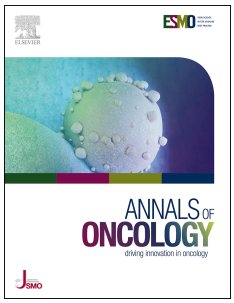


The effect of age on the acquisition and selection of cancer driver mutations in sun-exposed normal skin

B. Hernando, M. Dietzen, G. Parra, M. Gil-Barrachina, G. Pitarch, L. Mahiques, F. Valcuende-Cavero, N. McGranahan, C. Martinez-Cadenas

PII: S0923-7534(20)43199-0

DOI: https://doi.org/10.1016/j.annonc.2020.11.023

Reference: ANNONC 425

To appear in: *Annals of Oncology*

Received Date: 9 July 2020

Revised Date: 18 November 2020

Accepted Date: 29 November 2020

Please cite this article as: Hernando B, Dietzen M, Parra G, Gil-Barrachina M, Pitarch G, Mahiques L, Valcuende-Cavero F, McGranahan N, Martinez-Cadenas C, The effect of age on the acquisition and selection of cancer driver mutations in sun-exposed normal skin, *Annals of Oncology* (2021), doi: https://doi.org/10.1016/j.annonc.2020.11.023.

Article type: Original article

Title: The effect of age on the acquisition and selection of cancer driver mutations in sun-exposed normal skin

B. Hernando[1], M. Dietzen[2,3], G. Parra[4], M. Gil-Barrachina[1], G. Pitarch[5], L. Mahiques[5], F. Valcuende-Cavero[6], N. McGranahan[2,3,*], C. Martinez-Cadenas[1,*]

[1]Department of Medicine, Jaume I University of Castellon, Castellon, Spain

[2]Cancer Genome Evolution Research Group, University College London Cancer Institute, London, United Kingdom

[3]Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, United Kingdom

[4]Centre Nacional d'Anàlisi Genòmica-Centre de Regulació Genòmica (CNAG-CRG), Barcelona, Spain.

[5]Department of Dermatology, Castellon University General Hospital, Castellon, Spain.

[6]Department of Dermatology, La Plana University Hospital, Villarreal, Spain.

Correspondence:

*Dr. Nicholas McGranahan.

Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute 72 Huntley St, Bloomsbury, London WC1E 6DD, United Kingdom.

E-mail: nicholas.mcgranahan.10@ucl.ac.uk

*Dr. Conrado Martinez-Cadenas

Department of Medicine, Jaume I University of Castellon

Ave Sos Baynat 12071, Castellon, Spain

Telephone: 0034 964387607. E-mail:ccadenas@uji.es

ABSTRACT

Background: The accumulation of somatic mutations contributes to ageing and cancer. Sunlight is the principal aetiological factor associated with skin cancer development. However, genetic and phenotypic factors also contribute to skin cancer risk. This study aimed at exploring the role of photoaging, as well as other well-known epidemiological risk factors, in the accumulation of somatic mutations in cancer-free human epidermis.

Material and Methods: We deeply sequenced 46 genes in normal skin biopsies from 127 healthy donors, from which phenotypic data (including age, pigmentation-related genotype and phenotype) and sun exposure habits were collected. We determined the somatic mutational burden, mutational signatures, clonal selection and frequency of driver mutations in all samples.

Results: Our results reveal an exponential accumulation of UV-related somatic mutations with age, matching skin cancer incidence. The increase of mutational burden is in turn modified by an individual's skin phototype. Somatic mutations preferentially accumulated in cutaneous squamous cell carcinoma (cSCC) cancer genes and clonally expanded with age, with distinct mutational processes underpinning different age groups. Our results suggest loss of fidelity in transcription-coupled repair later in life.

Conclusion: Our findings reveal that aging is not only associated with an exponential increase in the number of somatic mutations accumulated in normal epidermis, but also with selection and expansion of cancer-associated mutations. Aged, sun-exposed normal skin is thus an extended mosaic of multiple clones with driver mutations, poised for the acquisition of transforming events.

KEY WORDS

Somatic mutation; Aging; Skin Phototype; Next-generation sequencing; UV exposure; normal epidermis; carcinogenesis; mutational spectrum

HIGHLIGHTS

- As skin cancer incidence increases, UV-related somatic mutations accumulate exponentially with increasing age in normal epidermis
- UV-related mutations do not only accumulate with age, but also repair processes are less efficient in elderly individuals
- Apart from age, an individual's skin phototype is key in the build-up of somatic mutations in healthy skin
- Distinct mutational processes may operate in different age groups
- Aging is associated with positive selection and expansion of clones with cancer mutations in sun-exposed normal skin

INTRODUCTION

Cancers arise as a result of somatic alterations occurring in the genomic DNA sequence of 'normal' cells. Somatic genomic alterations accumulate spontaneously in cells throughout a person's life as a result of errors occurring during cell replication and after exposure to mutagenic agents, such as certain chemicals in tobacco smoke or UV radiation from sunlight [1, 2].

The accumulation of clones of cells harbouring mutations across tissues may be expected to have functional consequences on the physiology of normal cells, contributing to ageing and promoting disease progression, as in the case of cancer. In this regard, recent sequencing studies have revealed that, in general, the somatic mutational burden and profile of physiological normal tissues seem to be surprisingly similar to those found in tumours from the same tissue [3–6]. These results suggest that the majority of somatic alterations may pre-date tumour formation, and only a small fraction of all somatic mutations in a cancer genome are, therefore, directly relevant in carcinogenesis, disease classification and treatment.

In the case of skin, a recent study analysing cancer-free epidermal samples from four individuals showed that the frequency of driver mutations in physiologically normal skin cells is remarkably high [7]. Sun-exposed epidermal cells carried a multitude of genetic alterations, and about 25-30% of these normal skin cells had already acquired at least one driver mutation, indicating that cancer-causing mutations are under strong positive selection even in epidermis maintaining normal physiological functions. The mutational profile found in eyelid samples, a chronically sun-exposed area, was similar to cutaneous squamous cell carcinomas (SCCs), but distinct from the profile characteristically observed in cutaneous melanoma [7]. This may be due to the low number of epidermal melanocytes (cells from which melanoma develops) in relation to keratinocytes [8], but also due to the fact that melanoma appears more frequently in sporadically, rather than chronically, sun-exposed areas of the body – those that are usually covered by clothing. This apparently paradoxical fact is attributable to injuries caused by an intermittent pattern of intense and acute sun exposure associated with recreational activities [9].

Skin cancer incidence worldwide reveals a clear relationship between pigmentation traits and sunlight damage, with individuals with fair skin and inability to tan showing greater cancer susceptibility. Cutaneous sensitivity to sunlight exposure (ability to tan *versus* tendency to burn) is defined by certain genetically determined pigmentation traits [10]. Thus, individuals carrying genetic variants associated with increased sun sensitivity should have higher somatic mutational rates, as they have reduced melanin levels, insufficient to protect the genome of epidermal cells from the mutagenic action of UV light [11, 12].

While previous studies have confirmed that seemingly normal cells harbour mutations, the key factors that determine which individuals are more prone to acquire and therefore accumulate somatic mutations remains unclear. In this study, we aimed at increasing our understanding of the accumulation of somatic mutations in the skin because of different patterns of sunlight exposure (Figure S1). Furthermore, we explored if there is an added risk to accumulate somatic mutations according to intrinsic characteristics of individuals (including age, and pigmentation-related genotype and phenotype) and to sun exposure habits.

MATERIALS AND METHODS

We focused on sequencing healthy skin samples from a large cohort of 123 cancer-free individuals, aged 11 to 92 years (mean of 58.50 years) (Table S1). Skin samples were collected from different body areas, classified according to the pattern of sunlight exposure as chronically-photoexposed (n=44) and intermittently-photoexposed (n=79). Only one sample was collected from each donor due to ethical reasons.

Deep sequencing of 46 genes implicated in skin cancers was performed on each epidermal sample, obtaining an average on-target coverage across samples of 923.44x (range 377.96-1657.37x).

To identify somatic mutations in the skin biopsies, we developed a bespoke pipeline. In brief, we first applied the Mutect2 tumour-only mode, due to the absence of a matched control sample. Putative germline variants and technical artefacts were subsequently removed by applying several key filtering steps, which were applied with the aim of having the maximum specificity. To validate the specificity of our pipeline, we applied it to an external dataset with available germline data. We found a very low rate of false positives (<0.05%) (Figure S2A). The proportion of mutations likely missed after applying our stringent filtering procedure was estimated as likely, between 10-20%.

Further details of sampling, sequencing, variant calling and filtering, and data analyses (including mutational burdens, mutational signatures, clonal selection, frequency of driver mutations, and copy number aberrations) are provided in Supplementary Methods.

RESULTS

Variation of mutational burden across samples

A total of 5,214 somatic mutations were identified in our dataset, with an average of 42.39 mutations per sample (range 2-169) (Figure 1A), corresponding to an average rate of 132.47 mutations per megabase (range 6.25-528.13). Consistent with previous studies of normal tissue, most mutations were present only in a small fraction of cells, evidenced by the fact that nearly all mutations exhibited a variant allele frequency (VAF) lower than 5% (Figure S2B).

The size of our cohort offered us a unique opportunity to directly quantify the key factors associated with mutational burden in sun-exposed skin samples. Indeed, several phenotypic and behavioural risk factors were collected from the participants, including sex, age, Fitzpatrick's skin phototype, history of sunlight exposure, body site pattern of sun exposure, signs of sun damage in the skin area biopsied and *MC1R* genotype.

To empirically assess the relative importance of each potential risk factor in contributing to the mutational burden of skin lesions, we first evaluated which type of model could best represent the data. A log-linear multivariate model was identified as providing the best fit, compared to linear, non-linear, quadratic and cubic models (Figure S3). When using this log-linear model, the total variance of mutational burden explained by all variables combined was high (adjusted-R^2 = 49.88%). Age explained the largest proportion of the total variance (55.16%; Figure 1B). An individual's skin phototype was, surprisingly, the second strongest predictor, explaining 17.92% of the mutational burden variance across samples. Our results suggested a significant decrease (β<0) in the number of somatic mutations accumulated in skin samples from individuals with high skin phototypes (skin types that normally tan after sunlight exposure), as compared with individuals with low skin phototypes (skin types that normally burn rather than tan after sunlight exposure) (Table S2). The lack of tanning capacity confers greater susceptibility to develop skin cancers, due in part to the inability to protect against UV-related DNA damage [10].

Contrary to previous studies [13], no association, in univariate or multivariate analysis, was found between the somatic mutational load and the genotype in *MC1R*, a key pigmentation-related gene determining the ability to respond to UV exposure [14]. Moreover, a lack of significant association between mutational burden and the pattern of body site photoexposure (chronic *versus* intermittent) was observed after including all risk factors in the model, explaining only 7.82% of the variance of the total number of mutations accumulated across samples. Our data suggested a comparable effect of intermittent exposure to sunlight, perhaps while on recreational activities, and continuous exposure, through spending a large amount of daylight time outdoors, with regards to the accumulation of somatic mutations in normal epidermis. None of the other risk factors remained significantly associated with mutational burden in the multivariate model (Figure 1C).

Given that age and phototype were the two most significant contributors to explain mutational burden, we quantitatively assessed age-related mutational burden according to skin phototype. For each skin

phototype, we determined the mutation rate increased per year of life (β) in the panel of 46 genes (0.32 Mb) using a log-linear model (Figure 1D). Nonparametric bootstraps (1000 runs) were conducted to estimate the 95% confidence intervals (CI95) of the effect of age. Somatic accumulation rates per year of life increased across skin phototypes. Strikingly, by the age of 65 years old, we expect 68.63 mutations (CI95: 40.65-122.73) to have accumulated in the 46 selected genes for phototype I, whereas only 14.22 mutations (CI95: 8.16-20.54) are expected for type IV.

Taken together, these results suggest a profound influence of age on the accumulation of somatic mutations in normal skin. This age-associated rise of mutational burden is in turn modified by an individual's skin phototype, reflecting the inability of UV-sensitive individuals to protect against UV-related DNA damage. We note that although sequencing coverage had an impact on the number of mutations called, our overall conclusions were not affected by discrepancies in sequencing coverage across samples (Supplementary Text and Figure S4).

Aging and the rise of UV-associated mutations

To explore the mutational processes underpinning the accumulation of mutations in sun-exposed samples, we considered the specific substitution types. In the cohort as a whole, we observed a predominance of C>T and CC>TT mutations at dipyrimidine sites (TpC or CpC), likely reflecting repair of 6,4-photoproducts and the production of cyclobutane pyrimidine dimers (CPDs) in response to UV-induced DNA damage [15](Figure 2A). These C>T substitutions were preferentially accumulated on the non-transcribed compared to the transcribed strand (*P*-value = 4.21×10^{-4}), consistent with the activity of transcription-coupled nucleotide excision repair (Figure 2B). In addition, we also observed an enrichment of T>C/A>G mutations at CTT sites, potentially caused by indirect DNA damage after UV radiation – greater incorporation of G, rather than A, opposite thymidine and cytidine photodimers by translesion polymerases (Figure 2A). This is in line with the fact that these mutations tended to accumulate in the transcribed rather than in the non-transcribed strand (*P*-value = 5.43×10^{-5}) (Figure 2B).

To quantify the presence of specific mutational signatures, which may reflect underlying mutational processes, we applied deconstructSigs [16] to the combined set of mutations within the cohort. The specific mutational spectrum observed in our samples mainly resembles the COSMIC base substitution signatures related to UV radiation (mutational signatures SBS7a-d) (Figure S5B).

The relative contribution of UV-related mutational pathogenesis varied significantly with age (Figure 2C). Mutational signatures directly related to UV contributed 68.14% of all mutations detected after the age of 63 (the cohort median age), but only 46.59% of the mutational burden in individuals younger than 63 (Figure S5B). This observation is in accordance with the fact that non-melanoma skin tumours typically occur at advanced age and are related to cumulative sun exposure [17]. In fact, we observed that the increase of UV-mutations in normal skin with age follows a similar exponential trend than skin cancer incidence in Spain (data downloaded from the Global Cancer Observatory, http://gco.iarc.fr)(Figure 2D). This exponential growth of UV-related mutations was not affected by the frequency of variant reads (Figure S4D). The main difference between the elderly and younger individuals in terms of mutational spectra, apart from the fraction of UV-related mutations, was related to the fraction of T>C/A>G mutations at CTT contexts, being proportionally higher in younger individuals. This mutational pattern closely matched the SBS17a signature, a COSMIC signature with unknown aetiology that has been shown to contribute to cutaneous melanoma [18]. To gain insight into the context dependency of this T>C/A>G substitutions, we explored the local sequence contexts (from -5 to +5 positions) and observed a specific pattern of contextual preference (CTTTT) in normal skin samples biopsied from younger individuals (Figure S7A). Additionally, the degree of transcriptional strand bias detected for T>C/A>G substitutions was substantially higher in younger individuals (Figure S7B). These differences in the mutational spectra between different age groups can be attributed to the exponential increase of C>T substitutions with age, coupled with a relatively constant, linear increase in T>C/A>G mutations with age

(Figure S7D). These observations suggest that, after the age of 60, the mutagenic process related to sunlight is the predominant contributor to the accumulation of somatic mutations in skin.

To confirm that the differences shown in mutational spectra with age was not solely due to unequal representation of sunlight exposure patterns between subgroups (the proportion of chronically-photoexposed skin samples is higher in the elderly group than in the young group), we implemented deconstructSigs splitting samples of each age group according to the pattern of sunlight exposure of the skin tissue biopsied. These results confirmed that the sunlight exposure of the skin sample was not a confounding factor in the age-specific mutational pattern observed (Figure S7E).

Taken together, our results reveal that mutational processes may sculpt the genome in distinct ways as we age. While SBS17, a signature that seems to accumulate at a constant rate during life, UV-related mutational processes appear to become the dominant cause of mutational acquisition with age. This suggests a decline in the ability to repair UV-induced mutagenic lesions later in life. Consistent with this, we observed a trend for decreased transcription coupled repair, as captured by transcription strand bias at C>T mutations, in older compared to younger individuals (*P*-value = 0.078; Figure S7C).

Positive selection of driver mutations in normal skin

We next considered whether protein-altering somatic mutations were subjected to positive selection and the interplay between age and selection. We evaluated the footprint of positive selection in two orthogonal ways, by quantifying the excess of non-synonymous mutations as well as by estimating clone sizes.

The majority of the normal sun-exposed skin samples harboured multiple protein-altering mutations, even though these epidermal samples were histologically benign (Figure 3A). The catalogue of recurrently mutated genes in normal skin was almost identical to that of cSCC, with *TP53*, *NOTCH1*, *NOTCH2* and

FAT1 being the most recurrently mutated genes. We also identified samples with canonical hotspot mutations with therapeutic relevance (according to the database of curated mutations) in several oncogenes such as *BRAF* (F595L, V600E), *HRAS* (G12D, G12V), *PIK3CA* (P471L, E542K, E726K, M1043I, H1047L), and *FGFR3* (R248C, S249C, Y373C, A391E). In total, thirty-four (53.12%) skin samples from elderly individuals carried at least one of the 50 different relevant disease-causing mutations identified. Conversely, only eight (12.70%) young individuals harboured at least one of these canonical mutations (Figure 3A).

In order to quantify the extent of positive selection driving minor clonal expansion in normal skin samples in both elderly and young donors, we considered the ratio of missense, nonsense and essential splicing mutations compared to synonymous mutations using the dNdScv package [19]. Our results provided evidence of significant positive selection when considering mutations in all known cancer genes as a whole, while a lack of selection pressure is observed in non-cancer genes (Figure 3B). Indeed, using the dN/dS ratios, we found that a considerable percentage of mutations (50.42%) accumulated in cancer genes may confer a growth advantage and thus may be positively selected in normal skin. However, non-cancer genes seemed to accumulate neutral mutations that may not be affected by natural selection and, hence, their growth/expansion can be attributed to genetic drift (Figure S9B).

At a single gene level, we had power to detect significant positive selection in three out of 46 sequenced genes for normal skin samples collected from both young and elderly individuals (*TP53, NOTCH1,* and *FAT1*) (Figure 3C). Notably, all three genes were also shown to have a significant excess of non-synonymous mutations in normal sun-exposed epidermis [7], and have been shown to be drivers of cSCC [20, 21]. Interestingly, *NOTCH2* and *CDKN2A*, two tumour suppressor genes recurrently mutated in skin cancers [20–22], only had a significant excess of truncating mutations in samples from elderly individuals. *NOTCH2* was also the gene most frequently subject to copy number aberrations (Supplementary Text and Figure S8).

The majority of sun-exposed samples (79.67%) carried at least one non-synonymous mutation in a cancer gene under positive selection (*TP53*, *NOTCH1*, and *FAT1*). Samples harbouring mutations in *TP53*, *NOTCH1* and *FAT1*, as well as other canonical hotspot mutations, had a significant increase in the overall number of mutations. Although this increase seemed to be largely influenced by the participant's age, the increase in mutation count associated with carrying protein-altering mutations in these positively selected genes was observed in both age groups (Figure S10A).

To confirm whether these protein-altering mutations are subject to positive selection, thereby resulting in clonal expansions, we scrutinized the VAF spectra of somatic mutations in each age group. The average VAF of non-synonymous mutations per gene was significantly higher than that of synonymous mutations in normal skin samples from both young (*P*-value = 9.96×10^{-6}) and elderly donors (*P*-value = 6.21×10^{-8}; Figure 3D). The highest average VAF was observed in cancer-associated genes under significant positive selection (*NOTCH1*, *FAT1* and *TP53*), suggesting that these selectively advantageous mutations may appear early and expand with age in sun-exposed normal skin. However, the modest VAF of even these somatic mutations (mean VAF of 0.014 and 0.005 for elderly and young donors, respectively) suggested that they are present in only a small subset of skin cells, therefore remaining in a minority of sun-exposed cells.

Overall, frequencies of both non-synonymous and synonymous mutations significantly increased in epidermal samples from elderly donors compared to those collected from young donors (*P*-value = 8.30×10^{-9} and *P*-value = 1.53×10^{-7}, respectively; Figure 3D and S10B), suggesting tolerance and selection for larger clones with age. Intriguingly, while frequencies of synonymous mutations were found to exhibit a linear relationship with patient age, an age-dependent exponential growth was observed for non-synonymous mutations (Figure 3E). Conceivably, the linear expansion of synonymous mutations (which are generally expected to be neutral) with age reflects fixation by drift, while the exponential growth of

clones carrying protein-affecting mutations (which are generally expected to increase cell fitness) with age indicates positive selection. Although there is a broad VAF spectrum in all samples, the association of clonal expansion with age and mutation effect was also observed when all mutations were analysed (Figure S9A).

Taken as a whole, our data suggest that sun-exposed normal skin already harbours protein-affecting mutations in cancer genes, with 5 genes subject to statistically significant positive selection. Furthermore, we find that skin from elderly individuals not only harbours more mutations, but a larger fraction of these reflect positive selection and expansion.

DISCUSSION

Like other tissues, skin undergoes chronological aging that might lead to its functional decline, which is accelerated by chronic sun damage. This study aimed to explore the role of photoaging, as well as other well-known epidemiological risk factors, in the accumulation of somatic mutations in cancer-free human epidermis. Our experimental design was focused on analysing the mutational landscape in a large cohort of subjects with a wide range of ages and phenotypic characteristics.

A profound variability in terms of mutational burden and driver mutations was observed among individuals, which was largely explained by age. An exponential increase of mutation count with increasing age was observed in the panel of 46 genes. In addition, normal skin samples collected from donors with low skin phototypes (individuals with fair skin who tend to burn rather than tan after being exposed to sunlight) tended to accumulate a higher number of somatic mutations over time. Fair-skinned individuals are more severely affected by photoaging [24]. Therefore, the increased risk of developing skin cancer, at least for sporadic cancers, associated with these phenotypes may be due to the presence of a higher reservoir of mutant cells waiting to acquire more cancer-driving mutations, evade clonal growth control and initiate malignant transformation. By contrast, sun exposure habits were not significantly

correlated to mutational burden variability across samples. Note that our samples were collected from Spanish individuals living in a region with relatively high UV index and pleasant weather throughout the year. Perhaps a higher effect of sun exposure pattern by body site would be found in populations from more northerly latitudes, exposed to sunlight mainly during summer vacations. Furthermore, the retrospective and subjective nature of some behavioural risk factors, particularly those related to sun exposure habits, raises the potential for recall bias. Therefore, validation in independent and larger cohorts will be needed to further analyse the association between behavioural risk factors and mutational burden. Additionally, future efforts should be done to explore other factors (i.e. repair mechanisms efficiency, antioxidant capability, etc.) that may impact on somatic mutation acquisition among individuals.

Analysing a large cohort with a wide age range has allowed us to investigate the timing of somatic mutation accumulation in normal skin. The number of T>C mutations accumulated in normal skin appeared to accumulate at a steady linear rate. In contrast, an exponential rise of UV-related mutations with increasing age was evident in normal epidermis. This exponential increase of UV-related mutations, together with the fact that the ratio of these mutations accumulated in the non-coding *versus* the coding strand tend to decrease with age, could be related to the decline of nucleotide excision repair (NER) capacity with age [25]. NER is a highly evolutionarily conserved mechanism for repairing bulky DNA lesions resulting, among others, from sunlight exposure. The importance of NER activity in the prevention of skin cancer is denoted by the extreme sensitivity to sunlight and severe predisposition to UV-induced skin cancers of patients with the inherited disorder xeroderma pigmentosum, in which genes encoding for the different components of the NER cascade are mutated [26].

The age-dependent exponential increase of skin cancer incidence in Spain followed a similar trend than the accumulation of UV-related mutations in normal skin. Apart from the decline of NER function, skin photoaging has also been linked to impaired skin homeostasis [27]. It is thought that strategies for tissue

maintenance have a noteworthy impact on cancer incidence, consistent with the dramatic increase of cancer incidence with tissue and stem cell fitness decline [28]. Aged/damaged tissue microenvironment may therefore provide an opportunity for clones with a selective advantage to expand, and that may be the reason why mutational profile notably differs between normal skin samples from young and adult donors.

Our results revealed an enrichment of driver mutations in most normal sun-exposed skin samples, especially in those collected from elderly individuals. In addition, there was a marked overrepresentation of protein-altering mutations in several cSCC driver genes, especially in *NOTCH1*, *TP53* and *FAT1*, likely reflecting positive selection. As previously seen in blood [29], the quantitative analysis of selection by measuring the excess of non-synonymous mutations (dN/dS ratios) were in line with clone size distributions. Analysing the VAF spectra of mutations accumulated in normal skin, we noted that clone size seems to be closely related to age (time during which it has been expanding) and to the variant impact on cell growth (rate at which it has been expanding).

However, because these positively-selected genes have been shown to be frequently mutated in normal skin [7, 23], it seems unlikely that these mutations alone confer a sufficient growth advantage to engender cancer development. Indeed, although clones carrying these selectively advantageous mutations have expanded in normal skin, it is notable that only 17 out of 123 donors (13.82%) carried more than one mutation with clinical relevance in their skin. These mutant clones also seem to be relatively small, suggesting even in these cases the somatic alterations are mostly in distinct clones. Indeed, using dN/dS ratio, the number of canonical mutations estimated per cell was only 0.21 (range of 0.07-0.72) in samples carrying at least two clinically relevant mutations. We speculate that most cells carrying driver mutations may not have yet acquired the right combination of mutated genes for expanding and culminating in the development of malignancy.

Cellular senescence, an irreversible proliferative arrest triggered by exogenous and endogenous stresses, may be a plausible explanation for the limited expansion of clones carrying cancer-causing mutations in normal tissues. Oncogenic-induced senescence is considered a crucial protective mechanism against cell transformation. Several reports from animal models support the idea that cell senescence may occur in tissues after acquiring a mitogenic mutation, preventing carcinogenesis at an initial step [30, 31]. The biology of naevi is a clear example of cellular senescence following an initial activating oncogenic mutation, normally in *BRAF* or *NRAS* [32, 33]. Melanocytic nevi are clonal proliferations of non-malignant melanocytic cells, which can remain non-growing for many years, but also can act as precursors of melanoma if cells overcome senescence. Interventions that favour oncogene-induced senescence may help restrict the growth of clones carrying cancer-causing mutations in normal tissues and thus tumour progression.

The question arising from our observations, together with those shown in different sequencing studies previously performed on healthy tissues [3–5, 7, 34, 35], is whether or not targeting these early mutations recurrently found in normal tissues will be relevant for preventing carcinogenesis. Further efforts should be done to delineate the succession of genetic alterations needed for malignant transformation of physiologically normal tissues to premalignant precursor lesions, and finally to tumours, with the aim of discriminating drivers of the disease from the non-pathogenic mutational landscape.

ACKNOWLEDGMENTS

We are extremely grateful to all the volunteers for giving their consent to take part in this study, as well as to all the medical specialists for supervising phenotype collection of all samples. This work is supported in part by the Jaume I University of Castellon (UJI-A2016-13). B.H. is funded by the Jaume I University of Castellon under a Postdoctoral Research contract (POSDOC-A/2018/07), and received additional funding for conducting a research exchange at the UCL Cancer Institute (E-2019-34). N.M is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (Grant Number 211179/Z/18/Z), and also receives funding from CRUK, Rosetrees, and the NIHR BRC at University College London Hospitals, and the CRUK University College London Experimental Cancer Medicine Centre.

AUTHOR CONTRIBUTIONS

B.H. treated skin tissues, performed DNA extraction, prepared samples for sequencing, conducted bioinformatics analysis, and wrote the manuscript. M.D. helped with mutational signatures analysis. G.P. helped with variant calling and filtering, and copy number aberration analysis. M.G-B. helped with lab work. G.P, L.M. and F.V-C collected tissue samples and phenotypic data of donors. N.M. provided expertise in interpreting the results, supervised bioinformatics analyses, and wrote the manuscript. C.M-C conceived the project, supervised the study, and wrote the manuscript. All co-authors provided critical feedback and contributed to manuscript preparation.

DECLARATION OF INTERESTS

N.M. has received consultancy fees and has stock options in Achilles Therapeutics. All remaining authors have declared no conflicts of interest.

DATA AVAILABILITY

Sequencing data has been deposited at the European Genome–phenome Archive (EGA) under the

accession code EGAS00001004279. The somatic mutations found in all samples are listed in the Supplementary Dataset S1. Clinical data of each donor can be found in the Supplementary Dataset S2. Any other relevant data can be obtained from the corresponding authors upon reasonable request.

FIGURE LEGENDS

Figure 1. Risk factors contributing to the accumulation of somatic mutations in normal epidermis. **(A)** Total number and type of somatic mutations detected across the 46 genes sequenced in each sample. Clinical and demographic characteristics are presented below. **(B)** Relative importance of predictors included in the log-linear multivariate regression model. The total variance explained by the model (adjusted-R^2 = 49.88%) is decomposed in order to know the individual contribution (effect size) of each predictor. Asterisk denotes significant predictors. **(C)** Heatmap showing the *P*-values of univariate and multivariate log-linear model coefficients from the analyses-of-variance (ANOVA) tables. **(D)** A log-linear regression is used for analysing the age effect on the accumulation of somatic mutations for each skin phototype. Solid lines represent the bootstrapped mean of the slope, and shaded areas its bootstrapped 95% confidence intervals (CI95). Regression coefficients (β) for each skin phototype are also presented.

Figure 2. Spectrum of somatic mutations in normal epidermis. (A) Bar plot showing the fraction of single (top) and doublet (bottom) base substitutions found in each of the possible 96 trinucleotide contexts (strand independent). **(B)** Relative number of each substitution type present on the transcribed (dark shading) and untranscribed strand (light shading). Asterisks indicate significant transcriptional strand asymmetries (Poisson test). **(C)** Mutational spectra in samples from young (left) and elderly donors (right). Heatmaps show the fraction of each trinucleotide change in each sample (middle). Bar plots represent the contribution mean of each 96-mutation type per age group (top). Clinical and demographic characteristics are presented next to each sample (right). **(D)** Age-dependent increase of both UV-

mutation accumulation and skin cancer incidence. Data of skin cancer incidence in Spain was downloaded from the Global Cancer Observatory (http://gco.iarc.fr).

Figure 3. Occurrence, positive selection and expansion of driver mutations in normal skin with age. **(A)** Heatmap showing the distribution of recurrent non-synonymous mutations per coding kilobase of sequence for each one of the 46 genes targeted across all normal skin samples. Genes are sorted from higher (top) to lower (bottom) number of non-synonymous mutations per kb. Percentage of normal skin samples carrying at least one non-synonymous mutation in each gene is shown in brackets. Clinical and demographic characteristics are presented above each sample. The *KMT2B* gene is not included in this plot since none non-synonymous mutation was found across samples. Red asterisks denote samples harbouring at least one canonical mutation in a specific gene. **(B)** Global dN/dS values (top) and frequency of driver mutations (bottom) calculated by taking together all cancer and non-cancer genes in normal skin biopsied from both young and elderly donors. Error bars represent 95% confidence intervals. Percentage of driver mutations was only calculated when dN/dS ratios denoted positive selection (dN/dS > 1). **(C)** dN/dS ratios for each of the 46 target genes. Genes under significant positive selection in both age groups are coloured in red, while genes positively selected only in the elderly group are coloured in blue (overall q-value < 0.05). Genes are sorted from higher (bottom) to lower (top) significant value in the elderly group. **(D)** Distribution of VAFs of somatic non-synonymous and synonymous mutations per gene. Red dots denote positively selected genes in both young and elderly groups (*FAT1*, *NOTCH1*, and *TP53*), and blue dots indicate genes under positive selection only in the elderly group (*NOTCH2* and *CDKN2A*). Dots representing the other genes sequenced are coloured in grey. A Wilcoxon-Mann-Whitney test is used for testing differences among mutation types in each age group. **(E)** Expansion of non-synonymous and synonymous mutations with age. Each dot represents the mean VAF of somatic non-synonymous and synonymous mutations per sample. Fitted lines show that the variant allele frequencies of non-synonymous mutations (with a positive fitness effect) grew exponentially with age

denoting selection, while synonymous mutations (with a neutral fitness effect) expanded linearly with age indicating fixation by drift.

SUPPLEMENTAL MATERIAL

Supplemental material includes supplementary methods and references, eleven figures, three tables and two datasets.

Figure S1. Schematic overview of the experimental design.

Figure S2. Evaluation of variant calling and filtering. (A) Validation of filtering procedure efficiency in an independent dataset comprising tumour and adjacent benign FFPE samples collected from six melanoma patients. A noteworthy decrease of false positive rates (proportion of germline variants in the set of somatic mutations) was denoted after applying the procedure for filtering out germline variants. **(B)** Histogram of somatic mutations identified by VAF. Most somatic mutations remain in a subclonal state with low VAFs (VAF << 5%). **(C)** Spectra of mutation sets removed after applying a specific filtering step. All mutational spectra are very different from the typical UV-related mutational spectrum, indicating that the filtered variants are unlikely to be real somatic mutations. **(D)** Global dN/dS ratios estimated before and after mutation filtering called with Mutect2 tumour-only mode. The global dN/dS << 1 denotes contamination of germline variants and/or technical artefacts in the non-filtered dataset of somatic mutations. This problem seems to be solved after applying the different filtering steps (dN/dS > 1). Error bars denote 95% confidence interval. **(E)** Results of applying a log-linear regression model in the non-filtered mutation dataset for predicting the number of mutations per sample. The low variance explained by the model (adjusted-R^2 = 6.08%) denotes that the non-filtered list of mutations includes a large number of likely false positive calls.

Figure S3. Selection of the best model explaining the age-dependent increase of somatic mutations in normal skin. Model selection was performed using the Akaike information criterion (AIC).

Figure S4. Evaluation of the impact of mutation detectability on mutational burden variability across samples. (A) Scatter plots showing a high correlation between the number of mutations predicted from the original dataset and from each down-sampled dataset. β values denote the gradient of impact of coverage metric on mutational burden estimates. **(B)** Box plots showing the ratio of increase, expressed as fold change (FC), in mutational burden estimates for each increase in coverage metrics. **(C)** Heatmap showing the analyses-of-variance (ANOVA) *P*-values of multivariate log-linear model coefficients of each dataset. DS, down-sampled. N, sample size of the dataset. **(D)** Scatter plot showing the mean VAF and the number of all mutations found per sample. This plot shows that mutation detectability did not significantly influence the number of mutations found across samples. **(E)** Heatmap showing the *P*-values of univariate log-linear model coefficients from the ANOVA tables. The normalized mutational burden of each sample was calculated by dividing the number of mutations per sample by the mean VAF of all mutations found in the sample.

Figure S5. Mutational spectra in normal skin. (A) Heatmap showing the fraction of each 96-mutation type per sample. Clinical and demographic characteristics are presented above each sample. **(B)** Percentage of substitutions attributed to each one of the six mutational signatures for all mutations from all 127 samples together (Total), as well as for all mutations included in each age subgroup.

Figure S6. Linearization of the exponential increase of both UV-mutation accumulation and skin cancer incidence with age. Logarithmic transformation of data displayed in Figure 2D. The relatively high R-squared values denote that a high proportion of the total variance in UV-mutation accumulation (blue dots) and in skin cancer incidence (black dots) is explained by the respective log-linear model.

Figure S7. Age-related mutational spectra in normal skin. (A) Local mutational context of T>C substitutions in samples biopsied from young and elderly donors. **(B)** Relative number of each substitution type present on the transcribed (dark shading) and untranscribed strand (light shading) in samples biopsied from young and elderly donors. Asterisks indicate significant transcriptional strand asymmetries (Poisson test). **(C)** Percentage of C>T mutations per strand in young and elderly individuals. ANOVA test used for comparing the ratio of non-coding/coding C>T mutations between age groups. **(D)** Age-dependent increase of T>C substitutions. **(E)** 96-barplot depicting the number of mutations observed at each trinucleotide context taking together all samples biopsied from young and elderly individuals (Total), as well as splitting samples of each age group by the body site pattern of sun exposure (Chronically- and Intermittently-photoexposed).

Figure S8. Occurrences of copy number alterations in the 46 cancer genes across samples. (A) Heatmap showing the significant copy number events detected in our cohort. **(B)** Scatter plots of four samples showing allelic imbalances in *NOTCH2*. The b-allele fraction (BAF) and 95% confidence interval of each germline heterozygous polymorphism in *NOTCH2* is shown. Red dots denote a deviation of the observed fraction of reads supporting the minor allele from the expected fraction (dashed lines), which is calculated by averaging the BAFs of all germline heterozygous SNPs in each sample and in all samples.

Figure S9. Mutation effect in cell fitness, selection and clonal expansion. (A) VAF spectra of non-synonymous and synonymous mutations in both driver and non-driver genes according to age. **(B)** Global dN/dS values (top) and frequency of driver mutations (bottom) estimated in driver and non-driver genes according to mutation frequencies. Percentage of driver mutations was only calculated when dN/dS ratios denoted positive selection (dN/dS > 1). Mutations were divided into four equal parts according to their VAF. VAF Q1, mutations with VAF values below the first quartile. VAF Q2, mutations with VAF values

between the first and second quartiles. VAF Q3, mutations with VAF values between the second and third quartiles. VAF Q4, mutations with VAF values above the third quartile.

Figure S10. Clonal expansion of clones with oncogenic mutations. (A) Number of non-synonymous mutations per sample in normal skin samples non-carriers or carriers of one or multiple non-synonymous mutations in *NOTCH1*, *TP53* and *FAT1*, as well as in normal skin without or with canonical hotspot mutations. Each dot represents a sample and is coloured according to the donor's age. For avoiding the confounding effects of age, samples were stratified according to donor age for statistical analyses. In panels comparing more than two groups, a Kruskall-Wallis (KW) test is used for testing differences among groups. In panels comparing two groups, a Wilcoxon-Mann-Whitney (WMW) test is used for testing differences among groups. **(B)** Heatmap showing the mean VAF of all non-synonymous mutations found per gene across normal samples collected from young and elderly individuals.

Figure S11. Coverage and mutational burden across genes and samples. (A) Plot showing the number of mutations per gene across all samples (bar plot, top) and the mean coverage per gene and sample (box plot, bottom). Genes in the x-axis sorted by mean coverage across samples. Blue line indicates the mean coverage across all samples. **(B)** Plot showing the number of mutations per sample (bar plot, top) and the mean coverage per sample (bar plot, bottom). Samples in the x-axis sorted by mean coverage across all sequenced regions. Blue line indicates the mean coverage across all samples. **(C)** Scatter plot showing the coverage and number of mutations per gene. **(D)** Scatter plot showing the coverage and number of mutations per sample coloured by skin phototype (left) and per age group (right). These plots show that coverage did not significantly influence the number of mutations found across genes and/or across samples.

Table S1. Demographic and clinical data of all Spanish donors.

Table S2. Log-linear modelling of the accumulation of somatic mutations in normal skin.

Table S3. Information from literature about the function and the role in carcinogenesis of the list of genes sequenced in this study.

Dataset S1: List of all somatic mutations included in the final list

Dataset S2: Clinical and phenotypic data of donors

REFERENCES

1. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science 2015; 349(6255):1483–1489.

2. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature 2009; 458(7239):719–724.

3. 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.

4. 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.

5. 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.

6. Martincorena I. Somatic mutation and clonal expansions in human tissues. Genome Med. 2019; 11(1):35.

7. Martincorena I, Roshan A, Gerstung M et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science 2015; 348(6237):880–886.

8. Hoath SB, Leahy DG. The organization of human epidermis: functional epidermal units and phi proportionality. J. Invest. Dermatol. 2003; 121(6):1440–1446.

9. Gandini S, Sera F, Cattaruzza MS et al. Meta-analysis of risk factors for cutaneous melanoma: II. Sun exposure. Eur. J. Cancer Oxf. Engl. 1990 2005; 41(1):45–60.

10. Scherer D, Kumar R. Genetics of pigmentation in skin cancer--a review. Mutat. Res. 2010; 705(2):141–153.

11. Rijken F, Bruijnzeel PLB, van Weelden H, Kiekens RCM. Responses of black and white skin to solar-simulating radiation: differences in DNA photodamage, infiltrating neutrophils, proteolytic enzymes induced, keratinocyte activation, and IL-10 expression. J. Invest. Dermatol. 2004; 122(6):1448–1455.

12. Brenner M, Hearing VJ. The Protective Role of Melanin Against UV Damage in Human Skin. Photochem. Photobiol. 2008; 84(3):539–549.

13. Robles-Espinoza CD, Roberts ND, Chen S et al. Germline MC1R status influences somatic mutation burden in melanoma. Nat. Commun. 2016; 7:12064.

14. García-Borrón JC, Abdel-Malek Z, Jiménez-Cervantes C. MC1R, the cAMP pathway, and the response to solar UV: extending the horizon beyond pigmentation. Pigment Cell Melanoma Res. 2014; 27(5):699–720.

15. Ravanat JL, Douki T, Cadet J. Direct and indirect effects of UV radiation on DNA and its components. J. Photochem. Photobiol. B 2001; 63(1–3):88–102.

16. Rosenthal R, McGranahan N, Herrero J et al. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016; 17:31.

17. Leiter U, Eigentler T, Garbe C. Epidemiology of skin cancer. Adv. Exp. Med. Biol. 2014; 810:120–140.

18. Alexandrov LB, Kim J, Haradhvala NJ et al. The repertoire of mutational signatures in human cancer. Nature 2020; 578(7793):94–101.

19. Martincorena I, Raine KM, Gerstung M et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell 2017; 171(5):1029-1041.e21.

20. Inman GJ, Wang J, Nagano A et al. The genomic landscape of cutaneous SCC reveals drivers and a novel azathioprine associated mutational signature. Nat. Commun. 2018; 9(1):3667.

21. Pickering CR, Zhou JH, Lee JJ et al. Mutational landscape of aggressive cutaneous squamous cell carcinoma. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 2014; 20(24):6582–6592.

22. Hayward NK, Wilmott JS, Waddell N et al. Whole-genome landscapes of major melanoma subtypes. Nature 2017; 545(7653):175–180.

23. Yizhak K, Aguet F, Kim J et al. RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. Science 2019. doi:10.1126/science.aaw0726.

24. Fisher GJ, Kang S, Varani J et al. Mechanisms of Photoaging and Chronological Skin Aging. Arch. Dermatol. 2002; 138(11):1462–1470.

25. Gorbunova V, Seluanov A, Mao Z, Hine C. Changes in DNA repair during aging. Nucleic Acids Res. 2007; 35(22):7466–7474.

26. Lehmann J, Seebode C, Martens MC, Emmert S. Xeroderma Pigmentosum - Facts and Perspectives. Anticancer Res. 2018; 38(2):1159–1164.

27. Panich U, Sittithumcharee G, Rathviboon N, Jirawatnotai S. Ultraviolet Radiation-Induced Skin Aging: The Role of DNA Damage and Oxidative Stress in Epidermal Stem Cell Damage Mediated Skin Aging. Stem Cells Int. 2016; 2016:7370642.

28. Rozhok AI, DeGregori J. The evolution of lifespan and age-dependent cancer risk. Trends Cancer 2016; 2(10):552–560.

29. Watson CJ, Papula AL, Poon GYP et al. The evolutionary dynamics and fitness landscape of clonal hematopoiesis. Science 2020; 367(6485):1449–1454.

30. Collado M, Gil J, Efeyan A et al. Tumour biology: senescence in premalignant tumours. Nature 2005; 436(7051):642.

31. Braig M, Lee S, Loddenkemper C et al. Oncogene-induced senescence as an initial barrier in lymphoma development. Nature 2005; 436(7051):660–665.

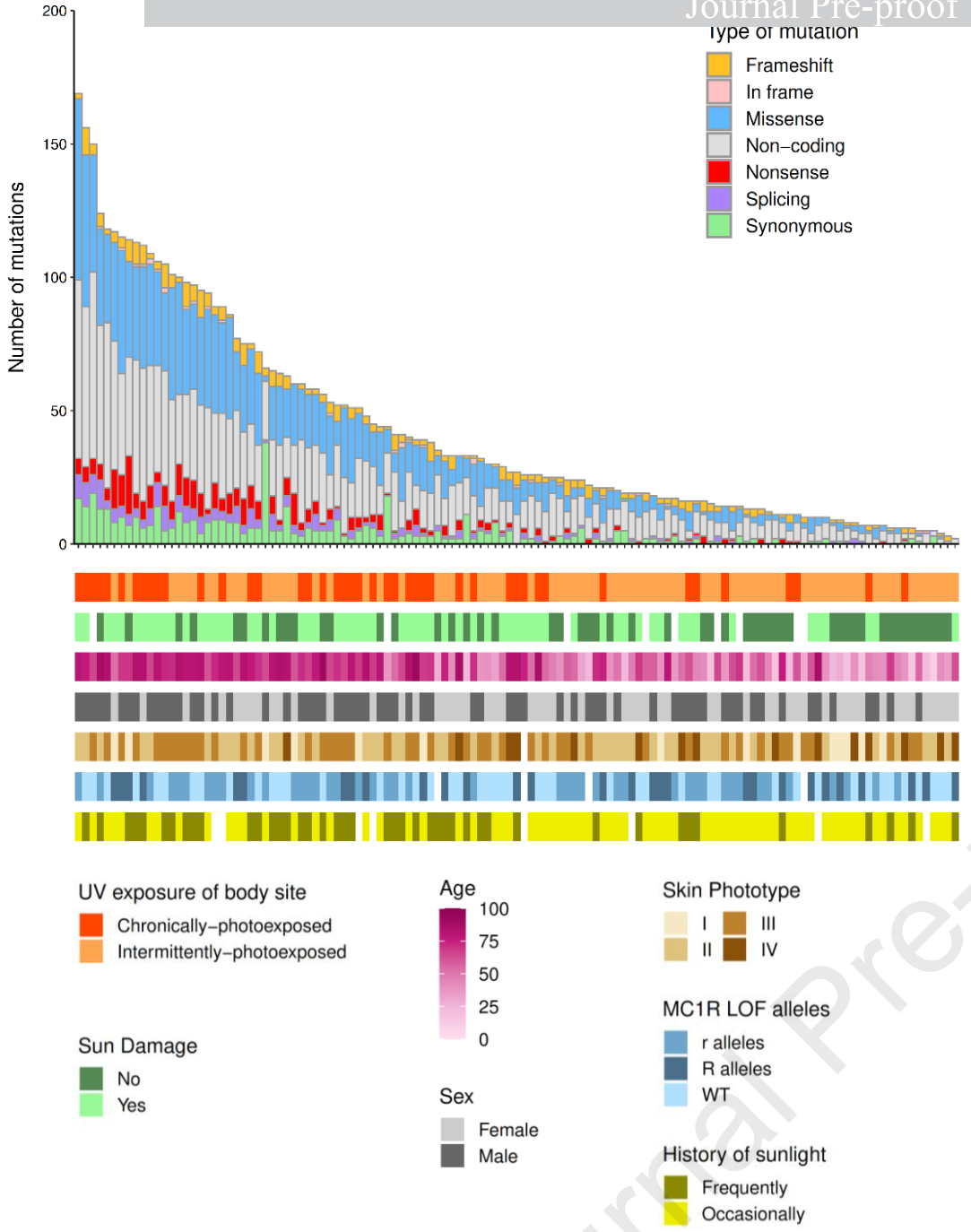
32. Bennett DC. Human melanocyte senescence and melanoma susceptibility genes. Oncogene 2003; 22(20):3063–3069.

33. Bennett DC, Medrano EE. Molecular regulation of melanocyte senescence. Pigment Cell Res. 2002; 15(4):242–250.

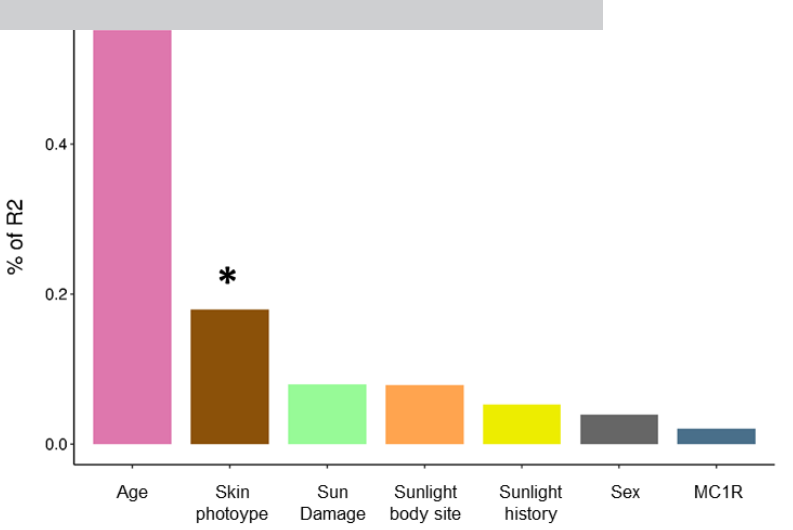
34. Martincorena I, Fowler JC, Wabik A et al. Somatic mutant clones colonize the human esophagus with age. Science 2018; 362(6417):911–917.

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.

A



B



C

	Multivariate log-linear model	Univariate log-linear model
Age	3.84e−15	2.53e−14
MC1R	2.29e−01	3.58e−01
Sex	7.05e−01	2.76e−03
Skin phototype	7.34e−04	4.59e−03
Sun Damage	3.78e−01	1.40e−04
Sun History	4.76e−01	2.64e−03
Sunlight body site	2.58e−01	2.10e−05

ANOVA *P*-values: 1, 0.05, 0.01, 0.001, 0.0001, 0.00001

D