

Published in final edited form as:

Science. 2015 May 22; 348(6237): 880–886. doi:10.1126/science.aaa6806.

High burden and pervasive positive selection of somatic mutations in normal human skin

Iñigo Martincorena¹, Amit Roshan², Moritz Gerstung¹, Peter Ellis¹, Peter Van Loo^{1,3,4}, Stuart McLaren¹, David C. Wedge¹, Anthony Fullam¹, Ludmil B. Alexandrov¹, Jose M. Tubio, Lucy Stebbings¹, Andrew Menzies¹, Sara Widaa¹, Michael R. Stratton¹, Philip H. Jones^{2,*}, and Peter J. Campbell^{1,5,*}

¹Wellcome Trust Sanger Institute, Hinxton CB10 1SA, Cambridgeshire, UK

²MRC Cancer Unit, Hutchison-MRC Research Centre, University of Cambridge, Cambridge, UK

³Francis Crick Institute, London, UK

⁴Department of Human Genetics, University of Leuven, Leuven, Belgium

⁵Department of Haematology, University of Cambridge, Cambridge, UK

Abstract

How somatic mutations accumulate in normal cells is central to understanding cancer development, but is poorly understood. We performed ultra-deep sequencing of 74 cancer genes in small (0.8–4.7mm²) biopsies of normal skin. Across 234 biopsies of sun-exposed eyelid epidermis from four individuals, the burden of somatic mutations averaged 2–6 mutations/megabase/cell, similar to many cancers, and exhibited characteristic signatures of ultraviolet light exposure. Remarkably, multiple cancer genes are under strong positive selection even in physiologically normal skin, including most of the key drivers of cutaneous squamous cell carcinomas. Positively selected ‘driver’ mutations were found in 18–32% of normal skin cells at a density of ~140/cm². We observed variability in the driver landscape among individuals and variability in sizes of clonal expansions across genes. Thus, aged, sun-exposed skin is a patchwork of thousands of evolving clones, with over a quarter of cells carrying cancer-causing mutations while maintaining the physiological functions of epidermis.

The standard narrative of tumor evolution depicts accumulation of driver mutations in cancer genes, causing waves of expansion of progressively more disordered clones (1, 2). Central to this model is the presumption that randomly distributed somatic mutations must accumulate in normal cells before transformation (3), but directly observing them has proved challenging due to the polyclonal composition of normal tissue. Retrospective reconstructions of clonal evolution from sequencing of tumors give only partial insights, leaving us with fundamental gaps in our understanding of the earliest stages of cancer development. Critical, but largely unanswered, questions include the burden of somatic mutations in normal cells, which mutational processes are operative in normal tissues, the extent of positive selection among competing clones within an organ, and the patterns of

*Correspondence to: phj20@mrc-cu.cam.ac.uk; pc8@sanger.ac.uk.

clonal expansion induced by the very first driver mutations (4, 5). These questions have been partially addressed in blood cells, where somatic mutations, including some driver mutations, have been found to accumulate at a low rate with increasing age (6-10).

To study the burden, mutational processes and clonal architecture of somatic mutations in normal non-hematological tissue, we focused on sun-exposed skin. Previous studies have reported the existence of clonal patches of skin cells carrying *TP53* mutations (11-15). Motivated by this, we designed a sequencing strategy capable of detecting such clones by performing ultra-deep sequencing of small biopsies and adapting algorithms to detect mutations in a small fraction of cells. We used eyelid epidermis because of its relatively high levels of sun exposure and being one of the few body sites to have normal skin excised (blepharoplasty). This procedure is performed for age-related loss of elasticity of the underlying dermis, which can cause eyelid drooping sometimes severe enough to occlude vision, although the epidermis remains physiologically and histologically normal. From four individuals undergoing bilateral blepharoplasty, we obtained the resected eyelids, all of which had normal epidermis free of macroscopic lesions. The donors, three female and one male, ranged from 55 to 73 years of age and had variable history of sun exposure (table S1). Three were of Western European origin, and one was of South Asian origin. We separated the underlying dermis and took multiple biopsies of the epidermis from each eyelid (Fig. 1A,B). In total, 234 biopsies of 0.79-4.71mm² in area were analyzed. We sequenced the coding exons of 74 genes implicated in skin and other cancers to an average effective coverage of 500× (supplementary methods S1.2, Fig. S7). We also performed whole genome sequencing to ~147× depth on one biopsy in which a predominant clone was found by the targeted gene screen.

Mutational signature of ultraviolet light exposure in normal skin

To identify somatic mutations in the skin biopsies, we adapted an algorithm designed to detect subclonal variants in cancers (16) (supplementary methods S1.3, Figs. S3,S6), based on building a per-base model of background sequencing errors and identifying loci that had statistical excess of mismatched base calls (code released in the 'deepSNV' R package). This allowed us to detect mutations present in as few as 1% of the cells of a biopsy, detecting mutant clones ranging from 0.01mm² to several mm² in size. Overall, we identified 3,760 somatic mutations across the 234 biopsies (Fig. 1C; Dataset S1). Several lines of evidence confirm that the overwhelming majority of these variant calls are genuine somatic mutations (supplementary methods S1.3.2, Figs. S1,S2).

The pattern of mutations we identified closely matched that expected for ultraviolet (UV) light exposure, and that seen in skin cancers (Fig. 1D-F, Fig. S8). There was a predominance of C>T mutations, especially when the mutated cytosine was preceded by another pyrimidine (namely, TpC or CpC context), and there were high rates of CC>TT dinucleotide substitutions. This signature is consistent with the known chemistry of sunlight-induced damage to DNA, in which ultraviolet rays catalyze the formation of cyclobutane dimers from adjacent pyrimidines (17-20). C>T and CC>TT mutations were significantly more frequent on the non-transcribed strand of genes (Fig. 1D,E), consistent with transcription-coupled repair (21).

We also observed enrichment of C>A (G>T) mutations, with no obvious sequence context but a strong bias toward higher rates of C>A mutations on the transcribed strand (Fig. 1D). Assuming the strand bias results from transcription-coupled repair, this indicates that the damaged base is the guanine in the C:G pairing. This signature is also seen in cutaneous squamous cell carcinoma (cSCC) cancers, particularly in those with a relatively low mutation burden (Fig. S8), but less frequently in basal cell carcinomas (BCCs) and melanomas. A significant fraction of mutations seen after *in vitro* exposure of cells to UV rays are not the canonical transitions at dipyridime sites, with C>A transversions being prominent (20). One hypothesis for the pathogenesis of this signature is oxidation of guanine residues (typically 8-oxoGuanine) by reactive oxygen species generated by sunlight (22). Notably, 8-oxoGuanine is subject to transcription-coupled repair (23), consistent with the strand bias we see.

Pervasive positive selection of somatic mutations in normal skin

In the Darwinian model of cancer evolution, clones with driver mutations in cancer genes have a selective advantage over those without. In genomic data across multiple tumors, this manifests as an enrichment of protein-altering mutations in cancer genes compared to that expected for the background mutational rate. To explore whether clonal selection is operative in normal skin cells, we adapted a dN/dS model that accounts for the context-dependent mutation spectrum and that estimates the background mutation rate of each gene separately using synonymous mutations (24) (Fig. 2, supplementary methods S1.4, Fig. S6). One major advantage of this approach is that the mutation rate is estimated locally, thus inherently correcting for the variation in mutation rate across the genome, differences in read depth across the genes surveyed and the mutational spectrum observed in each individual. Genes under positive selection can be identified and the number of driver mutations can be quantified from the excess of nonsynonymous mutations (24).

Remarkably, six genes had a significant excess of protein-altering base substitutions after correcting for multiple hypotheses testing (Fig. 2), with five of these also showing excess rates of indels and/or dinucleotide subs (Fig. 2D, supplementary methods S1.4). *NOTCH1* was the most frequently mutated gene in the cohort and showed the highest observed/expected ratios of missense, nonsense and essential splice site mutations. *NOTCH2* and *NOTCH3* also carried significant excess of protein-altering mutations. NOTCH receptors are key regulators of stem cell biology in a number of organs (25), and are a frequent target of inactivating mutations in epithelial cancers (26-29) and activating mutations in lymphoid malignancies (30, 31). The distribution of somatic mutations within the *NOTCH1* and *NOTCH2* genes was not random, with striking clustering of amino acid substitutions in the extracellular EGF-like domains and large numbers of protein-truncating mutations distributed throughout the genes, matching that observed in cutaneous and head and neck SCCs (Fig. 2I). The density of positively selected driver mutations was surprisingly high. We estimated the excess of protein-altering mutations to be 57.1/cm² (CI_{95%}, 51-61/cm²) for *NOTCH1*, 24.6/cm² (19-28/cm²) for *NOTCH2* and 1.3/cm² (0.6-1.6/cm²) for *NOTCH3* (Fig. 2C, supplementary methods S1.4.2). Thus, on average there are 83 driver mutations in NOTCH genes for every square centimeter of aged, sun-exposed skin.

In SCCs of the skin and other organs, both copies of *NOTCH1* are frequently inactivated (28, 29), typically through point mutation combined with copy number alteration. We developed an algorithm to identify small populations of cells with copy number alterations across the genes targeted for sequencing by phasing heterozygous SNPs (32) (supplementary methods S1.6, Fig. S4). *NOTCH1* was the gene most frequently subject to copy number changes (Fig. 3), with 27/234 biopsies having detectable alterations (Fig. 3B). Only occasional copy number alterations were detected in other genes, although our power to detect these was variable due to differences in numbers of heterozygous SNPs sequenced. Strikingly, when we estimate the percentage of cells carrying a *NOTCH1* copy number change in a biopsy, we find that there is often a *NOTCH1* point mutation apparently occurring in the same fraction of cells in the biopsy (Fig. 3C). This overlap, which occurs much more frequently than expected by chance ($P < 10^{-5}$, supplementary methods S1.6.1), demonstrates that biallelic inactivation of *NOTCH1* is already frequent in normal skin cells, and not restricted to SCCs.

FAT1 showed a statistically significant excess of inactivating mutations across all classes, including nonsense and essential splice site substitutions and short indels ($q = 8 \times 10^{-11}$, 9×10^{-6} and 2×10^{-4} , respectively; Fig. 2B-D; supplementary methods S1.4.5). *FAT1* is a cadherin-like protein that suppresses tumor growth by blocking β -catenin signaling and is recurrently mutated in a range of cancers (33), including cutaneous (table S2) and head and neck squamous cell carcinomas (34, 35). Consistent with previous analyses of mutant clones in normal skin (11), *TP53* carried an estimated 9.5 driver mutations/cm² (4.6-11.8/cm²; $q = 4 \times 10^{-6}$). In addition, we saw canonical hotspot mutations in several oncogenes, including *KRAS*, *NRAS* and *HRAS*.

Interestingly, we found evidence of positive selection in other genes that have not previously been implicated in skin cancer. *RBM10* is an RNA-binding protein that is subject to recurrent inactivating mutations in lung adenocarcinoma (36) – we also see excess of protein-truncating mutations in normal skin ($q = 0.009$; Fig. 2B). *RBM10* is not a known skin cancer gene, although it may conceivably emerge as a rare driver in cSCCs with further sequencing. Additionally, in an analysis for excess mutations at hotspots, *FGFR3* showed significant recurrence at two canonical residues (supplementary methods S1.4.3). The same hotspot mutations have been found in ~40% of seborrheic keratoses (37). These skin growths have an incidence 15-fold higher than skin cancers (38), but never become invasive or malignant. This observation suggests that there may be a class of genes in which somatic mutations give a clonal selective advantage in normal tissue, but do not cause, or could even inhibit, hallmarks of the cancer phenotype such as invasion or dissemination.

We compared the catalogue of significantly mutated genes in normal skin to published exome sequencing studies from the three commonest classes of skin cancer, namely cSCCs (28, 39, 40), BCC (41) and melanoma (42). When analyzed using the same statistical methodology, there was little overlap in positively selected genes in normal skin compared to either BCC or melanoma (Fig. 2G,H; supplementary results S2.2; tables S2-S4). In contrast, we found that the pattern in normal skin closely matched that of cSCC, with *NOTCH1*, *NOTCH2*, *FAT1* and *TP53* all being significantly mutated in the latter (Fig. 2F). Point mutations in *CDKN2A* were not found to be under positive selection in normal skin,

despite this gene being a frequent driver in cSCC, inactivated by point mutations or homozygous deletions. Although our design does not allow us to reliably detect homozygous deletions, we found only three *CDKN2A* point mutations (2 missense and 1 synonymous) across all 234 samples of normal skin, whereas ~31% (CI_{95%}: 14-52%) of cSCCs carry non-synonymous point mutations in the gene. These data suggest that the selective forces acting on physiologically normal skin resemble those in squamous cell carcinomas, with remarkable similarities between the driver mutations in each. However, *CDKN2A* inactivation appears to be specific to cancer clones, suggesting that its loss confers no selective advantage until more advanced stages of cancer evolution. The absence of mutations characteristic of melanomas is consistent with the fact that around 95% of the cells in the epidermis are keratinocytes (43), while melanomas originate from melanocytes. The absence of the *PTCH1* mutations seen in BCC is notable, especially given that BCC has a threefold higher incidence than cSCC in populations of European ancestry (44). This may be consistent with BCC originating from cells infrequent or absent in the eyelid epidermis, such as from hair follicles (45), although our data cannot rule out other explanations.

Surprisingly, one of the four individuals in our series contributed a disproportionate number of mutations in *NOTCH2* (39% of all mutations in *NOTCH2* compared to 24% in other genes). A formal test of heterogeneity confirmed that *NOTCH2* showed a variable rate of driver mutations among individuals ($q=0.0005$; Fig. 2J; supplementary methods S1.5; Figs. S9-S10). Since the dN/dS method used inherently accounts for gene-specific coverage and patient-specific mutation spectrum, this finding is likely to reflect a true biological difference among the four individuals rather than a bias arising from some aspect of the experimental design. One conceivable explanation is that some difference in the local eyelid environment provides a stronger pressure for *NOTCH2* mutations; another, more likely explanation is that the genetic background of each individual could lead to differences in the strength of selective advantage across genes. The patient with different selection strength for *NOTCH2* was of South Asian ancestry, while the other three were Western Europeans, although this needs considerably larger sample sizes to address formally. Nonetheless, these data illustrate the exciting potential of such study designs to detect inter-individual differences in the driver landscape that cannot be extracted from sequencing a single established cancer per patient.

Mutant clonal expansions

Together with the mutation rate, the size of the clonal expansions induced by driver mutations in normal tissue is critical to understanding the evolution of cancer, since both factors together determine the size of the pool of cells that can acquire sequential hits (4, 5). With our experimental design, the observed fraction of sequencing reads reporting a mutation correlates accurately with the fraction of cells in a biopsy that carry the mutation, once we correct for the local copy number at that locus, enabling us to estimate clone sizes (32) (supplementary sections S1.7, S2.6.1). For the majority of mutations identified here, the variant allele fraction was <5% (Fig. 1B), indicating that most mutations were seen in only a small proportion of cells in the biopsy, typically <10%, with many mutations seen in only 1-2% of cells. There were exceptions, however, and some biopsies carried somatic mutations found in most of the cells. We find that the distribution of mutant clone sizes in

aged, sun-exposed skin has a heavy right tail (Fig. 4A; Fig. S11A), with some clones as large as several mm² in surface area.

To estimate the average burden of somatic substitutions per skin cell, we can integrate the estimated fraction of mutant cells across the biopsies from each of the four subjects (supplementary methods S1.8). This reveals that the mutation burden estimated from coding sequence is at least 2-6 somatic mutations/Mb/cell (Fig. 4B). This estimate is at the lower end of the burden of mutations in cSCCs (1-380/Mb) and melanomas (0.5-200/Mb), and higher than the average mutation burdens seen in many adult solid tumors (Fig. 4B). Using the variant allele fraction, we estimate that 14%-21% of skin cells carry *NOTCH1* mutations, with 5-7% having *NOTCH2* and 2-3% *NOTCH3* mutations (Fig. 2E). *TP53* mutations and *FAT1* mutations are present in 3-5% of skin cells, remarkably similar to the estimate of 4% from immunohistochemical studies of *TP53* clones in human skin (11). Thus, about a quarter of all skin cells in these biopsies carried NOTCH mutations, the vast majority of which are driver mutations.

In current models of cancer development, driver mutations cause clonal expansions, widening the pool of cells that is susceptible to further driver mutations until enough accumulate to drive transformation and invasion. We compared the clone size of mutations in driver genes against that of synonymous mutations in non-driver genes, which are likely selectively neutral (Fig. 4C). We find that whereas the average clone size for neutral mutations was 0.15mm² (CI_{95%}:0.13-0.17), it was significantly larger for driver mutations in *NOTCH1* (average 0.23mm²; q=0.002), *TP53* (0.33mm²; q=0.009) and *FGFR3* (0.69mm²; q=0.0007; permutation test). Clone sizes for *FAT1*, *NOTCH2* and *NOTCH3* mutations were not significantly increased. Although some putatively neutral mutations in this dataset may be hitchhiking in clones with driver mutations, the difference in clone sizes between driver and neutral mutations is unexpectedly small. The large excess of truncating mutations in these genes demonstrates that clones carrying these mutations must have had a strong selective advantage at some stage. Indeed, lineage tracing in mice has revealed that clones carrying *TP53* mutations grow nearly exponentially in UV-exposed epidermis (13). Yet, exponential growth must slow relatively early in the expansion of the clones to explain both the limited range of clone sizes observed here (Fig. 4C) and their similarity across individuals of different ages (Fig. 1C). Such constraints on clonal growth are likely to represent a critical protection against progressive accumulation of driver mutations and cancer. The physiological mechanisms underpinning this are unknown, but ‘imprisonment’ of *Tp53*-mutant clones has been observed in murine epidermis (46), possibly driven by interactions between the clone and surrounding cells and density-dependent growth constraints.

In contrast to the relatively small clone sizes of canonical cSCC driver genes, clones with activating *FGFR3* mutations were among the largest observed. It is striking that the driver mutations inducing the largest clonal expansions in normal skin were those associated with benign tumors, namely seborrheic keratosis. This shows that the size of clonal expansion induced by a somatic mutation need not correlate with its potential to induce malignant transformation.

Our data reveal notable similarities between normal and cancer cells, with normal cells carrying thousands of mutations, including oncogenic driver mutations subject to strong positive selection. A major difference between the normal cells sequenced here and cancer cells seems to be the number of driver mutations per cell (Fig. 4D). Using dN/dS, we estimate that normal cells in the skin of these four subjects carry an average of 0.27 (CI_{95%}: 0.19-0.35) driver point mutations per cell. Using the same method for cSCCs, we estimate an average of 2.7 (CI_{95%}:0.91-4.65) driver point mutations per tumor in the genes sequenced in this study.

At an average of 0.27 driver mutations/cell, there may be many normal cells with several drivers coexisting. When clones represent a large enough fraction of the biopsy, we can apply deductive reasoning to demonstrate co-occurrence of mutations in the same clone of cells (32). In our data, there were six large clones for which this was possible (Fig. 4E), with three showing two or more likely driver mutations in the same subclone. In one massive clone that spanned six adjacent biopsies, we found all cells carrying a canonical activating mutation in *FGFR3* together with a known driver mutation in *TP53*, and two separate subclonal expansions (Fig. 4F).

To obtain a more comprehensive picture of the mutational landscape of normal cells, we performed whole genome sequencing to 147× depth on a biopsy containing this clone. This identified 73,904 base substitutions and 2,248 small indels, with a mutation signature largely dominated by UV light exposure (Fig. S12B,C). About 14,000 of these were clonal (~4.6/Mb), presumably hitchhiking with the *FGFR3* and *TP53* mutations, but the rest were subclonal, often in <20% cells (Fig. S12A). Integrating the allele frequencies, we estimate an average of 21,102 mutations per genome per cell (~7/Mb) in this sample. The mutation rate was found to vary along the genome, with higher rates in lowly expressed genes and in repressed chromatin (Fig. S13), as observed in cancer (47) and human evolution (48).

Discussion

We found the frequency of driver mutations in physiologically normal skin cells surprisingly high. For example, there were more *NOTCH1* mutations in just 5cm² of aged, sun-exposed skin analyzed here than have been identified in more than 5,000 cancers sequenced by TCGA (The Cancer Genome Atlas). About 20% of normal skin cells carry driver mutations in *NOTCH1*, with some but not overwhelming enrichment in the matching cancer (60% of cSCCs have *NOTCH1* mutations). Several other cancer genes were under positive selection in normal skin, and we found clones carrying 2-3 driver mutations that had not acquired malignant potential, raising the question of what combinations of events are sufficient for transformation. These observations may not be entirely unexpected – for cancers to occur with the frequency they do in the general population, there may be a vast underlying reservoir of competing clones part or much of the way to malignant transformation. A rather sobering corollary is that if we had a systemic targeted therapeutic that killed all cells with inactivated *NOTCH1*, we might successfully treat 60% of cSCCs but with considerable collateral damage to physiologically normal skin.

Studying tumor evolution by sequencing established cancers is akin to inferring the rules of a musical talent quest by identifying similarities across the show's annual winners. Successful aspirants undoubtedly have common properties that identify necessary criteria for victory, but there is no substitute for directly observing the competition in its raw, early, local heats. Here, we have found hundreds of evolving clones per square centimeter of skin (Fig. 4G); thousands of mutations per skin cell; variability among individuals in profiles of driver mutations; and variability among cancer genes in clonal dynamics. Scaled up across the range of organ systems, cell types and mutational exposures – and encompassing ageing, predisposing diseases and genetic backgrounds – such studies promise to reveal fundamental insights into the earliest stages of cancer development.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank M.S. Kolodney, R.J. Cho and M. Dimon for kindly providing their published exome sequencing data on BCC and cSCC. This work was supported by the Wellcome Trust (077012/Z/05/Z) and by an MRC Centennial award (PHJ). PJC has a Wellcome Trust Senior Clinical Research Fellowship (WT088340MA). PHJ is supported by an MRC grant in aid and Cancer Research UK (Program grant C609/A17257). IM is supported by fellowships from EMBO (1287-2012) and Queens' College, Cambridge. AR acknowledges support from the CRUK-Cambridge Cancer Centre Clinical Research Fellowship. The sequencing data have been deposited in EGA (EGAS00001000860, EGAS00001000515, EGAS00001000603). The R code used for variant calling (*ShearwaterML*) is available in Bioconductor (*deepSNV* package).

References and Notes

1. Vogelstein B, et al. Cancer genome landscapes. *Science*. Mar 29.2013 339:1546. [PubMed: 23539594]
2. Garraway LA, Lander ES. Lessons from the cancer genome. *Cell*. Mar 28.2013 153:17. [PubMed: 23540688]
3. Tomasetti C, Vogelstein B, Parmigiani G. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proceedings of the National Academy of Sciences of the United States of America*. Feb 5.2013 110:1999. [PubMed: 23345422]
4. Nowell PC. The clonal evolution of tumor cell populations. *Science*. Oct 1.1976 194:23. [PubMed: 959840]
5. Jackson AL, Loeb LA. The mutation rate and cancer. *Genetics*. Apr.1998 148:1483. [PubMed: 9560368]
6. Welch JS, et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell*. Jul 20.2012 150:264. [PubMed: 22817890]
7. Jaiswal S, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *The New England journal of medicine*. Dec 25.2014 371:2488. [PubMed: 25426837]
8. Genovese G, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *The New England journal of medicine*. Dec 25.2014 371:2477. [PubMed: 25426838]
9. Xie M, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nature medicine*. Dec.2014 20:1472.
10. McKerrell T, et al. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell reports*. Mar 3.2015 10:1239. [PubMed: 25732814]
11. Jonason AS, et al. Frequent clones of p53-mutated keratinocytes in normal human skin. *Proceedings of the National Academy of Sciences of the United States of America*. Nov 26.1996 93:14025. [PubMed: 8943054]

12. Ziegler A, et al. Sunburn and p53 in the onset of skin cancer. *Nature*. Dec 22-29.1994 372:773. [PubMed: 7997263]
13. Klein AM, Brash DE, Jones PH, Simons BD. Stochastic fate of p53-mutant epidermal progenitor cells is tilted toward proliferation by UV B during preneoplasia. *Proceedings of the National Academy of Sciences of the United States of America*. Jan 5.2010 107:270. [PubMed: 20018764]
14. Nakazawa H, et al. UV and skin cancer: specific p53 gene mutation in normal skin as a biologically relevant exposure measurement. *Proceedings of the National Academy of Sciences of the United States of America*. Jan 4.1994 91:360. [PubMed: 8278394]
15. Ling G, et al. Persistent p53 mutations in single cells from normal human skin. *The American journal of pathology*. Oct.2001 159:1247. [PubMed: 11583952]
16. Gerstung M, Papaemmanuil E, Campbell PJ. Subclonal variant calling with multiple samples and prior knowledge. *Bioinformatics*. May 1.2014 30:1198. [PubMed: 24443148]
17. Alexandrov LB, et al. Signatures of mutational processes in human cancer. *Nature*. Aug 22.2013 500:415. [PubMed: 23945592]
18. Pfeifer GP, You YH, Besaratinia A. Mutations induced by ultraviolet light. *Mutation research*. Apr 1.2005 571:19. [PubMed: 15748635]
19. Pfeifer GP, Besaratinia A. UV wavelength-dependent DNA damage and human non-melanoma and melanoma skin cancer. *Photochem Photobiol Sci*. Jan.2012 11:90. [PubMed: 21804977]
20. Brash DE. UV signature mutations. *Photochemistry and photobiology*. Jan-Feb;2015 91:15. [PubMed: 25354245]
21. Shuck SC, Short EA, Turchi JJ. Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. *Cell research*. Jan.2008 18:64. [PubMed: 18166981]
22. Besaratinia A, Synold TW, Xi B, Pfeifer GP. G-to-T transversions and small tandem base deletions are the hallmark of mutations induced by ultraviolet a radiation in mammalian cells. *Biochemistry*. Jun 29.2004 43:8169. [PubMed: 15209513]
23. Guo J, Hanawalt PC, Spivak G. Comet-FISH with strand-specific probes reveals transcription-coupled repair of 8-oxoGuanine in human cells. *Nucleic acids research*. Sep.2013 41:7700. [PubMed: 23775797]
24. Greenman C, Wooster R, Futreal PA, Stratton MR, Easton DF. Statistical analysis of pathogenicity of somatic mutations in cancer. *Genetics*. Aug.2006 173:2187. [PubMed: 16783027]
25. Liu J, Sato C, Cerletti M, Wagers A. Notch signaling in the regulation of stem cell self-renewal and differentiation. *Current topics in developmental biology*. 2010; 92:367. [PubMed: 20816402]
26. Wang NJ, et al. Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*. Oct 25.2011 108:17761. [PubMed: 22006338]
27. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. Sep 27.2012 489:519. [PubMed: 22960745]
28. Durinck S, et al. Temporal dissection of tumorigenesis in primary cancers. *Cancer discovery*. Jul. 2011 1:137. [PubMed: 21984974]
29. Agrawal N, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science*. Aug 26.2011 333:1154. [PubMed: 21798897]
30. Puente XS, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. Jul 7.2011 475:101. [PubMed: 21642962]
31. Weng AP, et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science*. Oct 8.2004 306:269. [PubMed: 15472075]
32. Nik-Zainal S, et al. The life history of 21 breast cancers. *Cell*. May 25.2012 149:994. [PubMed: 22608083]
33. Morris LG, et al. Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation. *Nature genetics*. Mar.2013 45:253. [PubMed: 23354438]
34. Stransky N, et al. The mutational landscape of head and neck squamous cell carcinoma. *Science*. Aug 26.2011 333:1157. [PubMed: 21798893]

35. India Project Team of the International Cancer Genome Consortium. Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. *Nature communications*. 2013; 4:2873.
36. Imielinski M, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*. Sep 14.2012 150:1107. [PubMed: 22980975]
37. Logie A, et al. Activating mutations of the tyrosine kinase receptor FGFR3 are associated with benign skin tumors in mice and humans. *Human molecular genetics*. May 1.2005 14:1153. [PubMed: 15772091]
38. Harvey I, Frankel S, Marks R, Shalom D, Nolan-Farrell M. Non-melanoma skin cancer and solar keratoses. I. Methods and descriptive results of the South Wales Skin Cancer Study. *British journal of cancer*. Oct.1996 74:1302. [PubMed: 8883422]
39. South AP, et al. NOTCH1 mutations occur early during cutaneous squamous cell carcinogenesis. *The Journal of investigative dermatology*. Oct.2014 134:2630. [PubMed: 24662767]
40. Pickering CR, et al. Mutational landscape of aggressive cutaneous squamous cell carcinoma. *Clin Cancer Res*. Dec 15.2014 20:6582. [PubMed: 25303977]
41. Jayaraman SS, Rayhan DJ, Hazany S, Kolodney MS. Mutational landscape of basal cell carcinomas by whole-exome sequencing. *The Journal of investigative dermatology*. Jan.2014 134:213. [PubMed: 23774526]
42. Hodis E, et al. A landscape of driver mutations in melanoma. *Cell*. Jul 20.2012 150:251. [PubMed: 22817889]
43. McGrath, JA.; Eady, RA.; Pope, FM. *Rook's Textbook of Dermatology*. 7th ed.. Blackwell Publishing; 2004.
44. Lomas A, Leonardi-Bee J, Bath-Hextall F. A systematic review of worldwide incidence of nonmelanoma skin cancer. *The British journal of dermatology*. May.2012 166:1069. [PubMed: 22251204]
45. Grachtchouk M, et al. Basal cell carcinomas in mice arise from hair follicle stem cells and multiple epithelial progenitor populations. *The Journal of clinical investigation*. May.2011 121:1768. [PubMed: 21519145]
46. Zhang W, Remenyik E, Zelteman D, Brash DE, Wikonkal NM. Escaping the stem cell compartment: sustained UVB exposure allows p53-mutant keratinocytes to colonize adjacent epidermal proliferating units without incurring additional mutations. *Proceedings of the National Academy of Sciences of the United States of America*. Nov 20.2001 98:13948. [PubMed: 11707578]
47. Schuster-Bockler B, Lehner B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature*. Aug 23.2012 488:504. [PubMed: 22820252]
48. Martincorena I, Luscombe NM. Non-random mutation: the evolution of targeted hypermutation and hypomutation. *BioEssays*. Feb.2013 35:123. [PubMed: 23281172]

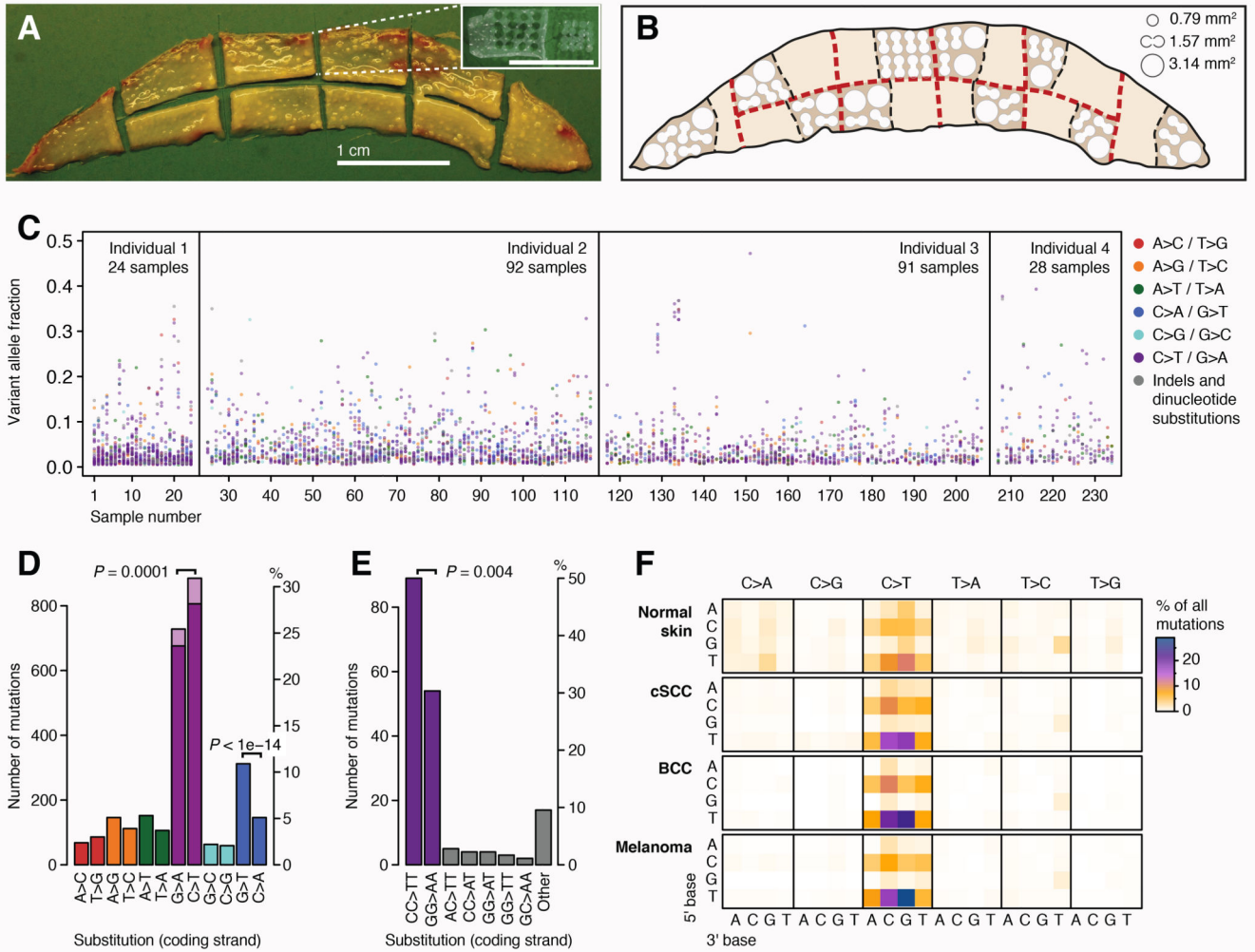


Fig. 1. Burden and spectrum of mutations in normal human skin

(A) Excised human eyelid viewed from the dermal surface. The inset shows a sample region of epidermis after the dermis has been removed and biopsies taken. (B) Locational map of harvested areas from an eyelid showing locations of 0.79mm², 1.57mm² and 3.14mm² biopsies. (C) Distribution of the variant allele fraction (*i.e.* the fraction of sequencing reads reporting the mutation of all reads across the locus) for the 3,760 mutations found across the 234 samples from 4 individuals, colored by mutation type. (D, E) Total counts in the coding (untranscribed) versus the non-coding (transcribed) strand for single base substitutions (D) and dinucleotides (E). The counts of C>T (G>A) mutations in a dipyrimidine context are shown in dark purple. *P*-values reflect the transcription strand asymmetry (Exact Poisson test). (F) Heat map of the relative rates of each mutation type depending on the nucleotides upstream and downstream of the mutated base. Rates are normalized for sequence composition of the targeted genes.

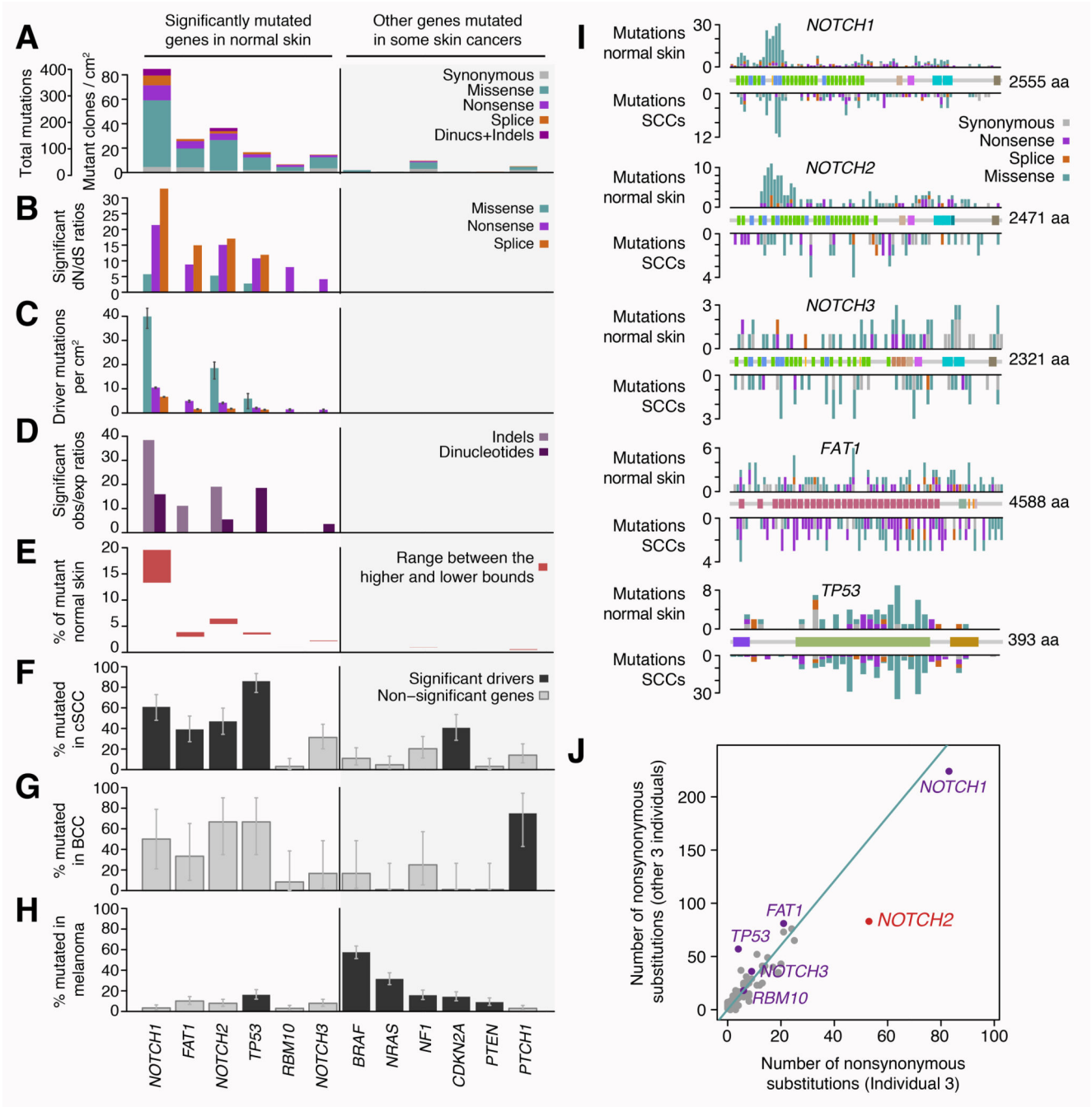


Fig. 2. Pervasive positive selection of oncogenic mutations in normal skin

(A-E) Patterns of selection in six genes recurrently mutated in normal skin and in six other genes frequently implicated in skin cancers. (A) Number of mutations per gene classified by their functional impact. (B) dN/dS ratios for genes under significant positive selection (only statistically significant ratios are shown). (C) Estimated number of driver mutations per cm² of normal skin. (D) Enrichment of indels and dinucleotides in driver genes (bars show significant observed/expected ratios only). (E) Estimated percentage of cells in normal skin carrying mutations in each gene. Lower bound estimates were obtained assuming the

possibility of up to two driver mutations per cell, while higher bound estimates are obtained by allowing only one driver mutation per gene per cell. **(F-H)** Percentage of cSCC, BCC and melanoma tumors that carry a non-synonymous point mutations in each gene. Genes found to be recurrently mutated in each cancer type are shown in black (supplementary results S2.2). **(I)** Distribution of mutations across five driver genes in normal skin (above the gene diagrams) and in SCCs (below), including 67 cutaneous SCCs and 319 TCGA head and neck cancer exomes. The gene diagrams show the location of encoded protein domains. **(J)** Differential selection in *NOTCH2* across individuals (supplementary methods S1.5).

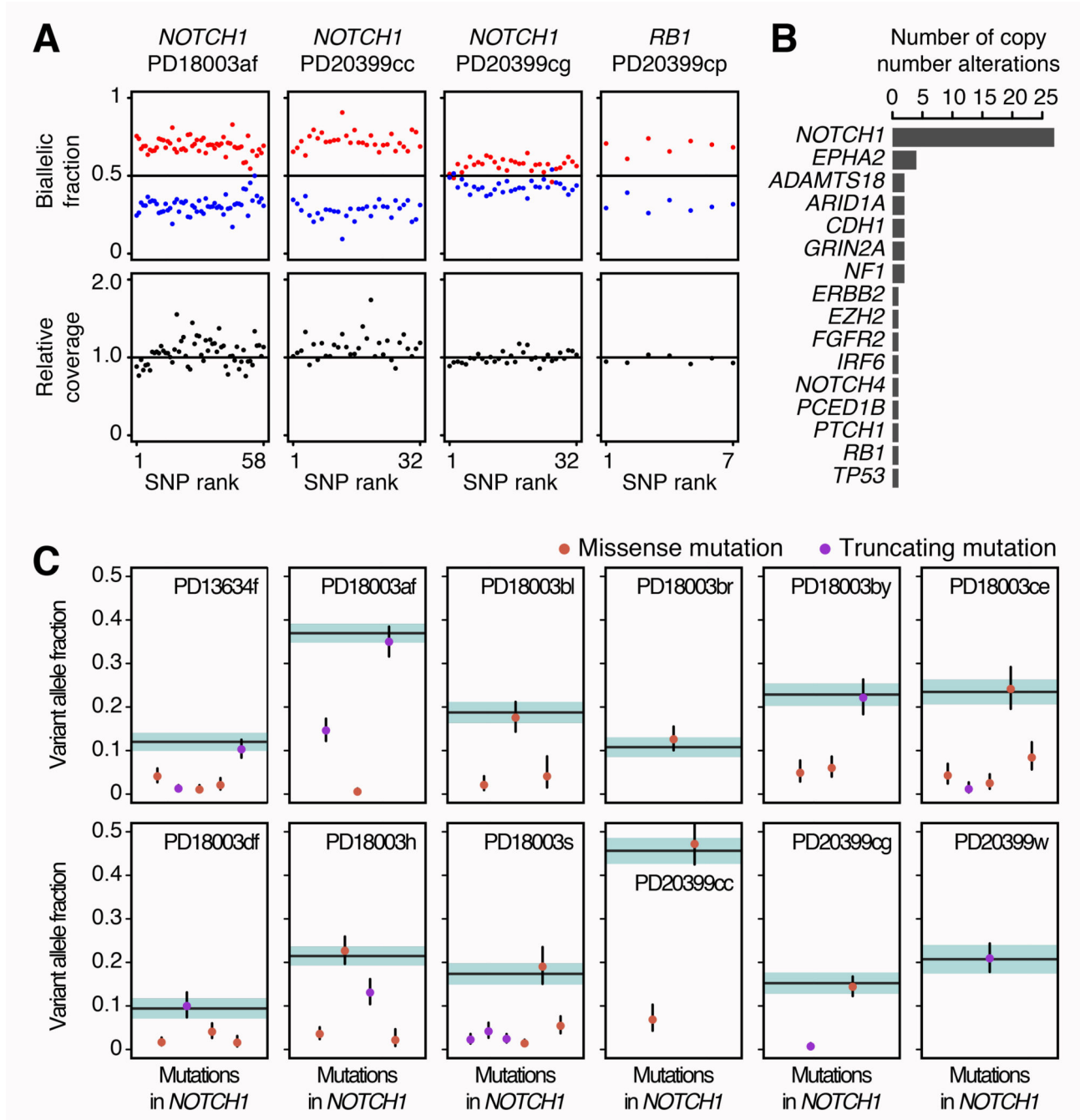


Fig. 3. Frequent copy number aberrations and biallelic loss of *NOTCH1* in normal skin

(A) Example of four skin samples with subclonal copy number aberrations in *NOTCH1* and *RB1*. Every point represents a heterozygous SNP within the affected gene and aberrations manifest as allelic imbalances, with a higher fraction of reads (biallelic fraction) supporting one of the alleles of the gene (in red). The extent of the deviation from 0.5 depends on the number of gene copies gained or lost and on the proportion of the biopsy occupied by the subclone (supplementary methods S1.6). (B) Number of copy number aberrations detected per gene. (C) In *NOTCH1*, a substitution is often found in the same fraction of cells as a

deletion of the other allele (dot colocalizing with a horizontal band), showing that the loss of both copies of *NOTCH1* is frequent in normal skin cells. Horizontal lines represent the expected variant allele fraction for a mutation inactivating the only remaining allele of a gene in the same clone, with colored shadows representing 95% confidence intervals. Orange and purple dots represent the allele fraction of missense and nonsense mutations in the biopsy, with 95% confidence intervals (supplementary methods S1.6.1, Fig. S5).

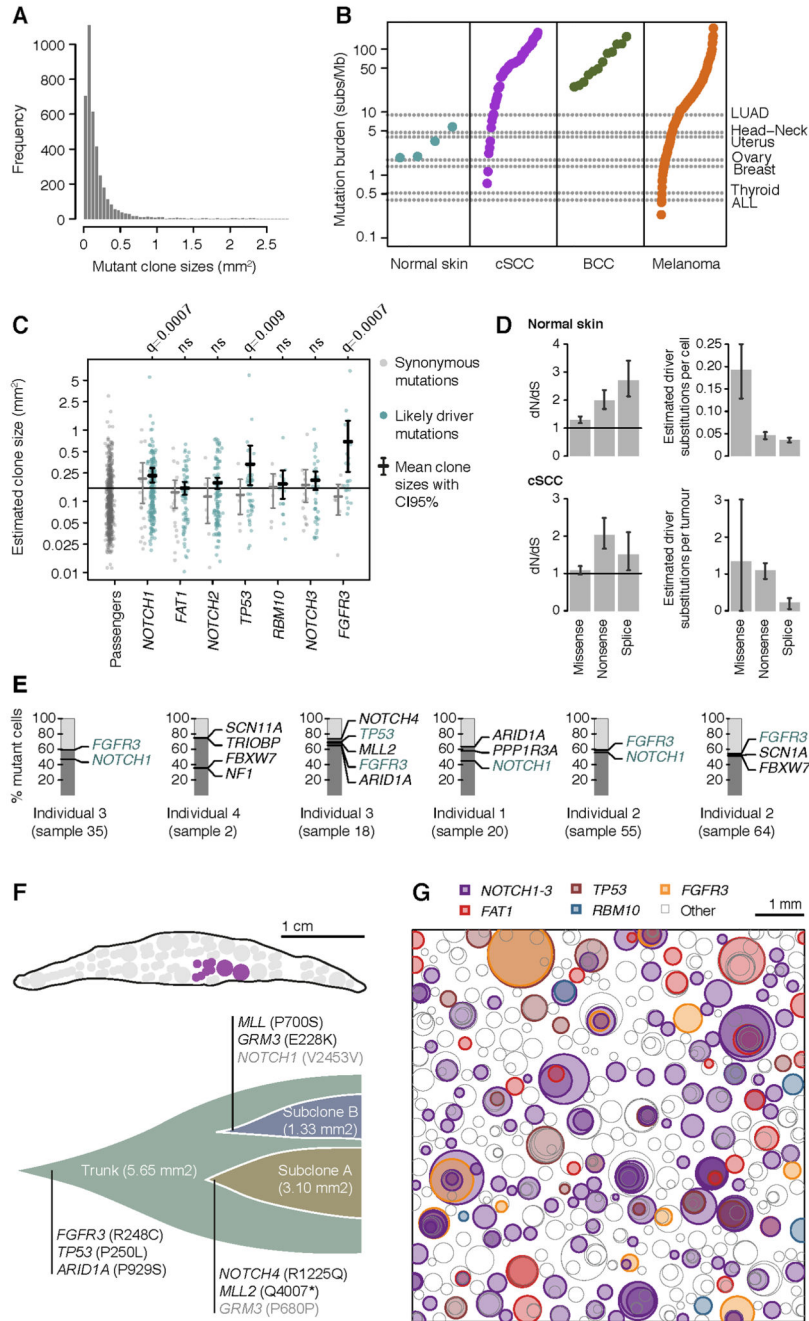


Fig. 4. Mutant clone sizes and clonal dynamics in normal skin

(A) Distribution of clone sizes of all mutations. (B) Mutation burden per megabase in the normal skin of four individuals and in a range of human cancers (supplementary methods S1.8). (C) Clone sizes of likely driver and passenger mutations in normal skin. Driver mutations are defined as those mutation types found to be under significant positive selection in each gene (Fig. 2). Confidence intervals and FDR-corrected q-values were obtained using 10,000 random permutations of the gene labels assigned to each mutation. (D) Global dN/dS estimates across all 74 genes analyzed in the study in normal skin and

cSCC. This allows us to estimate the number of driver mutations per normal cell or per tumor as the number of mutations fixed by positive selection (supplementary methods S1.4.2). **(E)** Identification of mutations co-occurring in the same sub-clone using the pigeonhole principle (32). **(F)** Subclonal structure of a large clone found to overlap with six biopsies (shown in purple in the eyelid locational map). **(G)** Schematic representation of the mutant clones in an average 1 cm² of normal eyelid skin. To generate the figure, a number of biopsies were randomly selected to amount to 1cm² of sequenced skin and all clones observed in these biopsies were represented as circles randomly distributed in space. The density, size and the simulated nesting of clones are all based on the sequencing data obtained in this study.