly, this approach has enabled study of the distribution of mutations across the same genome and their possible interactions, which is critically important when analyzing their potential collective functional effect.

Although highly informative, quantitative analysis of somatic mutation landscapes in single cells clonally expanded *in vitro* suffers from an important limitation: *de novo* mutations in postmitotic cells remain out of reach. Moreover, by its very nature, this approach exclusively analyzes mutations in stem or progenitor cells, which generally have a lower mutation rate than differentiated cells. This is best illustrated by a recent direct comparison of human liver stem cells with fully differentiated hepatocytes, showing an approximately 2-fold higher average mutation frequency in the latter (Brazhnik et al., 2020). In addition, fully differentiated cells also displayed a much higher cell-to-cell variation in mutation frequencies, which is probably a reflection of their position in the hierarchy of cellular differentiation.

Direct sequencing of single cells without the selection involved in clonal outgrowth *in vitro* is the method of choice because it can be applied to every cell type, including postmitotic cell types. However, single-cell methods require whole-genome amplification, which is fraught with error. Great progress has recently been made to overcome amplification-induced errors by modification of whole-genome amplification methods and introduction of computational ways of detecting artifacts and filtering them out (Bohrson et al., 2019; Chen et al., 2017; Dong et al., 2017; Zong et al., 2012). These improved methods showed that the total load of *de novo* somatic base substitution mutations in different normal human tissues appeared to be between several hundred and well over 5,000, depending on the tissue, the individual cell, and the age of the donor (Brazhnik et al., 2020; Lodato et al., 2018; Zhang et al., 2019; for a review, see Zhang and Vijg, 2018).

So far, all sequencing data, from somatic clonal expansion to single cells, yielded information about single-nucleotide variants (SNVs) and, occasionally, small insertions and deletions (INDELs). SNVs are not likely to cause functional changes except non-synonymous SNVs in protein-coding regions. Hence, these mutations are found in fairly large numbers in normal cells (see above). Small INDELs, most of them only 1-bp INDELs, can be detected using the same principle as for SNV detection. Their frequency is typically one-tenth of SNVs in somatic cells (Franco et al., 2018). The mutations most likely to be deleterious are genome structural variations (SVs); i.e., deletions, insertions, inversions and rearrangements (such as interchromosomal and intrachromosomal translocations), mobile element insertions (such as retrotranspositions), and loss of telomere repeats at chromosome ends. SVs, whose size can vary enormously from dozens of base pairs to whole chromosomes, readily affect cellular function because of their capacity to rearrange gene-regulatory sequences. This is reflected by the much greater role of SVs than of SNVs in the diversity and evolution of human genomes, something that was already realized in the 1970s (King and Wilson, 1975).

Because of their high functional effect, SVs are now studied extensively. So far, this has been mainly limited to germline variation, and it seems likely that an average individual may differ from another by well over 10,000 SVs (Audano et al., 2019). Although that is much less than the four to five million SNVs between two individuals, they could collectively affect as many as 20 million bp of sequence (Auton et al., 2015). We know very little about SVs in normal somatic cells. Similar to SVs in the germline genome, the functional effect of SVs in the somatic genome is likely to be much higher than that of SNVs.

Single-cell sequencing data are available for only two types of SVs, CNVs and retrotranspositions; the frequency of both appears to be limited to a few events per cell (Cai et al., 2014; Evrony et al., 2016; McConnell et al., 2013). So far, there are no single-cell data on other forms of SVs (such as translocations, inversions, and large deletions or insertions). It remains technically very difficult to derive accurate SV frequencies from whole-genome sequencing data, even in the case of bulk DNA (Kosugi et al., 2019). However, large genome rearrangements have been shown to increase with age in tissues such as the heart and liver of mice using the reporter gene approach (Dollé

et al., 1997) and in human and mouse lymphocytes using FISH (see above). Importantly, using a panel of 18 rodent species with diverse lifespans, it has been shown recently that more robust DNA double-strand break (DSB) repair coevolved with longevity (Tian et al., 2019). Because SVs are mostly a consequence of erroneous repair of DNA DSBs, it is tempting to suggest that the rate of SV accumulation is an important factor in determining the aging rate.

In summary, based on groundbreaking technological advances, we are now able to extensively study somatic mutagenesis and its resulting genome mosaicism in essentially any human cell type. This is especially true for base substitutions, which can now be accurately called in single cells. Other types of mutations, most notably SVs, still require new and improved analytical approaches. Although this is still ongoing, we can already begin to address the next question: what, if any, is the functional effect of somatic mutations and genome mosaicism? Observations of reduced proliferative capacity of primary cells with increased mutation burden (Franco et al., 2018) and selection against age-related accumulation of mutations in the functional part of the genome (Zhang et al., 2019) suggest that somatic mutations could be causally related to aging. Below we discuss the mechanisms that could underlie the pathogenic effects of somatic mutations and genome mosaicism.

Pathogenic Mechanisms

The main argument against a causal role of random somatic mutations in aging and aging-associated disease has been that the spontaneous mutation frequency, even at an old age, is too low to impair cellular function. The exception is cancer, where particular driver mutations are selected for a growth advantage. Mutation frequencies in somatic cells have been considered to be low because estimates were based on the mutation frequencies observed in the germline of various species, including humans, as deduced from heritable changes in proteins. Indeed, a germline mutation frequency of about 1×10^{-8} per base pair per generation was found for humans (Drake et al., 1998), which proved to be almost right when assessed later by whole-genome sequencing of parents and offspring, where all mutations found in the latter but not in the genome of the parents (about 60 new SNVs) were used to calculate the mutation rate per generation (Kong et al., 2012). In this case, the mutation rate is almost entirely determined by males, and particularly older males, because of the much larger number of germline cell divisions in sperm than in oocytes.

However, as discussed above, the new single-cell methods found many more mutations per cell in the soma; i.e., up to several thousands of SNVs, depending on the age of the subject and the cell type. This suggests that the somatic mutation rate is higher than the germline mutation rate. Indeed, in a direct comparison of germline and somatic mutation rates in humans and mice, the somatic mutation rate was found to be almost 2 orders of magnitude more frequent than the germline mutation rate (Miholland et al., 2017). To some extent, this can be explained by selection against deleterious mutations in the germline. However, most random mutations have no effect. Indeed, mutation frequencies in single sperm have also been found to be low (Wang et al., 2012), albeit about 3 times as high as those

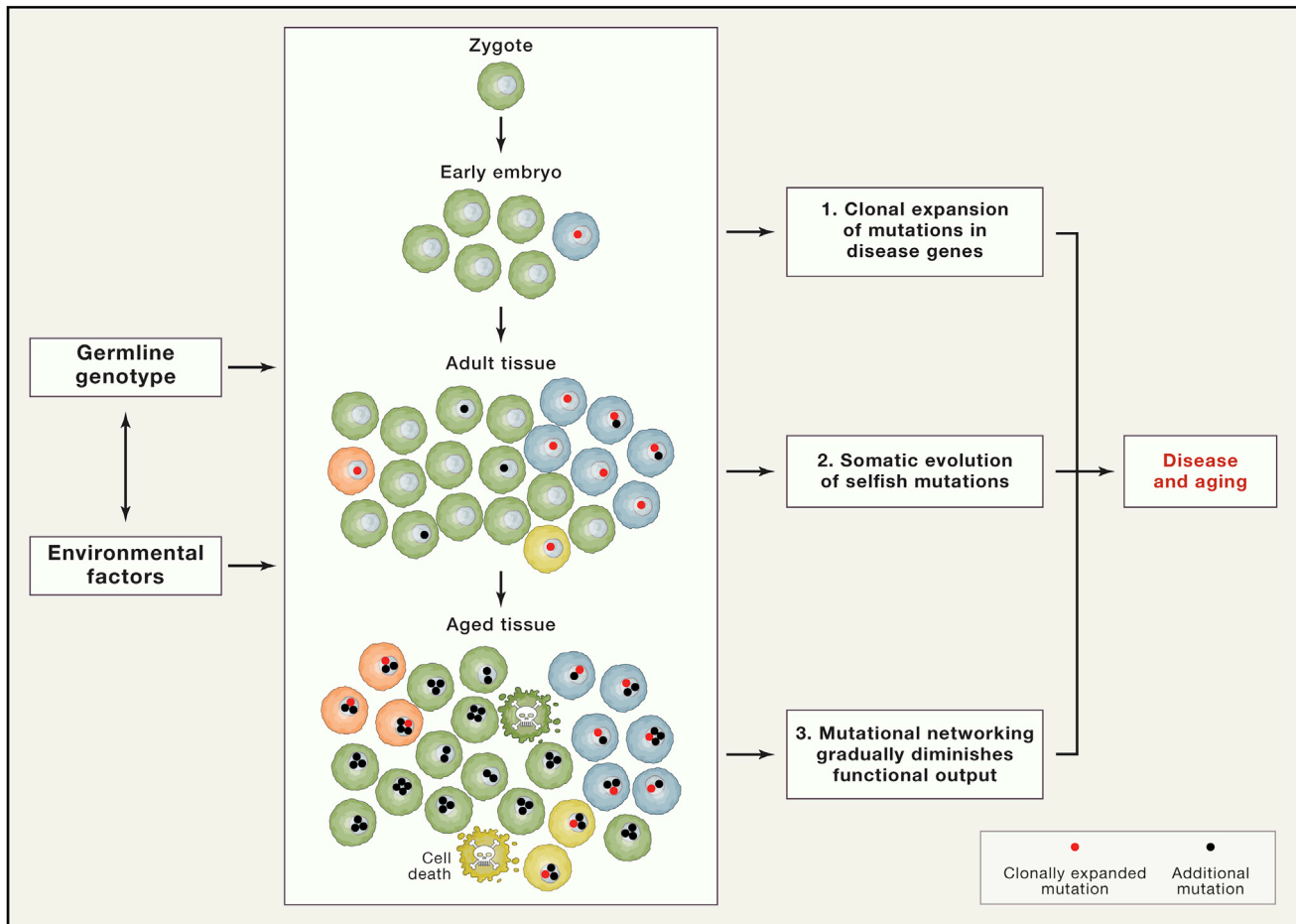


Figure 2. Possible Pathogenic Mechanisms of Somatic Genome Mosaicism

De novo mutations occur at all stages during the lifespan of an organism.

(A) Mutations that occur early in development can expand clonally, resulting in substantial proportions of cells in a tissue carrying the same mutation (blue cells), even in the absence of selection. When such mutations disrupt genes associated with Mendelian diseases, their adverse effects may be similar to that of the germline mutations, depending on the proportion of cells affected by the mosaicism.

(B) Cells in adult tissues that acquire *de novo* mutations may clonally expand in a process called somatic evolution (orange and yellow cells). Such selfish mutations can form substantial clonal lineages, which may cause cancer and other age-related diseases that involve loss of proliferative homeostasis, such as atherosclerosis.

(C) In all cells during aging, mutations accumulate across its genome, including in sequences that participate in functional networks (Figure 3). This may result in transcriptional noise, which, in turn, could adversely affect the signaling networks that are critical for maintaining organismal homeostasis.

obtained from genome-sequenced pedigree data. This suggests that other factors, including higher germ cell DNA repair rates, could also play a role. Evidence of the latter has come from recent work that identified a process called transcriptional scanning that maintains DNA sequence integrity specifically in the testes (Xia et al., 2020).

That random mutations can have an adverse effect on the fitness of cell populations, however, has been shown in an experiment using 3 groups of 75 isogenic strains of *E. coli* (Elena and Lenski, 1997). Each group harbored one, two, or three randomly induced mutations. The results indicated a log-linear decline of average fitness with the increase in the number of mutations. Of course, this result cannot be easily extrapolated to mammalian cells because mammalian genomes are much larger and contain more redundant sequences. So far there is very little

insight into the mechanism through which random somatic mutations could be pathogenic in aging mammals. Here we propose that there are essentially three such mechanisms (Figure 2).

Clonal Expansion of Mutations in Human Disease Genes

It has been known for some time that many Mendelian genetic diseases have a somatic mutational counterpart. The best examples are overgrowth syndromes in children, such as Proteus syndrome, neurofibromatosis 1 and 2, McCune-Albright syndrome, and CLOVES syndrome (Erickson, 2003, 2010). A fraction of these diseases, and sometimes all cases (because the germline variant would be lethal), are caused by somatic mutations that occur early enough in embryonic development to affect a substantial fraction of the tissue in which the disease manifests. Somewhat surprisingly, the fraction of cells in a tissue harboring the disease-causing mutation can be as low as 1% and still show

disease, as described for Proteus syndrome (Lindhurst et al., 2011). In many cases, the somatic mutation confers a growth advantage to mutant cells (hence the term “overgrowth syndromes”), but often the mutation is simply clonally amplified by chance (Figure 2A). The most dramatic example of clonal amplification of a human disease gene is an individual with sporadic early-onset Alzheimer’s disease who showed somatic mosaicism for a presenilin 1 gene mutation. The degree of mosaicism in this individual at the age of presentation was 8% in peripheral lymphocytes and 14% in the cerebral cortex with a gene dosage effect (Beck et al., 2004).

In many cases, Mendelian diseases are caused by a germline and a somatic mutation, each affecting an allelic copy of the same gene, such as in autosomal dominant polycystic kidney disease, in which one allele of the polycystic kidney disease 2 (*ADPKD2*) gene has been inactivated in the germline and the other by a somatic mutation (Koptides et al., 1999). The cysts in this disease are products of clonal expansion of the second, somatic mutation. Nephron cysts also occur in the absence of disease, during normal aging, in which case their prevalence increases with age, perhaps because of an increase in somatic mutations. This is one of many examples where the difference between disease and aging is blurred, with a possible overlap in mechanism.

Postzygotic mosaic mutation as a cause of disease is not limited to monogenic Mendelian disease. Indeed, in the complex disease autism spectrum disorder (ASD), more than 7% of *de novo*, causal mutations have been identified as postzygotic (Lim et al., 2017). These mutations were found in blood with an alternative allele frequency of 40% or less. Interestingly, a large portion of the affected genes were clinically relevant to ASD. The authors hypothesized that the postzygotic mutations arose early in development, which is supported by the observed enrichment of expression of the affected genes in the prenatal brain, most notably the amygdala, a part of the brain involved in emotion.

To gain an impression of the potential importance of somatic mutation expansion in human disease genes, we estimated the prevalence of random, postzygotic mosaic mutations affecting any known human disease genes in early development by chance alone. There are 4,200 human genes in which mutations cause Mendelian disorders or complex diseases, as reported in the OMIM database on January 1, 2020 (Amberger et al., 2019). Based on a reported somatic mutation rate of 2.66×10^{-9} per base pair per mitosis (Milholland et al., 2017), the number of amino acids encoded per gene (Brocchieri and Karlin, 2005), and codon use (Nakamura et al., 2000), the chance of a human embryo to acquire a heterozygous loss-of-function (LOF) mutation at its first zygotic division is 0.058%. This same chance increases when the number of cells increases. For example, at the seventh cell division, the chance that an embryo is affected increases to 1.6%. Such a mutation (i.e., at the seventh cell division), could expand to 1% of human somatic cells only by genetic drift, enough to cause disease, as described above (Lindhurst et al., 2011).

Of course, human development and aging cannot be explained by a simple series of cell divisions, like a cell line in culture, but is subject to complex and hierarchically dictated

schemes, with some cells dividing much more frequently than others and others becoming subject to apoptosis. Nevertheless, accidental somatic mutation early in development could be a significant mechanism in the etiology of human disease, alone or in combination with a germline variant. In such cases, the phenotypic effects are straightforward and associated with the known role of the target gene(s) as a germline defect. However, combinations of low-frequency disease gene mosaics could occur, in which case the phenotypic effects in terms of aging phenotypes in organs and tissues are difficult to predict. Finally, an important question is how a clonally expanded disease gene mutation ending up as a very small fraction of cells can still have pathogenic effects. One possible mechanism could be protein aggregation, a hallmark of normal aging and of some age-related diseases, including Alzheimer’s disease (AD). Age-related deposition of extracellular β -amyloid ($A\beta$) is associated with AD pathogenesis, and the resulting aggregates have been proposed to seed their own progression and spread (Eisele et al., 2015). In this case, the pathogenic effects would be caused by a combination of clonal expansion and prion-like self-propagation of protein misfolding.

Somatic Evolution

Evolution does not only occur in populations of organisms but also in populations of cells (Figure 2B), which are genetically heterogeneous because of *de novo* mutations. For example, variations have been observed in gene copy number and in gene expression and protein levels among 14 HeLa cell strains from 13 international laboratories, cultured under uniform conditions, with consequences for their phenotype, such as responses to *Salmonella* infection (Liu et al., 2019). Similar findings were reported in an earlier study of 27 strains of the breast cancer cell line MCF7 and other cell lines (Ben-David et al., 2018). Importantly, single-cell-derived clones quickly re-developed genetic heterogeneity. This cell line evolution occurred as a consequence of positive clonal selection that was sensitive to culture conditions (Ben-David et al., 2018).

Not surprisingly, since the landmark paper by Nowell (1976), most attention has been focused on evolution of somatic cells in relation to the well-documented, age-related increase in cancer incidence and mortality. For example, specific TP53 gene mutations normally found in human skin tumors have also been detected in normal human skin cells, but only after UV irradiation (Nakazawa et al., 1994). Evidence was provided that this could only be a result of selective clonal expansion of TP53 mutant cells. More recently, by performing targeted ultra-deep sequencing of hundreds of sun-exposed normal human skin biopsies, Martincorena et al. (2015) discovered somatic mutations in multiple cancer genes, including genes encoding key drivers of cutaneous squamous cell carcinomas. These mutations could be detected because of their somatic amplification into thousands of evolving clones due to a selective advantage over other cells without such mutations. At this stage, the skin cell clones harboring the mutations were still normal skin cells. These results were mostly interpreted in terms of cancer, and tumors can be driven by mutations occurring in the still normal cells when exposed to a natural mutagen. However, cancer genes were the only genes analyzed, and a more objective way of describing the phenomenon would be as an example of somatic evolution of

mutations expanded in a tissue because they provide the normal cell with a growth and/or survival advantage.

Another example of mutations enriched by somatic evolution in tissues exposed to a mutagen is the observation of cancer driver mutations in normal lung tissue of smokers. Franklin et al. (1997) observed a single identical point mutation in TP53 in microdissected bronchial epithelium from multiple sites in both lungs of a subject showing dysplastic changes in the respiratory epithelium but no cancer. The mutation was not found in blood or other organs and must have come from an expanded bronchial epithelial clone in the bronchial mucosa. Indeed, using bronchial epithelial cell clones for detecting somatic mutations across the genome, an approach described above, Yoshida et al. (2020) demonstrated that smoking increases the overall somatic mutation frequency, with 25% of the cell clones sequenced carrying cancer driver mutations. They found a number of known cancer driver mutations, most notably NOTCH1, TP53, and ARID2, under positive selection in normal bronchial epithelium. Naturally, these somatically expanded mutations in cancer driver genes were assumed to eventually lead to cancer.

However, there is evidence that somatic evolution also causally contributes to age-related diseases other than cancer. It has been proposed that acquired somatic mutations in cancer driver genes that have been found frequently in human bronchial epithelium of smokers, such as TP53, Ras, and PTEN, enhance the inflammation and oxidative stress associated with chronic obstructive pulmonary disease (COPD), which is caused mostly by cigarette smoking but without known etiology (Anderson and Bozinovski, 2003).

Somatic evolution has also been considered a potential mechanism for cardiovascular disease, which, like cancer, is a major age-related disorder (Andreassi, 2003). The pathogenesis of atherosclerosis has been proposed to involve an inflammatory response to injury and, thus, cell proliferation. The monoclonal nature of the smooth muscle cells in human atherosclerotic plaques (Benditt and Benditt, 1973) could reflect expansion of mutant cells. Evidence of genomic instability in atherosclerotic cells has indeed been reported (Arvanitis et al., 2005; McCaffrey et al., 1997), leading to the hypothesis that expansion of mutant cells could be a major causal factor in cardiovascular disease. However, the nature of the mutant genes(s) remains unknown (Andreassi, 2003). Interestingly, advanced atherosclerotic lesions contain senescent cells, which can further drive pathology through an increase in expression of atherogenic and inflammatory cytokines and chemokines (Childs et al., 2016). Cellular senescence is permanent cessation of cellular proliferation, which is often induced by DNA damage and genome instability (Andriani et al., 2019; Campisi, 2013).

More recent data show that the inflammatory endpoint in cardiovascular disease can also be reached through somatic evolution in blood cells. Indeed, partial bone marrow reconstitution in mice with Tet2-deficient cells led to clonal expansion of these cells and a marked increase in atherosclerotic plaque size (Fuster et al., 2017). Although the mechanism of its selective advantage is not entirely clear, all Tet2-deficient cells appear to have enhanced proliferation, which explains their enrichment during aging in hematopoietic cells (Jaiswal et al., 2014).

In addition to a cell growth advantage, mutations can be positively selected by preventing cell death. Using ultra-deep whole-exome sequencing of post-mortem hippocampus from AD patients, clonally expanded somatic mutations were found to increase with age and were significantly enriched for phosphatidylinositol 3-kinase (PI3K)-protein kinase B (AKT), mitogen-associated protein kinase (MAPK), and AMP-activated protein kinase (AMPK) pathway genes, which are involved in tau phosphorylation (Park et al., 2019). Hyperphosphorylation of tau, which is the main component of neurofibrillary tangles and a hallmark of AD, prevents apoptosis of neuronal cells in rodents (Li et al., 2007). This mechanism could promote the survival of these mutant cells. As we have seen, even if only a few cells in the hippocampus are affected by these mutations, the protein aggregates could spread from each cell to its surrounding neighbors, as discussed above for A β .

In summary, in tissues of mammals, the adaptive landscape of somatic evolution during aging is similar to the adaptive landscape of evolution (Wright, 1931) but from a different perspective. Indeed, in the aging soma, selection for fitness among individual cells tends to move them away from their optimal peak of functioning, in concert with other cells in their host, to a more selfish pattern of genetic variation. This pushes the aging process toward loss of functionality and increased risk of disease, most notably loss of proliferative homeostasis; e.g., neoplasia, fibrosis, and inflammation, long recognized as a major aging-related phenomenon (Martin, 2007).

Mutational Networking

As we have discussed, there is now ample evidence that somatic cells carrying mutations in known human heritable disease genes can clonally expand during development and can contribute to diseases other than cancer. In addition, certain acquired gene mutations that are not by themselves disease causing can confer a selective advantage to the cell, which expands and gradually erodes organ and tissue functioning because of increasingly selfish behavior. Although the magnitude of the adverse effects of these events in aging still await more extensive studies, there is a third possible mechanism by which randomly accumulating mutations eventually affect cell fitness. This does not require clonal outgrowth and depends on the penetration of such mutations in the DNA sequence components of the gene-regulatory networks (GRNs) that provide function to a mammalian organism throughout its life (Figures 2C and 3A). Virtually all mutations would accumulate not in the about 1% protein-coding part of the genome but in the gene-regulatory regions that make up approximately 11% of all genome sequences (Kellis et al., 2014).

The topology of GRNs is scale free, which makes them mutationally robust and minimizes the possibility of a deleterious mutation severely affecting function (Albert et al., 2000). However, GRNs are not immune to mutational effects. Indeed, their structure, which dampens the effects of mutations, also serves specific biological functions that could be affected by mutations. For example, the most abundant network motif in GRNs, from fruit flies to humans, is the “feedforward loop” (Boyle et al., 2014). This network motif serves to accelerate or delay the response of a cell to external stimuli (Mangan and Alon, 2003). Even a slightly deleterious mutation could change the kinetics of this process, affecting the time-accurate manner of the

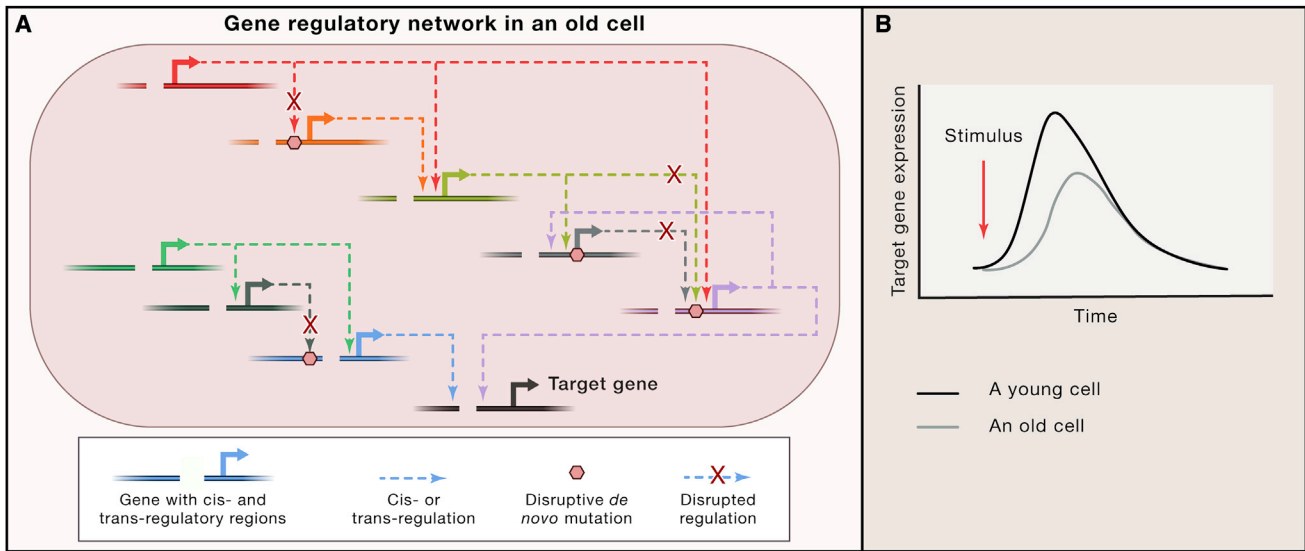


Figure 3. How Can Mutation Accumulation Affect Functional Output of GRNs?

Cellular and organismal function is organized in complex gene-regulatory networks (GRNs) involving interactions between large numbers of genes and their regulators, typically genes and their *cis*- and *trans*-acting elements.

(A) Mutations inevitably accumulate in elements of each functional network in a stochastic manner. Although initially insufficient in number to have an effect, eventually these mutations collectively diminish the functional output of the GRN in an increasingly large number of cells, possibly through increased transcriptional noise (see text).

(B) Increased noise in GRNs could explain defects in kinetics of cell signaling during aging, in which the response to external stimuli is blunted.

response (Figure 3B). Because GRNs consist of many genes and many more gene-regulatory sequences, the chance that such deleterious mutations might affect network functions is not insignificant. Different sets of random mutations could adversely affect the same GRN through various synergistic interactions. Indeed, a network of 193 noncoding loci has been identified in which different sets of mutations disrupt target gene expression in 88% of tumors in 930 patients across 22 cancer types (Zhang et al., 2018).

Accumulated mutations in GRNs could explain the defects in cell signaling that have been observed with age (Greer and Brunet, 2008). Cells respond to environmental challenges such as temperature changes, infections, and a variety of other stressors through GRNs and their networks of regulatory interactions. Although the dynamics of these complex networks in humans are far from understood, their actions are ultimately based on genes and the regulatory sequences that control their expression. Each of the genes in a GRN that is active in a given cell has to work in concert to ensure the specific function, fitness, and survival of that cell. All of these genes must be expressed at the proper time and in the proper amounts to ensure the appropriate functional output. Of note, normal aging is not complete abrogation of particular life processes but gradual erosion of the efficiency of such processes. Examples are thermoregulation, vaccine responses, and short-term memory, which are all reduced in the elderly but far from completely ablated. Such a decline in function could well be explained by gradual accumulation of weakly deleterious mutations in GRN components (Figure 3). Naturally, the key question is how mutation accumulation in GRNs would translate in phenotypic deficiencies. One possibility is through transcriptional noise.

Transcriptional noise—increased cell-to-cell transcriptional heterogeneity in a tissue, organ, or cell population—is a predicted consequence of mutation accumulation in GRNs. Elevated transcriptional noise was first discovered in house-keeping and cell-type-specific genes in cardiomyocytes of old mice (Bahar et al., 2006). Later, increased transcriptional noise was demonstrated in activated T lymphocytes of aged mice (Martinez-Jimenez et al., 2017), and more recently it was shown that most cell types in the mouse lung show increased transcriptional noise at old age (Angelidis et al., 2019). Transcriptional noise has also been demonstrated in the human pancreas among islet endocrine cells from older donors (Enge et al., 2017). In this study, transcriptional noise and somatic mutation load were shown to be increased in the aged human pancreas. Interestingly, cells with high levels of transcriptional noise displayed an increase in non-cell-type-specific hormone expression. The authors suggested that such a drift in cell fate could explain the decrease in fitness and organ function associated with aging.

One of the most widely studied molecular changes during aging is telomere attrition, which can cause cellular senescence (Bodnar et al., 1998; Campisi, 2013). Accumulation of senescent cells with age can have adverse effects (Campisi, 2013). Interestingly, telomere attrition can also contribute to transcriptional noise. Using isogenic pairs of human primary cells in culture with long or short telomeres, expression profiles of genes close to the telomeres were found to be altered with telomere length (Robin et al., 2014). Although it is still unclear whether this finding is relevant *in vivo*, this mechanism could increase transcriptional noise without the need for telomeres to become critically shortened and cause senescence or cell death.

Apart from non-cell-type-specific gene expression and loss of cell-to-cell coordination when responding to external stimuli, transcriptional noise can also lead to stoichiometric imbalances in protein macromolecular assembly, which is essential for most biological processes. The subsequent unpartnered components may misfold and accumulate, leading to protein aggregation, which is associated with multiple age-related diseases, especially neuronal degeneration (Lindner and Demarez, 2009). Crude examples of this are the dose imbalances caused by CNVs and by LOF mutations (Veitia et al., 2017). For example, *PIN1* LOF somatic mutations in the brains of AD patients cause haploinsufficiency, which increases tau phosphorylation and aggregation (Park et al., 2019).

Although amply confirmed and likely to explain at least some of the adverse effects of aging, it remains unclear whether transcriptional noise is caused by somatic mutations. Indeed, although the somatic mutation rate is at least an order of magnitude greater than originally assumed based on estimated germline mutation frequencies (Milholland et al., 2017), even a few thousand SNVs genome-wide may not affect the expression of a large enough number of genes to cause transcriptional noise. Of course, SNVs are the only mutation type for which we know accurate numbers per cell, with other mutations not yet quantitatively analyzed to the same extent; e.g., SVs could also accumulate with age and have a much greater functional effect on gene expression patterns.

Alternatively, epimutations, i.e., alterations in the epigenome that control gene expression patterns through chromatin alteration, such as DNA methylation or histone modification, but do not involve changes in the DNA coding sequence, have been proposed to play a more important causal role in aging and disease, possibly through transcriptional noise (Enge et al., 2017). Using single-cell bisulfite sequencing, epimutation frequencies of CpG DNA methylation have been observed in mouse liver hepatocytes to be at least two orders of magnitude higher than somatic mutation frequencies (Gravina et al., 2016). Although this suggests that increased epimutations are more likely to play a causal role in the observed increase in transcriptional noise, the effect of a single DNA sequence alteration, on average, could be much greater than that of a single epimutation. For example, CpG DNA methylation sites may act collectively in controlling gene expression. In other words, many epimutations may be needed to cause the same effect as one sequence alteration.

Nevertheless, epimutations have been identified as *bona fide* disease-causing events, sometimes directly in somatic cells and sometimes indirectly through genetic mutations in epigenetic modifiers that affect chromatin (Zoghbi and Beaudet, 2016). However, compared with mutations affecting DNA sequence, well-documented cases of epimutations as a cause of human disease are scarce, possibly because the effect of the average epimutation is not as high as that of a mutation. Naturally, compared with DNA sequence alterations, the field of epimutations as stochastic causes of disease and aging is still in its infancy, with little information regarding epimutational mosaicism. Ultimately, to compare the relative effects of mutations and epimutations on influencing the transcriptome would require simultaneous sequencing of DNA and RNA from the same cell as well as computational methods to evaluate the

(epi)mutational effect on target gene transcription and regulatory network kinetics, which are now being developed (Dey et al., 2015; Zhou et al., 2018).

In summary, the genome's distributed functional organization and high level of integration of the many sequence features that encode specific cellular functions provide robustness and a very high level of redundancy. Initially, therefore, genomes can tolerate fairly high levels of mutation. However, at some point, this type of genome functional organization will begin to amplify the effects of multiple random mutations. It is possible, therefore, that even with a linear increase in somatic mutations with age, their effects on health and mortality would be exponential and begin to increase rapidly in middle age, following the same Gompertz kinetics as the increase in mortality and disease incidence during the lifetime.

Conclusions

The considerable advances in DNA sequencing technologies over the last decade have provided us with the first accurate picture of the aging-related landscape of somatic mutagenesis. Increasing evidence has revealed that a continuous trickle of *de novo* mutations in the soma, from early embryogenesis to adulthood and old age, can cause disease—not only cancer but also a large variety of non-cancer-related disorders. The most surprising finding so far has been the ubiquity and high frequency of somatic mutations. Whether this is sufficiently high to explain a considerable part of age-related functional decline and disease, alone or in conjunction with epimutations, remains to be tested. Irrespective of a possible causal role of somatic mosaicism in aging, somatic mutations are not the only cause of aging.

Trying to explain aging in terms of a singular process would be in conflict with evolutionary theory. Even if loss of genome sequence integrity was the most conserved cause of aging, already active in the first replicators (Vijg, 2007), natural selection would allow a multitude of mutations with late adverse effects to accumulate in the germline, many of which would be positively selected for because of their beneficial effects early in life (Williams, 1957). In this respect, somatic mutation accumulation could be a conserved, inevitable cause of aging but superposed on multiple other processes that usually cause the earlier demise of an individual.

Nevertheless, it is clear that previous ideas of a stable genome have now been replaced by a much more dynamic view of considerable cell-to-cell variation in genome structure and function. The purpose of this Perspective was to focus attention on the recent flood of quantitative data on somatic mosaicism, driven by technological advances, that allows, for the first time, to test whether progressive volatility of genome sequence organization could be causally related to aging. The main challenge ahead is to fully characterize mosaic genomes in human tissues in relation to aging and disease. Continuous progress in single-cell sequencing technologies should provide us with such a comprehensive picture in the years ahead, allowing us to make much better estimates of the magnitude of genome mosaicism and its causal relation to human disease and aging. To understand the functional consequences of genome mosaicism, we need to learn how *de novo* mutations can affect phenotype. This will be greatly facilitated by the current explosion of

single-cell multi-omics technologies (Chappell et al., 2018), which will allow genome alterations to be directly connected to changes at the epigenome, transcriptome, and proteome level. Such an integrated analysis of the effect of genome sequence erosion on cell, tissue, and organismal phenotypes will eventually allow us to draw specific conclusions about the role of somatic mutations as a causal factor in aging and age-related diseases.

ACKNOWLEDGMENTS

This work was supported by the NIH (grants P01 AG017242, P01 AG047200, P30 AG038072, and U01 ES029519 to J.V. and K99 AG056656 to X.D.), the Glenn Foundation for Medical Research (to J.V.), and the Shanghai Jiao Tong University School of Medicine.

DECLARATION OF INTERESTS

J.V. and X.D. are co-founders and shareholders of SingulOmics Corp.

REFERENCES

Albert, R., Jeong, H., and Barabasi, A.L. (2000). Error and attack tolerance of complex networks. *Nature* 406, 378–382.

Amberger, J.S., Bocchini, C.A., Scott, A.F., and Hamosh, A. (2019). OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.* 47 (D1), D1038–D1043.

Anderson, G.P., and Bozinovski, S. (2003). Acquired somatic mutations in the molecular pathogenesis of COPD. *Trends Pharmacol. Sci.* 24, 71–76.

Andreassi, M.G. (2003). Coronary atherosclerosis and somatic mutations: an overview of the contributive factors for oxidative DNA damage. *Mutat. Res.* 543, 67–86.

Andriani, G.A., Maggi, E., Piqué, D., Zimmerman, S.E., Lee, M., Quispe-Tintaya, W., Maslov, A., Campisi, J., Vijg, J., Mar, J.C., and Montagna, C. (2019). A direct comparison of interphase FISH versus low-coverage single cell sequencing to detect aneuploidy reveals respective strengths and weaknesses. *Sci. Rep.* 9, 10508.

Angelidis, I., Simon, L.M., Fernandez, I.E., Strunz, M., Mayr, C.H., Greiffo, F.R., Tsitsiridis, G., Ansari, M., Graf, E., Strom, T.M., et al. (2019). An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. *Nat. Commun.* 10, 963.

Arvanitis, D.A., Flouris, G.A., and Spandidos, D.A. (2005). Genomic rearrangements on VCAM1, SELE, APEG1 and AIF1 loci in atherosclerosis. *J. Cell. Mol. Med.* 9, 153–159.

Audano, P.A., Sulovari, A., Graves-Lindsay, T.A., Cantsilieris, S., Sorensen, M., Welch, A.E., Dougherty, M.L., Nelson, B.J., Shah, A., Dutcher, S.K., et al. (2019). Characterizing the Major Structural Variant Alleles of the Human Genome. *Cell* 176, 663–675.e19.

Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R.; 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature* 526, 68–74.

Bae, T., Tomasini, L., Mariani, J., Zhou, B., Roychowdhury, T., Franjic, D., Pleitikos, M., Pattni, R., Chen, B.J., Venturini, E., et al. (2018). Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science* 359, 550–555.

Baer, C.F., Miyamoto, M.M., and Denver, D.R. (2007). Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat. Rev. Genet.* 8, 619–631.

Bahar, R., Hartmann, C.H., Rodriguez, K.A., Denny, A.D., Busuttill, R.A., Dollé, M.E., Calder, R.B., Chisholm, G.B., Pollock, B.H., Klein, C.A., and Vijg, J. (2006). Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* 441, 1011–1014.

Beck, J.A., Poulter, M., Campbell, T.A., Uphill, J.B., Adamson, G., Geddes, J.F., Revesz, T., Davis, M.B., Wood, N.W., Collinge, J., and Tabrizi, S.J. (2004). Somatic and germline mosaicism in sporadic early-onset Alzheimer's disease. *Hum. Mol. Genet.* 13, 1219–1224.

Ben-David, U., Siranosian, B., Ha, G., Tang, H., Oren, Y., Hinohara, K., Strathdee, C.A., Dempster, J., Lyons, N.J., Burns, R., et al. (2018). Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* 560, 325–330.

Benditt, E.P., and Benditt, J.M. (1973). Evidence for a monoclonal origin of human atherosclerotic plaques. *Proc. Natl. Acad. Sci. USA* 70, 1753–1756.

Blokzijl, F., de Ligt, J., Jager, M., Sasselli, V., Roerink, S., Sasaki, N., Huch, M., Boymans, S., Kujik, E., Prins, P., et al. (2016). Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* 538, 260–264.

Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C.P., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S., and Wright, W.E. (1998). Extension of life-span by introduction of telomerase into normal human cells. *Science* 279, 349–352.

Bohrson, C.L., Barton, A.R., Lodato, M.A., Rodin, R.E., Luquette, L.J., Viswanadham, V.V., Gulhan, D.C., Cortés-Ciriano, I., Sherman, M.A., Kwon, M., et al. (2019). Linked-read analysis identifies mutations in single-cell DNA-sequencing data. *Nat. Genet.* 51, 749–754.

Boveri, T. (1914). Zur Frage der Entstehung maligner Tumoren (Gustav Fischer).

Boyle, A.P., Araya, C.L., Brdlik, C., Cayting, P., Cheng, C., Cheng, Y., Gardner, K., Hillier, L.W., Janette, J., Jiang, L., et al. (2014). Comparative analysis of regulatory information and circuits across distant species. *Nature* 512, 453–456.

Brazhnik, K., Sun, S., Alani, O., Kinkhabwala, M., Wolkoff, A.W., Maslov, A.Y., Dong, X., and Vijg, J. (2020). Single-cell analysis reveals different age-related somatic mutation profiles between stem and differentiated cells in human liver. *Sci Adv.* 6, eaax2659.

Brocchieri, L., and Karlin, S. (2005). Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.* 33, 3390–3400.

Cai, X., Evrony, G.D., Lehmann, H.S., Elhosary, P.C., Mehta, B.K., Poduri, A., and Walsh, C.A. (2014). Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep.* 8, 1280–1289.

Campisi, J. (2013). Aging, cellular senescence, and cancer. *Annu. Rev. Physiol.* 75, 685–705.

Chappell, L., Russell, A.J.C., and Voet, T. (2018). Single-Cell (Multi)omics Technologies. *Annu. Rev. Genomics Hum. Genet.* 19, 15–41.

Chen, C., Xing, D., Tan, L., Li, H., Zhou, G., Huang, L., and Xie, X.S. (2017). Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science* 356, 189–194.

Childs, B.G., Baker, D.J., Wijshake, T., Conover, C.A., Campisi, J., and van Deursen, J.M. (2016). Senescent intimal foam cells are deleterious at all stages of atherosclerosis. *Science* 354, 472–477.

Crowley, C., and Curtis, H.J. (1963). The development of somatic mutations in mice with age. *Proc. Natl. Acad. Sci. USA* 49, 626–628.

Day, F.R., Ruth, K.S., Thompson, D.J., Lunetta, K.L., Pervjakova, N., Chasman, D.I., Stolk, L., Finucane, H.K., Sulem, P., Bulik-Sullivan, B., et al.; PRAC-TICAL consortium; kConFab Investigators; AOCs Investigators; Generation Scotland; EPIC-InterAct Consortium; LifeLines Cohort Study (2015). Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat. Genet.* 47, 1294–1303.

Dey, S.S., Kester, L., Spanjaard, B., Bienko, M., and van Oudenaarden, A. (2015). Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotechnol.* 33, 285–289.

Di Noia, J.M., and Neuberger, M.S. (2007). Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* 76, 1–22.

Dollé, M.E., Giese, H., Hopkins, C.L., Martus, H.J., Hausdorff, J.M., and Vijg, J. (1997). Rapid accumulation of genome rearrangements in liver but not in brain of old mice. *Nat. Genet.* 17, 431–434.

- Dong, X., Zhang, L., Milholland, B., Lee, M., Maslov, A.Y., Wang, T., and Vijg, J. (2017). Accurate identification of single-nucleotide variants in whole-genome-amplified single cells. *Nat. Methods* *14*, 491–493.
- Drake, J.W., Charlesworth, B., Charlesworth, D., and Crow, J.F. (1998). Rates of spontaneous mutation. *Genetics* *148*, 1667–1686.
- Dumanski, J.P., Lambert, J.C., Rasi, C., Giedraitis, V., Davies, H., Grenier-Bolley, B., Lindgren, C.M., Campion, D., Dufouil, C., Pasquier, F., et al.; European Alzheimer's Disease Initiative Investigators (2016). Mosaic Loss of Chromosome Y in Blood Is Associated with Alzheimer Disease. *Am. J. Hum. Genet.* *98*, 1208–1219.
- Duncan, A.W., Hanlon Newell, A.E., Bi, W., Finegold, M.J., Olson, S.B., Beaudet, A.L., and Grompe, M. (2012). Aneuploidy as a mechanism for stress-induced liver adaptation. *J. Clin. Invest.* *122*, 3307–3315.
- Eisele, Y.S., Monteiro, C., Fearn, C., Encalada, S.E., Wiseman, R.L., Powers, E.T., and Kelly, J.W. (2015). Targeting protein aggregation for the treatment of degenerative diseases. *Nat. Rev. Drug Discov.* *14*, 759–780.
- Elena, S.F., and Lenski, R.E. (1997). Test of synergistic interactions among deleterious mutations in bacteria. *Nature* *390*, 395–398.
- Enge, M., Arda, H.E., Mignardi, M., Beausang, J., Bottino, R., Kim, S.K., and Quake, S.R. (2017). Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. *Cell* *171*, 321–330.e14.
- Erickson, R.P. (2003). Somatic gene mutation and human disease other than cancer. *Mutat. Res.* *543*, 125–136.
- Erickson, R.P. (2010). Somatic gene mutation and human disease other than cancer: an update. *Mutat. Res.* *705*, 96–106.
- Evrony, G.D., Lee, E., Park, P.J., and Walsh, C.A. (2016). Resolving rates of mutation in the brain using single-neuron genomics. *eLife* *5*, e12966.
- Failla, G. (1958). The aging process and cancerogenesis. *Ann. N Y Acad. Sci.* *71*, 1124–1140.
- Franco, I., Johansson, A., Olsson, K., Vrtačnik, P., Lundin, P., Helgadottir, H.T., Larsson, M., Revêchon, G., Bosia, C., Pagnani, A., et al. (2018). Somatic mutagenesis in satellite cells associates with human skeletal muscle aging. *Nat. Commun.* *9*, 800.
- Franco, I., Helgadottir, H.T., Moggio, A., Larsson, M., Vrtačnik, P., Johansson, A., Norgren, N., Lundin, P., Mas-Ponte, D., Nordström, J., et al. (2019). Whole genome DNA sequencing provides an atlas of somatic mutagenesis in healthy human cells and identifies a tumor-prone cell type. *Genome Biol.* *20*, 285.
- Franklin, W.A., Gazdar, A.F., Haney, J., Wistuba, I.I., La Rosa, F.G., Kennedy, T., Ritchey, D.M., and Miller, Y.E. (1997). Widely dispersed p53 mutation in respiratory epithelium. A novel mechanism for field carcinogenesis. *J. Clin. Invest.* *100*, 2133–2137.
- Fuster, J.J., MacLauchlan, S., Zuriaga, M.A., Polackal, M.N., Ostriker, A.C., Chakraborty, R., Wu, C.L., Sano, S., Muralidharan, S., Rius, C., et al. (2017). Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science* *355*, 842–847.
- Gage, F.H. (2019). Adult neurogenesis in mammals. *Science* *364*, 827–828.
- Gravina, S., Dong, X., Yu, B., and Vijg, J. (2016). Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. *Genome Biol.* *17*, 150.
- Greer, E.L., and Brunet, A. (2008). Signaling networks in aging. *J. Cell Sci.* *121*, 407–412.
- Grist, S.A., McCarron, M., Kutlaca, A., Turner, D.R., and Morley, A.A. (1992). In vivo human somatic mutation: frequency and spectrum with age. *Mutat. Res.* *266*, 189–196.
- Jacobs, P.A., Brunton, M., Court Brown, W.M., Doll, R., and Goldstein, H. (1963). Change of human chromosome count distribution with age: evidence for a sex differences. *Nature* *197*, 1080–1081.
- Jacobs, P.A., Brunton, M., and Brown, W.M. (1964). Cytogenetic studies in leucocytes on the general population: subjects of ages 65 years and more. *Ann. Hum. Genet.* *27*, 353–365.
- Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burt, N., Chavez, A., et al. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* *371*, 2488–2498.
- Johannsen, W. (1911). The genotype conception of heredity. *Am. Nat.* *45*, 129–159.
- Kellis, M., Wold, B., Snyder, M.P., Bernstein, B.E., Kundaje, A., Marinov, G.K., Ward, L.D., Birney, E., Crawford, G.E., Dekker, J., et al. (2014). Defining functional DNA elements in the human genome. *Proc. Natl. Acad. Sci. USA* *111*, 6131–6138.
- King, M.C., and Wilson, A.C. (1975). Evolution at two levels in humans and chimpanzees. *Science* *188*, 107–116.
- Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., et al. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature* *488*, 471–475.
- Koptides, M., Hadjimichael, C., Koupepidou, P., Pierides, A., and Constantinou Deltas, C. (1999). Germinal and somatic mutations in the PKD2 gene of renal cysts in autosomal dominant polycystic kidney disease. *Hum. Mol. Genet.* *8*, 509–513.
- Kosugi, S., Momozawa, Y., Liu, X., Terao, C., Kubo, M., and Kamatani, Y. (2019). Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biol.* *20*, 117.
- Laurie, C.C., Laurie, C.A., Rice, K., Doherty, K.F., Zelnick, L.R., McHugh, C.P., Ling, H., Hetrick, K.N., Pugh, E.W., Amos, C., et al. (2012). Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat. Genet.* *44*, 642–650.
- Li, H.L., Wang, H.H., Liu, S.J., Deng, Y.Q., Zhang, Y.J., Tian, Q., Wang, X.C., Chen, X.Q., Yang, Y., Zhang, J.Y., et al. (2007). Phosphorylation of tau antagonizes apoptosis by stabilizing beta-catenin, a mechanism involved in Alzheimer's neurodegeneration. *Proc. Natl. Acad. Sci. USA* *104*, 3591–3596.
- Lim, E.T., Uddin, M., De Rubeis, S., Chan, Y., Kamumbu, A.S., Zhang, X., D'Gama, A.M., Kim, S.N., Hill, R.S., Goldberg, A.P., et al.; Autism Sequencing Consortium (2017). Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. *Nat. Neurosci.* *20*, 1217–1224.
- Lindhurst, M.J., Sapp, J.C., Teer, J.K., Johnston, J.J., Finn, E.M., Peters, K., Turner, J., Cannons, J.L., Bick, D., Blakemore, L., et al. (2011). A mosaic activating mutation in AKT1 associated with the Proteus syndrome. *N. Engl. J. Med.* *365*, 611–619.
- Lindner, A.B., and Demarez, A. (2009). Protein aggregation as a paradigm of aging. *Biochim. Biophys. Acta* *1790*, 980–996.
- Liu, Y., Mi, Y., Mueller, T., Kreibich, S., Williams, E.G., Van Drogen, A., Borel, C., Frank, M., Germain, P.L., Bludau, I., et al. (2019). Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nat. Biotechnol.* *37*, 314–322.
- Lodato, M.A., Rodin, R.E., Bohrsen, C.L., Coulter, M.E., Barton, A.R., Kwon, M., Sherman, M.A., Vitzthum, C.M., Luquette, L.J., Yandava, C.N., et al. (2018). Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* *359*, 555–559.
- Lofffield, E., Zhou, W., Graubard, B.I., Yeager, M., Chanock, S.J., Freedman, N.D., and Machiela, M.J. (2018). Predictors of mosaic chromosome Y loss and associations with mortality in the UK Biobank. *Sci. Rep.* *8*, 12316.
- Lynch, M., Ackerman, M.S., Gout, J.F., Long, H., Sung, W., Thomas, W.K., and Foster, P.L. (2016). Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* *17*, 704–714.
- Mangan, S., and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA* *100*, 11980–11985.
- Martin, G.M. (2007). The genetics and epigenetics of altered proliferative homeostasis in ageing and cancer. *Mech. Ageing Dev.* *128*, 9–12.
- Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D.C., Fullam, A., Alexandrov, L.B., Tubio, J.M., et al. (2015). Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* *348*, 880–886.

- Martincorena, I., Fowler, J.C., Wabik, A., Lawson, A.R.J., Abascal, F., Hall, M.W.J., Cagan, A., Murai, K., Mahbubani, K., Stratton, M.R., et al. (2018). Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911–917.
- Martinez-Jimenez, C.P., Eling, N., Chen, H.C., Vallejos, C.A., Kolodziejczyk, A.A., Connor, F., Stojic, L., Rayner, T.F., Stubbington, M.J.T., Teichmann, S.A., et al. (2017). Aging increases cell-to-cell transcriptional variability upon immune stimulation. *Science* **355**, 1433–1436.
- McCaffrey, T.A., Du, B., Consigli, S., Szabo, P., Bray, P.J., Hartner, L., Weksler, B.B., Sanborn, T.A., Bergman, G., and Bush, H.L., Jr. (1997). Genomic instability in the type II TGF- β 1 receptor gene in atherosclerotic and restenotic vascular cells. *J. Clin. Invest.* **100**, 2182–2188.
- McConnell, M.J., Lindberg, M.R., Brennand, K.J., Piper, J.C., Voet, T., Cowing-Zitron, C., Shumilina, S., Lasken, R.S., Vermeesch, J.R., Hall, I.M., and Gage, F.H. (2013). Mosaic copy number variation in human neurons. *Science* **342**, 632–637.
- Millholland, B., Dong, X., Zhang, L., Hao, X., Suh, Y., and Vijg, J. (2017). Differences between germline and somatic mutation rates in humans and mice. *Nat. Commun.* **8**, 15183.
- Morgan, T.H., and Bridges, C.B. (1919). The Origin of Gynandromorphs (Carnegie Inst.).
- Nakamura, Y., Gojobori, T., and Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**, 292.
- Nakazawa, H., English, D., Randell, P.L., Nakazawa, K., Martel, N., Armstrong, B.K., and Yamasaki, H. (1994). UV and skin cancer: specific p53 gene mutation in normal skin as a biologically relevant exposure measurement. *Proc. Natl. Acad. Sci. USA* **91**, 360–364.
- Nowell, P.C. (1976). The clonal evolution of tumor cell populations. *Science* **194**, 23–28.
- Odegard, V.H., and Schatz, D.G. (2006). Targeting of somatic hypermutation. *Nat. Rev. Immunol.* **6**, 573–583.
- Ono, T., Ikehata, H., Nakamura, S., Saito, Y., Hosoi, Y., Takai, Y., Yamada, S., Onodera, J., and Yamamoto, K. (2000). Age-associated increase of spontaneous mutant frequency and molecular nature of mutation in newborn and old lacZ-transgenic mouse. *Mutat. Res.* **447**, 165–177.
- Paquola, A.C.M., Erwin, J.A., and Gage, F.H. (2017). Insights into the role of somatic mosaicism in the brain. *Curr. Opin. Syst. Biol.* **1**, 90–94.
- Park, J.S., Lee, J., Jung, E.S., Kim, M.H., Kim, I.B., Son, H., Kim, S., Kim, S., Park, Y.M., Mook-Jung, I., et al. (2019). Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of tau phosphorylation. *Nat. Commun.* **10**, 3090.
- Pierre, R.V., and Hoagland, H.C. (1972). Age-associated aneuploidy: loss of Y chromosome from human bone marrow cells with aging. *Cancer* **30**, 889–894.
- Robin, J.D., Ludlow, A.T., Batten, K., Magdinier, F., Stadler, G., Wagner, K.R., Shay, J.W., and Wright, W.E. (2014). Telomere position effect: regulation of gene expression with progressive telomere shortening over long distances. *Genes Dev.* **28**, 2464–2476.
- Szilard, L. (1959). On the nature of the aging process. *Proc. Natl. Acad. Sci. USA* **45**, 30–45.
- Thompson, D.J., Genovese, G., Halvardson, J., Ulirsch, J.C., Wright, D.J., Terao, C., Davidsson, O.B., Day, F.R., Sulem, P., Jiang, Y., et al.; International Lung Cancer Consortium (INTEGRAL-ILCCO); Breast Cancer Association Consortium; Consortium of Investigators of Modifiers of BRCA1/2; Endometrial Cancer Association Consortium; Ovarian Cancer Association Consortium; Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) Consortium; Kidney Cancer GWAS Meta-Analysis Project; eQTLGen Consortium; Biobank-based Integrative Omics Study (BIOS) Consortium; 23andMe Research Team (2019). Genetic predisposition to mosaic Y chromosome loss in blood. *Nature* **575**, 652–657.
- Tian, X., Firsanov, D., Zhang, Z., Cheng, Y., Luo, L., Tomblin, G., Tan, R., Simon, M., Henderson, S., Steffan, J., et al. (2019). SIRT6 Is Responsible for More Efficient DNA Double-Strand Break Repair in Long-Lived Species. *Cell* **177**, 622–638.e22.
- Veitia, R.A., Govindaraju, D.R., Bottani, S., and Birchler, J.A. (2017). Aging: Somatic Mutations, Epigenetic Drift and Gene Dosage Imbalance. *Trends Cell Biol.* **27**, 299–310.
- Vijg, J. (2007). *Aging of the Genome*, Oxford.
- Wang, J., Fan, H.C., Behr, B., and Quake, S.R. (2012). Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* **150**, 402–412.
- Williams, G.C. (1957). Pleiotropy, Natural Selection, and the Evolution of Senescence. *Evolution* **11**, 398–411.
- Wright, S. (1931). Evolution in Mendelian Populations. *Genetics* **16**, 97–159.
- Xia, B., Yan, Y., Baron, M., Wagner, F., Barkley, D., Chiodin, M., Kim, S.Y., Keefe, D.L., Alukal, J.P., Boeke, J.D., and Yanai, I. (2020). Widespread Transcriptional Scanning in the Testis Modulates Gene Evolution Rates. *Cell* **180**, 248–262.e21.
- Yizhak, K., Aguet, F., Kim, J., Hess, J.M., Kübler, K., Grimsby, J., Frazer, R., Zhang, H., Haradhvala, N.J., Rosebrock, D., et al. (2019). RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science* **364**, eaaw0726.
- Yokoyama, A., Kakiuchi, N., Yoshizato, T., Nannya, Y., Suzuki, H., Takeuchi, Y., Shiozawa, Y., Sato, Y., Aoki, K., Kim, S.K., et al. (2019). Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **565**, 312–317.
- Yoshida, K., Gowers, K.H.C., Lee-Six, H., Chandrasekharan, D.P., Coorens, T., Maughan, E.F., Beal, K., Menzies, A., Millar, F.R., Anderson, E., et al. (2020). Tobacco smoking and somatic mutations in human bronchial epithelium. *Nature* **578**, 266–272.
- Zhang, L., and Vijg, J. (2018). Somatic Mutagenesis in Mammals and Its Implications for Human Disease and Aging. *Annu. Rev. Genet.* **52**, 397–419.
- Zhang, W., Bojorquez-Gomez, A., Velez, D.O., Xu, G., Sanchez, K.S., Shen, J.P., Chen, K., Licon, K., Melton, C., Olson, K.M., et al. (2018). A global transcriptional network connecting noncoding mutations to changes in tumor gene expression. *Nat. Genet.* **50**, 613–620.
- Zhang, L., Dong, X., Lee, M., Maslov, A.Y., Wang, T., and Vijg, J. (2019). Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. *Proc. Natl. Acad. Sci. USA* **116**, 9014–9019.
- Zhou, J., Theesfeld, C.L., Yao, K., Chen, K.M., Wong, A.K., and Troyanskaya, O.G. (2018). Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nat. Genet.* **50**, 1171–1179.
- Zhu, M., Lu, T., Jia, Y., Luo, X., Gopal, P., Li, L., Odewole, M., Renteria, V., Singal, A.G., Jang, Y., et al. (2019). Somatic Mutations Increase Hepatic Clonal Fitness and Regeneration in Chronic Liver Disease. *Cell* **177**, 608–621.e12.
- Zoghbi, H.Y., and Beaudet, A.L. (2016). Epigenetics and Human Disease. *Cold Spring Harb. Perspect. Biol.* **8**, a019497.
- Zong, C., Lu, S., Chapman, A.R., and Xie, X.S. (2012). Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622–1626.