Genetic evolution of keratinocytes to cutaneous squamous cell carcinoma

Bishal Tandukar[1 2 3 12] Delahny Deivendran[1 2 3 12], Limin Chen[1 2], Noel Cruz-Pacheco[1 2 3], Harsh Sharma[1 2 3], Albert Xu[1 2], Aravind K. Bandari[1 2], Daniel B. Chen[1 2 4], Christopher George[5 6 7], Annika Marty[1 2 8], Raymond J. Cho[1], Jeffrey Cheng[1], Drew Saylor[1], Pedram Gerami[9], Sarah T. Arron[10], Boris C. Bastian[1 2 3 11], A. Hunter Shain[1 2 3 13]

[1] Department of Dermatology, University of California San Francisco, San Francisco, CA, USA

[2] Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, USA

[3] Human Tumor Atlas Network (HTAN), National Cancer Institute, Bethesda, MD, USA

[4] School of Medicine, Case Western Reserve University, Cleveland, OH, USA

[5] Department of Dermatology, Erasmus MC, Rotterdam, Netherlands

[6] Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[7] Department of Dermatology, School of Medicine, Stanford University, Palo Alto, CA, USA

[8] Institute of Physiology, University of Zurich, Zurich, Switzerland

[9] Department of Dermatology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

[10] Peninsula Dermatology, Burlingame, CA, USA

[11] Department of Pathology, University of California San Francisco, San Francisco, CA, USA

[12] These authors contributed equally to this work

[13] Corresponding author. Email: hunter.shain@ucsf.edu

Authors declare that they have no competing interests. AHS has an industry sponsored research agreement with Kenvue that does not pertain to the work described here.

Abstract

We performed multi-omic profiling of epidermal keratinocytes, precancerous actinic keratoses, and squamous cell carcinomas to understand the molecular transitions during skin carcinogenesis. Single-cell mutational analyses of normal skin cells showed that most keratinocytes have remarkably low mutation burdens, despite decades of sun exposure, however keratinocytes with *TP53* or *NOTCH1* mutations had substantially higher mutation burdens. These observations suggest that wild-type keratinocytes (i.e. without pathogenic mutations) are able to withstand high dosages of cumulative UV radiation, but certain pathogenic mutations break these adaptive mechanisms, priming keratinocytes for transformation by increasing their mutation rate. Mutational profiling of squamous cell carcinomas adjacent to actinic keratoses revealed *TERT* promoter and *CDKN2A* mutations emerging in actinic keratoses, whereas additional mutations inactivating *ARID2* and activating the MAPK-pathway delineated the transition to squamous cell carcinomas. Surprisingly, actinic keratoses were often not related to their neighboring squamous cell carcinoma, indicating that collisions of unrelated neoplasms are common in the skin. Spatial variation in gene expression patterns was common in both tumor and immune cells, with high expression of checkpoint molecules at the invasive front of tumors. In conclusion, this study catalogues the key events during the evolution of cutaneous squamous cell carcinoma.

Introduction

Cutaneous squamous cell carcinoma is the second most common type of cancer(1) and is responsible for an estimated 2,500-15,000 deaths per year in the United States(2–4). These estimates vary widely because there are no cancer registries to officially track the mortality of cutaneous squamous cell carcinoma, but this range is on par with melanoma, gastric cancer, cervical cancer, liver cancer, and kidney cancer(5). Compared to other cancer subtypes with similar death tolls, the evolution of cutaneous squamous cell carcinoma remains poorly understood, posing a major obstacle towards improvement of prevention strategies and development of new therapeutic modalities.

Fully evolved cutaneous squamous cell carcinomas have somatic alterations disrupting the p53 and Notch signaling pathways(6). To a lesser extent, they also have alterations known to activate the MAPK/PI3-Kinase pathways, upregulate telomerase, perturb the SWI/SNF chromatin remodeling complex, abrogate cell-cycle checkpoint control, and/or activate the Hippo signaling pathway(6). The order in which these somatic alterations become selected during tumor evolution is not entirely known, but sequencing of normal skin and precursor lesions, such as actinic keratoses, provide some insights.

Cutaneous squamous cell carcinoma arises from keratinocytes of the epidermis. Clonal patches of keratinocytes with p53 or Notch mutations can be found in the epidermis, increasing in density with higher levels of cumulative sun exposure(7–15). Actinic keratoses are low-risk precursor lesions of cutaneous squamous cell carcinoma and probably arise from these patches. While several studies have sequenced actinic keratoses(16–18), their genetic drivers remain incompletely understood due to their complex clonal architecture, as we discuss in more detail below.

To better understand the genetic evolution of cutaneous squamous cell carcinoma, we profiled the mutational and transcriptional landscapes of individual keratinocytes from physiologically normal human skin. We also performed DNA sequencing and spatial transcriptomics of actinic keratoses that were adjacent to cutaneous squamous cell carcinomas. Our data reveal the key events driving the transformation of cutaneous squamous cell carcinomas from epidermal keratinocytes through the pre-neoplastic and pre-malignant stages of progression.

Results

Mutational landscapes of individual keratinocytes from normal human skin

We began by profiling the mutational landscapes of epidermal keratinocytes and comparing them to epidermal melanocytes and dermal fibroblasts from the same biopsies. It remains difficult to detect somatic mutations in an individual cell with high specificity and sensitivity(19–22). To achieve this goal, we adapted a workflow, previously designed to genotype melanocytes at single-cell resolution(23), to also work on keratinocytes and fibroblasts (see Fig. S1A for an overview). Briefly, we clonally expanded individual skin cells *ex vivo*, producing small colonies of daughter cells (typically 200 cells per colony), extracted their DNA and RNA, and further amplified the nucleic acids *in vitro*. The combination of clonal expansion of cells and *in vitro* amplification of DNA/RNA produced sufficient template material to sensitively detect mutations. To call somatic mutations at high specificity, we identified patterns in the sequencing data, as previously described (23), which can distinguish *bona fide* mutations from artifacts introduced during amplification.

A limitation to clonal expansion is that it may introduce a bias in the cells that grow out and are mutationally profiled. Such a critique would apply to other studies that measure the mutational landscapes of individual cells via clonal expansion(24–27), or more broadly, could be applied to any study that has ever performed tissue culture. It is a tradeoff that we believe is justified by the high-quality mutational calls, at single-cell resolution, that can be achieved with the help of clonal expansion. To minimize biases, we optimized culture conditions so that a high proportion of cells clonally expanded from each tissue (see methods). For example, we sorted cells with limited dilution in lieu of fluorescence activated cell sorting (FACS), despite considerable effort required. Nevertheless, bias may persist, so we compared the mutational landscapes of individual cells to bulk sequencing measurements and discuss, below, how each approach may skew mutational calls.

In total, we measured somatic mutations in single-cell expansions of 137 keratinocytes, 131 melanocytes, and 23 fibroblasts from 22 different skin biopsies from 15 unique donors (table S1, Fig. S1B). Donors ranged from 35-95 years of age, and skin was collected from body sites that experience different degrees of habitual sun exposure, including the buttocks, trunk, and head/neck area. The lineage of each cell was confirmed by their cytological features and gene expression profiles (Fig. S1C-D). Most clonal expansions were sequenced at exome resolution (95X coverage on average), including all keratinocytes.

The median mutation burden of keratinocytes was 1.14 mutations per megabase (mut/Mb), which was lower than the mutation burdens of melanocytes (3.91 mut/Mb) and fibroblasts (1.92 mut/Mb, Fig. 1A). These differences held up within most skin biopsies where multiple cell types were sequenced, thus reflecting cell type variation rather than donor-to-donor variation. Keratinocytes from sun-exposed skin had higher mutation burdens than those from sun-shielded skin, but the differences were smaller than in other cell populations (Fig. S2A). For instance, keratinocytes from the upper back had a median mutation burden of 1.70 mut/Mb versus 0.38 mut/Mb for keratinocytes from the buttocks. By contrast, melanocytes from the upper back had a median mutation burden of 14.81 mut/Mb versus 0.30 mut/Mb for melanocytes from the buttocks (Fig. S2A). Curiously, cells from the upper back, irrespective of type, had higher mutation burdens on average than from the head/neck area, seemingly at odds with the cumulative doses of UV exposure typically experienced at these sites. However, this finding is consistent with previous observations by our group(23) and others(11), warranting future studies to understand how cells from chronically sun-exposed body sites, such as the head/neck, keep their mutation burdens relatively low.

The low mutation burden of keratinocytes compared to melanocytes and fibroblasts is unexpected. Keratinocyte stem cells and melanocytes both reside in the basal layer of the epidermis and are expected to receive similar doses of UV radiation. Fibroblasts reside in the underlying dermis and thus would be expected to receive lower doses of UV radiation than either keratinocytes or melanocytes.

While most keratinocytes had low mutation burdens, some had mutation burdens as high as 49.71 mutations/Mb (Fig 1B-C). We annotated cells with mutations known to be pathogenic in cutaneous

squamous cell carcinoma(6), and every keratinocyte with more than 3.5 mutations/Mb had at least one pathogenic mutation. Among these, keratinocytes with missense mutations in *TP53*, which are known to confer a dominant negative effect on the protein(28), had the highest mutation burdens. These findings suggest that wild-type keratinocytes (i.e. cells without pathogenic mutations) are remarkably well-adapted to repair DNA damage or undergo cell death when levels of DNA damage are beyond repair. In support of this, Gilchrest and colleagues previously noted that keratinocytes are more likely than melanocytes to undergo apoptosis in sunburned skin(29), which would select, over time, for keratinocytes with relatively lower mutation burdens. However, pathogenic mutations, such as *TP53* mutations, eliminate the ability of keratinocytes to undergo apoptosis, which would produce, over time, a subpopulation of keratinocytes with hypermutated genomes.

Mutational signature analyses (30) revealed differences in the types of mutations between keratinocytes, melanocytes, and fibroblasts (Fig. 1B, S2B). SBS7a, which has been attributed to UV radiation, was present in keratinocytes, however it contributed to a lower proportion of mutations than in melanocytes and fibroblasts (Fig. 1D). Keratinocytes, instead, had higher proportions of mutations with clock-like signatures- SBS1 and SBS5 (Fig.1E), which are associated with aging and cumulative mitoses (31).

To better understand biases that may be introduced by our single-cell genotyping workflow, we compared the mutational landscapes of individual keratinocytes to bulk cell sequencing of epidermis. Encouragingly, the trinucleotide contexts of mutations from individual keratinocytes, aggregated together, were nearly identical to those observed in an independent study that performed bulk-cell sequencing of epidermis(11) (Fig. S3A).

While the types of mutations in our study were similar to bulk-cell measurements, the numbers of mutations per cell were somewhat lower. The average mutation burden per keratinocyte in our study was 3.65 mutations/Mb. Other studies have inferred, from bulk-cell sequencing of epidermis, average mutation burdens per cell that ranged from less than 1 to greater than 20 Mut/Mb, though most report mutation burdens of approximately 5 mutations/Mb (10–12). However, these measurements are not directly comparable because the data originated from different donors and body sites. Therefore, we performed bulk-cell sequencing of two epidermal microbiopsies, immediately adjacent to skin samples used for single-cell expansions (Fig. S4). In this patient-matched comparison, the average mutation burden per cell from single-cell measurements was approximately three-fold lower than the mutation burdens inferred from patient-matched bulk-cell data. This difference may be caused by a random sampling, given that individual cells display a wide range of mutation burdens. As another possibility, the epidermis contains other cell types, including melanocytes, that may skew its mutation burden higher in bulk-cell measurements. However, it is possible that keratinocytes with low mutation burdens are more likely to clonally expand. Even if this latter explanation were true, our single-cell genotyping workflow clearly reveals cell-to-cell mutation variation, which cannot be appreciated from bulk-cell sequencing.

Next, we analyzed skin cells for somatic copy number changes. Autosomal copy number alterations were infrequent, occurring in only 5.8% of keratinocytes, 13.7% of melanocytes, and 0% of fibroblasts. When copy number alterations were present, they typically affected a small portion of the genome (e.g. a single chromosomal arm), indicating that chromosomal instability is not a major mutational mechanism operating in normal skin cells.

Some keratinocytes shared a portion of their somatic mutations with other keratinocytes from the same biopsy, indicating that they are clonally related (Fig. 2A, S5). We inferred the area occupied by clones from the size of each biopsy and the proportion of cells with shared mutations. The median keratinocyte clone occupied 6.21 mm^2 (Fig. 2B). These surface area estimates are consistent with the upper end of clone sizes estimated by Martincorena and colleagues(10), who made their inferences from deep sequencing of bulk tissue. Our approach to clone detection is likely missing smaller clones, whose detection would require sequencing more cells per square millimeter. We compared biopsies in which keratinocytes and melanocytes were sampled at a similar density. Keratinocyte clones were more prevalent than

melanocyte clones and less likely to harbor pathogenic mutations (Fig. 2C-D). Interestingly, clones of keratinocytes with pathogenic mutations were not larger than clones of keratinocytes without pathogenic mutations (Fig. 2B). Martincorena and colleagues also found that clones with mutations in *NOTCH1*, *TP53*, or *FAT1* were only marginally larger than clones without pathogenic mutations(10).

Genetic alterations driving the transition from actinic keratosis to squamous cell carcinoma

The accumulation of mutational damage can induce a keratinocyte clone to grow into a neoplasm known as an actinic keratosis, which has the capacity to further progress to squamous cell carcinoma. To better understand these transitions, we studied a cohort of archival tissues from 16 patients with squamous cell carcinoma, each immediately adjacent to an actinic keratosis (Table S3, see Fig. 3A for an example). The histologically distinct regions were marked by a pathologist and dissected for DNA sequencing. Deep sequencing (380-fold coverage) was performed using a cancer gene panel (Table S4). We prioritized sequencing depth over a broader sequencing footprint because keratinocyte cancers tend to have substantial levels of stromal cell contamination(6), and the high sequencing depth was helpful in resolving the different populations of clonally related cells within each dissection.

After DNA sequencing, somatic point mutations were stratified by their allele frequencies in each area to uncover the relationship between the dissected tissue regions. To our surprise, the squamous cell carcinoma was often not related to the neighboring actinic keratosis. In 6 of the 16 cases, the actinic keratosis and adjacent squamous cell carcinoma did not share somatic alterations (see Fig. S6A for an example), suggesting that the lesions arose as independent clones, despite their proximity. In 4 other cases, there were no mutations exclusive to the squamous cell carcinoma (see Fig. S6B for an example), implying there was a single population of clonally related cells spanning both dissected tissue areas, with no identifiable mutations accounting for the progression to invasive carcinoma. For these cases, the actinic keratosis histology may represent an extension of the squamous cell carcinoma rather than a distinct precursor lesion.

From the original 16 squamous cell carcinomas, there were only five that clearly evolved from the neighboring actinic keratosis, evidenced by having both a cluster of shared and unshared mutations, as indicated in figures 3B and S7. We prioritized these *bona fide* cases of squamous cell carcinoma arising from an actinic keratosis for further analyses, illustrated by an example case shown in figure 3.

We annotated mutations in genes known to drive keratinocyte cancers, and in the example case, the actinic keratosis had loss-of-function mutations affecting *TP53*, *NOTCH1*, *NOTCH2*, and *CDKN2A* as well as a gain-of-function mutation affecting the *TERT* promoter (Fig. 3C). The squamous cell carcinoma additionally acquired loss-of-function mutations in *ARID2* and *CBL*. There were also some mutations that clustered separately from the dominant clones of the actinic keratosis and squamous cell carcinoma (Fig. 3B, grey data points). These mutations had low allele frequencies in the actinic keratosis and/or squamous cell carcinoma. This could be due to subclones of cells, or more likely contamination from unrelated clones of keratinocytes in the tissue sample; therefore, we did not include these mutations in our phylogenetic analyses. There were no discernible copy number alterations in the actinic keratosis or squamous cell carcinoma of the example case (Fig. 3D), though there was allelic imbalance of chromosomal arm 9p, affecting the *CDKN2A* gene (Fig. 3E). Based on the distribution of shared and unshared somatic alterations in the dominant clones, we inferred the order in which mutations occurred (Fig. 3F) and used immunohistochemistry to validate some of these observations. p53 immunoreactivity was present in both the actinic keratosis and squamous cell carcinoma (Fig. 3G), consistent with a missense mutation in *TP53* present in both areas. Higher phospho-MAPK signaling was observed in the squamous cell carcinoma (Fig. 3H), consistent with the *CBL* mutation in the squamous cell carcinoma. Similar phylogenetic analyses were also performed on the other four squamous cell carcinomas that evolved from actinic keratoses (Fig. 4A).

To supplement our cohort, we reanalyzed publicly available data from another study that sequenced 160 known cancer genes in cutaneous squamous cell carcinomas and adjacent skin(18). In that study, the adjacent skin biopsies were either: sun exposed skin, actinic keratosis, or squamous cell carcinoma *in situ*.

We observed patterns similar to our cohort in this data. In most cases (Fig. S8), the squamous cell carcinoma was unrelated to any clones in the adjacent skin. In other cases (Fig. S9A), the squamous cell carcinoma likely extended into the neighboring skin. Finally, there were 3 *bona fide* cases in which the squamous cell carcinomas evolved from neighboring precursor lesions (Fig. S9B).

We explored common patterns of evolution in the 8 squamous cell carcinomas that clearly evolved from precursor lesions (5 from our cohort [Fig. 3F and 4A] and 3 from Kim and colleagues [Fig. S9B]). As expected, mutations in the p53 and NOTCH signaling pathways typically resided on the trunks of phylogenetic trees (Fig. 4B), though some cases had more than one mutation in these pathways, with secondary hits residing on the branches of phylogenetic trees. Mutations that abrogate cell-cycle control checkpoints or upregulate telomerase also fell on the trunks of phylogenetic trees, indicating that these alterations contribute to the formation of actinic keratoses. By contrast, mutations that disrupt the SWI/SNF chromatin remodeling complex and mutations that activate the RAS/MAPK/PI3K-signaling cascade were most commonly observed at the transition to squamous cell carcinoma (Fig. 4B).

To complement the analyses of squamous cell carcinomas and matched precursor lesions, we also compared the frequency of driver mutations in publicly available data from unmatched cohorts of fully-evolved squamous cell carcinomas(6) and biopsies of normal skin(10) (Fig. 4C). *TP53* and *NOTCH* mutations were common in normal skin, confirming that they undergo selection early, even before neoplasms are present. *TP53* mutations were less common in normal skin than *NOTCH1* mutations but more common in squamous cell carcinoma, suggesting that *TP53* mutations endow keratinocytes with more malignant potential, as has previously been shown in the esophagus(32,33). Mutations in other genes, such as *CDKN2A* and *ARID2*, were rare in normal skin but common in squamous cell carcinoma, implying that they undergo selection comparatively later in tumor evolution. A limitation to these comparisons is that the normal skin biopsies were sequenced with a small gene panel, precluding a comprehensive comparison of mutation frequencies in all genomic loci, such as the *TERT* promoter.

Spatial transcriptomic analysis of actinic keratoses adjacent to squamous cell carcinomas

Bulk-cell RNA-sequencing has been performed on normal skin, actinic keratoses, and squamous cell carcinomas(16,17,34), providing insights into the gene expression changes that occur during tumor evolution. However, as a limitation to those studies, the data encompasses mixtures of clones whose phylogenetic relationships are unknown. Here, we performed spatial transcriptomics (10X Visium) on five of the squamous cell carcinomas adjacent to actinic keratoses, whose clonal relationships were resolved as outlined above.

The spatial transcriptomics data helped define the localization of tumor cells, revealing a complex spatial architecture of clones. We inferred copy number from individual spots of the Visium arrays, as previously described(35), and identified spots with copy number profiles similar to those observed in bulk-cell DNA-sequencing data (see Fig. S10A for an example). In each case, the main lesion of squamous cell carcinoma had concordant copy number alterations, but interestingly, the spatial data revealed satellite colonies of cells, physically distant from the main tumor (see Fig. S10C for an example). The presence of areas of squamous cell carcinoma outside the main lesion was consistent with DNA-sequencing data, in which we recurrently inferred low levels of cross-contamination between actinic keratosis and squamous cell carcinoma (Fig. 3B, S7).

Spots clustered primarily by cell type (Fig. S11B) and secondarily by cell state (Fig. S11C-D, S10B). Broadly, there were clusters of spots from stromal cells, immune cells, adnexal structures, or tumor cells. We selected spots overlying actinic keratosis or squamous cell carcinoma, aided by the distribution of copy number alterations, and performed differential gene expression analyses. On average, spots overlying squamous cell carcinoma expressed higher levels of stem cell, progenitor, and mesenchymal genes, in agreement with bulk-cell data(34) (Fig. S11C), but the spatial data revealed notable heterogeneity within these tumors.

Within actinic keratoses and squamous cell carcinomas, gene expression programs spanned a range of differentiation states (Fig. S11D-E). Spots with stem-like signatures were exclusive to the invasive front of squamous cell carcinoma, corresponding to the “tumor specific keratinocytes” or “TSKs”, defined by Ji and colleagues(36). TSKs express mesenchymal genes, which are typically observed during brief periods of epithelial development or after injury. There were also layers of cells in both the actinic keratoses and the interior portions of squamous cell carcinomas that expressed basal, suprabasal, spinous, and corneocyte gene expression programs. This spectrum of differentiation states was observed regardless of somatic mutation background. Taken together, there is a shift in the average state of epithelial cells towards a progenitor fate during tumor progression, however, hierarchies of differentiation are maintained, even in fully evolved tumors.

Finally, we observed spatial heterogeneity in gene expression of non-tumor cells. It was common for immune infiltrates to extend along the borders of both the actinic keratosis and squamous cell carcinoma, but immune cells expressed different gene expression programs, depending on their localization (Fig. 5, S12). We observed higher expression of immune checkpoint ligands (*PVR*, *NECTIN2*, *CD274*, *CD80*, and *CD86*) in the tumor cells at the invasive front of the squamous cell carcinomas. Concordantly, we observed higher expression of immune checkpoint proteins (*CTLA4*, *TIGIT*, and *PDCD1*) in the lymphocytes at the invasive front of squamous cell carcinomas.

Discussion

Our work provides different vantage points into the changes that occur during the transformation of keratinocytes to squamous cell carcinoma (Fig. 6), starting with individual keratinocytes of normal epidermis. Our single cell analysis revealed a surprisingly broad range of cellular mutation burdens. Cells with high mutation burdens harbored pathogenic mutations, typically affecting *TP53* or *NOTCH1*. p53 conveys DNA damage signals and induces cell cycle arrest to increase repair time or cell death when damage surpasses a threshold, such as after a sunburn(9). Cells with defective p53 are thus likely to accumulate DNA damage at a higher rate. *NOTCH1* mutations induce a stem/progenitor cell state in epithelial cells(37), which may also prevent cells from undergoing apoptosis upon excessive DNA damage. Loss-of-function mutations in *TP53* and *NOTCH1* are known to provide a fitness advantage to epithelial cells(32,38,39), but our findings suggest their contribution to tumor progression may primarily stem from the mutator phenotypes that they induce. Indeed, *TP53* and *NOTCH1* mutant clones were no larger than clones without pathogenic mutations. Taken together, the lower mutation burden of keratinocytes without these mutations compared to epidermal melanocytes and dermal fibroblasts may result from a lower threshold of keratinocytes to undergo apoptosis.

TP53 and *NOTCH1* mutations thus likely prime keratinocytes for transformation by increasing their mutation rates, but additional driver mutations are needed to form a neoplasm (Fig. 6). To uncover the secondary mutations and deduce the order in which they undergo selection, we compared the mutational landscapes of normal epidermis, actinic keratoses, and squamous cell carcinomas. *FAT1* mutations, which activate the Hippo pathway, were observed in individual keratinocytes of normal epidermis, though they were less common than *TP53* and *NOTCH1* mutations. Loss-of-function mutations of *CDKN2A* and *TERT* promoter mutations were rare in normal epidermal cells but were recurrently present in actinic keratoses. Finally, *ARID2* mutations and gain-of-function mutations in the RTK-RAS-MAPK pathway were enriched specifically in squamous cell carcinomas. While Notch-pathway mutations were typically present in the earliest phases of progression, most squamous cell carcinomas acquired additional hits to this pathway. There are multiple NOTCH receptors with varying degrees of functional redundancy(40), and therefore multiple hits may be required to fully ablate this signaling pathway. There were exceptions to these patterns, implying there is more than one route to squamous cell carcinoma. Nevertheless, the general patterns of mutations observed at each stage of evolution reveal the main barriers that evolving neoplasms must overcome during the transition from keratinocyte through pre-neoplastic and pre-malignant phases of tumorigenesis.

The spatial architecture of keratinocyte clones in tumor-bearing skin revealed a complex mosaic of neoplastic and non-neoplastic cells. Squamous cell carcinomas were often genetically unrelated to neighboring precursor lesions, suggesting that “collisions” of clonally unrelated neoplastic proliferations of keratinocytes are common. When we inferred the localization of clones in spatial transcriptomic data, the boundaries of keratinocyte clones were not contiguous in 2-dimensional space. Lineage-tracing experiments in mice show that clones of epithelial cells from surrounding tissue can invade, infiltrate, and mix with tumors as they grow(41,42), which may explain the complex patchwork of clones we observed. For clinical purposes, it is not safe to assume that keratinocyte lesions in close proximity are necessarily clonally related or that the histological boundaries are reliable measures of tumor extent.

Finally, we observed a remarkable degree of spatial heterogeneity in gene expression. A stratified hierarchy of differentiation, resembling that of normal epidermis, was maintained within actinic keratoses and squamous cell carcinomas, though cells from squamous cell carcinomas were, on average, less differentiated. The composition of immune cell infiltrates also varied within tumors. The immune cells near the leading edge of squamous cell carcinomas expressed high levels of immune checkpoint genes. These observations suggest that an immune response is mounted already at the actinic keratosis stage, possibly resulting in an equilibrium state of oncogene-mediated proliferation and immune cell-mediated elimination of partially transformed keratinocytes. However, that state is broken during tumor progression once cells of the carcinoma develop the ability to engage immune-cell checkpoints. It is unclear whether this immune evasion is driven by specific somatic mutations in tumor cells, but it is intriguing that mutations affecting the SWI/SNF chromatin remodeling complex are common at the transition from actinic keratosis to squamous cell carcinoma because there is an emerging view that these mutations drive immune evasion(43).

In closing, our study reveals key events during the transformation of keratinocytes to squamous cell carcinoma, and we provide a blueprint, for future studies(44), on how mutational and gene expression profiling with spatially aware technologies can be used to understand key transitions during tumorigenesis.

Methods

Data Availability

This study is part of the Human Tumor Atlas Network (HTAN), which is funded by the National Cancer Institute (U01 CA294536). The goal of HTAN is to catalog molecular transitions during the evolution of cancer. Raw and intermediate data are immediately available, as described below. These data will also be accessible through the HTAN data portal after the next data release (currently anticipated for Spring of 2025).

The DNA and RNA sequencing data of individual skin cells is available in dbGaP (phs001979.v1.p1 and phs003683.v2.p1). The DNA sequencing data and spatial transcriptomic data from the cutaneous squamous cell carcinomas in association with actinic keratoses are available in dbGaP (phs003282.v2.p1).

Intermediate levels of analysis are also available. Somatic mutation calls for individual cells are available in supplemental tableS2 and were deposited in cBioPortal. A summary of genetic alterations in each keratinocyte as well as copy number data from each cell is available on figshare: https://figshare.com/projects/Genetic_evolution_of_keratinocytes_to_cutaneous_squamous_cell_carcinoma/199837. Publicly available mutation data, covering the progression of squamous cell carcinoma from potential precursor lesions, was retrieved from TableS3 of Kim *et. al.*, *JID*, 2022(18), and the pertinent portions of their dataset, which supported our analyses, are reprinted as part of this publication in Table S6. Publicly available mutation data, covering the somatic point mutations in epidermal biopsies was retrieved from supplementary dataset S1 of Martincorena *et. al. Science*, 2015(10).

Collection of normal human skin samples for keratinocyte genotyping

Skin biopsies were collected from cadaver tissue through the UCSF Willed Body program or from patients seen by Dermatologists at the University of California San Francisco or Northwestern University. Living patients consented to participating in this study through approved protocols by their respective institutional review boards (UCSF IRB 22-36678 and Northwestern IRB STU00211546). Cadaver tissue came from donors who broadly consented, prior to their death as part of their living will, to the use of their tissues for medical research and/or educational purposes. In both instances, we typically took small shave biopsies (3-5mm in their longest dimension) of skin samples.

In addition to skin samples, which were used for single-cell genotyping, we collected a source of reference DNA from each donor. Living donors submitted a buccal swab, which was used as a reference source of germline DNA. Reference DNA from cadaver tissue came from a separate tissue biopsy, unrelated to and distant from the skin sample being processed for somatic mutation analyses.

Sequencing DNA/RNA from individual skin cells

Skin cells were genotyped as previously described(23). Below, we overview the genotyping workflow with an emphasis on modifications made in the present study.

Each skin biopsy was treated overnight in dispase (10mg/ml) to break up collagens holding the dermal layer of the skin to the epidermis. After treatment, the epidermis was separated from the dermis with tweezers, minced, and trypsinized to form a suspension of single cells. The single-cell suspension of epidermal cells was divided into two portions, and cells from each portion were cultured under different conditions. One portion of cells was plated in CNT40 media (CELLnTEC), which favors melanocyte proliferation, and the other portion of cells were seeded in KSFM media (Gibco, 10724-001), which favors keratinocyte proliferation. The remaining dermis was separately minced, broken down by 0.2 mg/ml collagenase (Roche) at 37˚C for 30 minutes and filtered through a 40μm nylon mesh. The resulting single-cell suspension of dermal cells was seeded in DMEM (Gibco, 11965092), favoring the outgrowth of fibroblasts.

The bulk cultures of cells were allowed to recover until they showed signs of stabilization and proliferation (typically 2-10 days). After their establishment, the bulk cells were manually single-cell sorted into individual wells of a 96-well plate using limited dilution. The cells were diluted so that on average one cell would be deposited in every other well (concentration of 0.5 cells/well). We chose limited dilution over the usage of flow cytometry because it yielded a higher proportion of surviving cells. When using the limited dilution method, a healthy population of bulk cells would successfully seed new plates at approximately 50% efficiency, suggesting that the seeded cells were representative; however, we cannot entirely rule out the possibility of a bias being introduced by this step.

Prior to single-cell sorting, the bulk cultures of cells were primarily composed of either fibroblasts, melanocytes, or keratinocytes, depending on the media in which they were maintained, but each culture, nevertheless, contained a mixture of cell types. To further enrich for melanocytes or keratinocytes, we performed differential trypsinization when sorting individual cells from bulk-cell cultures of epidermis. Melanocytes are less adherent than keratinocytes and can be separated with a quick (typically 3 minutes) treatment of trypsin (0.05%), whereas keratinocytes require a longer or sometimes multiple treatments to be released from the tissue culture dish. After single-cell sorting, each cell type was maintained in its optimal media and allowed to clonally expand.

Plates were screened a day after seeding and wells with more than 1 cell were not processed further to ensure that colonies started from a single cell. As an additional safeguard, we discarded data from colonies with mutant allele frequencies averaging less than 50%, revealed at the analysis step, indicating that these wells had more than one founder cell.

Cells were clonally expanded over a period of 2-4 weeks, typically forming colonies of approximately 200 cells before undergoing growth arrest. It is unclear why cells arrest after a brief period of growth, but we believe the daughter cells maintain a “memory” of mitotic signals, at least for a few population doublings, as previously described(45), before stress signals build up and induce growth arrest. While the increase in cell number was modest, it relieved a major bottleneck that ordinarily prohibits the detection of somatic mutations at high specificity and sensitivity from a single cell.

We considered the possibility that the brief period of tissue culture would introduce somatic mutations. However, we(23) and others(27) showed that the number of mutations from 2-10 days in tissue culture (the period of time in which the cells were maintained in culture prior to single-cell seeding) was extremely small (at most, 0.1mutations/Mb). Of note, somatic mutations that arose after single-cell seeding and during clonal expansion had subclonal allele frequencies and were removed.

Clonal expansion increased the starting template material, but the genomic material from 200 cells is insufficient to be directly sequenced with standard next-generation sequencing workflows. Therefore, we extracted and separated genomic DNA and mRNA using the G&T-seq protocol(46,47). The mRNA was amplified with a modified version of the SMART-seq2 protocol, as described(46,47). The genomic DNA was amplified with multiple displacement amplification (Qiagen REPLI-g Single Cell Kit, 150345) or via primary template amplification(48) (BioSkryb ResolveDNA Whole Genome Amplification Kit, 100136).

Amplified cDNA, amplified genomic DNA, or bulk-cell genomic DNA (reference tissue) samples were prepared for next-generation sequencing. Nucleic acids were sheared to a target size of 350bp (Covaris LE220), end-repaired, ligated to IDT 8 or 10 dual index adaptors and amplified using KAPA HyperPrep Kit (Roche, KK8504). Libraries were enriched for exomic sequences by hybridization with NimbleGen SeqCap EZ Exome + UTR (Roche, 06740294001) or KAPA HyperExome (Roche, 09062556001) baits, according to manufacturer’s protocols. Paired-end sequencing (either 100 or 150bp) was performed on one of the following Illumina instruments: Illumina HiSeq 2500 or NovaSeq 6000.

DNA-sequencing data was aligned to the hg19 version of the genome with BWA (v2.0.5)(49). Sequencing reads were deduplicated with Picard (v2.1.1) and further curated (indel realignment and base quality recalibration) with GATK (v4.1.2.0). RNA-sequencing data was aligned to the genome and transcriptome with STAR align (v2.1.0)(50) and deduplicated with Picard (v2.1.1). Gene-level read counts were quantified with RSEM (v1.2.0)(51).

Confirming the lineage of the cell type (Related to figure S1C-D)

In addition to using morphology to identify the lineage of each cell, we inferred cell identity based on gene expression patterns. A t-SNE plot was generated using Rtsne R package (v0.16), showing three distinct clusters of cells. Morphologically, the three clusters of cells corresponded to keratinocytes, melanocytes, and fibroblasts. We performed differential gene expression analysis to identify the top genes associated with each cluster using DESeq2 R package (v1.38.3)(52). The top genes are shown in figure S1 as a heatmap.

Point mutation calling from single-cell expansions

Somatic mutations were called from colonies of individual skin cells as previously described(23) and summarized below. Code related to these operations can be found here: https://github.com/elliefewings/Melanocytes_Tang2020.

MuTect2 (v4.1.2.0) was used to generate a candidate list of point mutations by comparing the aligned bam files of each single-cell expansion to the bam files representing the respective patient’s normal DNA. Pindel was used to generate a candidate list of short insertions and deletions using the same comparison. Pindel calls were filtered to identify candidate mutations with at least 4 reads of support, which were manually inspected to eliminate alignment artifacts. These steps removed sequencing- and alignment-induced artifacts but not artifacts induced during template amplification.

We used two strategies to distinguish *bona fide* mutations from amplification-induced artifacts. First, mutations that were present in both the DNA-sequencing and RNA-sequencing data were considered to be true mutations because it is unlikely the same artifact would be introduced when amplifying DNA and also when amplifying RNA. Mutations present in DNA-sequencing data but absent in RNA-sequencing data of at least 15X coverage were considered to be artifacts with some exceptions. An exception was made for truncating mutations (i.e. nonsense, splice-site, or frameshift mutations) because a truncating mutation, encoded in DNA, is likely to undergo nonsense-mediated decay after expression and may not be detectable in RNA-sequencing data. An exception was also made for X-chromosome mutations from female samples because mutations on the silenced X-chromosome may not be expressed.

Second, we considered a variant to be a true mutation if it occurred in complete linkage with one of the alleles from a nearby heterozygous SNP – a strategy that has also been validated and utilized by others(53). Mutations that were not in complete linkage with nearby SNPs were considered to be artifacts unless they occurred in a region with copy number gains. This strategy works well because we genotyped colonies of cells (as opposed to individual cells). Since there were multiple template molecules, corresponding to each allele, an artifact rarely appears in complete linkage with either haplotype after amplification.

The strategies, described above, enabled us to validate (or invalidate) variants within expressed genes and/or variants that could be phased into their respective haplotypes. We used the variant allele frequencies of these credentialed variants to establish a benchmark to determine the statistical likelihood that the remaining variants (those in poorly expressed genes and unphaseable portions of the genome) were bona fide mutations or artifacts. Most artifacts had low allele frequencies, whereas bona fide mutations tended to have allele frequencies of 50% (for heterozygous mutations) or 100% (for homo- or hemi-zygous mutations). See Tang *et al.* for more details(23) on how we arrived at a specific cutoff for each sample.

Copy number calling from single-cell expansions

Copy number was inferred from colonies derived from individual skin cells using CNVkit (v.0.9.6.2). Our copy number workflow is described in detail here(23). Briefly, CNVkit infers copy number from either DNA-sequencing(54) or RNA-sequencing(55) data. Since we produced matching DNA/RNA-sequencing data from each colony, we ran CNVkit in both modes. When running CNVkit on either DNA- or RNA-sequencing data, we generated a reference from large pools of samples that were of the same lineage and run in the same sequencing batch. We considered a copy number alteration to be true when it was detected in both the DNA- and RNA- sequencing data. Copy number inferences at the bin level (.cnr files) or segment level (.cns files) are available here: https://figshare.com/s/9474ef6f59d92dc082f8.

Allelic dropout from single cell expansions

A set of germline heterozygous SNPs was identified from the reference bam of each donor as described(23). Briefly, we called variants in the reference bam against the reference hg19 genome using FreeBayes (v.1.3.1), identified variants that had been observed in greater than 1% of participants from the 1000 Genomes Project, and required variants to have at least 5 reads supporting each allele and a variant allele frequency between 40-60% for each allele.

After establishing these sets of germline heterozygous SNPs for each donor, the number of reference and alternate reads were counted in the bam files from single-cell expansions of each donor. We calculated rates of mono-allelic and bi-allelic dropout for each colony. We also consulted the copy number data to distinguish between biological dropouts, resulting from a deletion, versus technical dropout, resulting from amplification biases during sample preparation. Dropout rates are listed in Table S1 in columns N and O. The median allelic dropout rate, across all single-cell expansions, was 0.06%, confirming our ability to sensitively detect single nucleotide variants in single-cell expansions.

Mutation burden and signature analysis (related to figures 1, S2, and S3)

Mutation burdens were calculated as mutations per megabase. The number of mutations was tabulated for each cell and divided by the captured footprint that was sequenced with 10X coverage or greater. The footprint of sequencing data with 10X coverage or greater was counted with the footprints software(23). A trinucleotide profile for each individual cell was generated using deconstructSigs R package (v1.9.0)(56). The Bioconductor library BSgenome.Hsapiens.UCSC.hg19 (v1.4.3) was first used to apply mutational context to all the single base substitution (SBS) mutations identified in each cell. The results for all the cells were combined for each cell type and visualized as trinucleotide context in figure S3. A custom forward stagewise algorithm using SigProfilerAssignment (v0.1.8) was applied to build a mutational profile based on 78 pre-defined COSMIC (v3.7) signatures previously extracted by SigProfiler (57). The minimum number of SBS mutations for the signature analysis is set at 10. The signatures for all the cells are depicted as stacked barplots in figure 1B (bottom panel) showing the fractions of top 7 signatures. Signatures present in less than 10% of cells were grouped into an "others" category.

Annotation of pathogenic mutations in individual cells (related to figure 1B and S2B)

As a guide for annotation of pathogenic mutations, we identified mutations in genes shown to be under selection from a meta-analysis of cutaneous squamous cell carcinoma(6). Three authors on this manuscript independently reviewed the mutation lists to nominate pathogenic mutations. After independent review, we consulted and agreed upon a single list, as annotated in column AA of Table S2. The full list of mutations, including passenger mutations, is also available in Table S2 for interpretation by the readers.

Construction of phylogenetic trees from individual skin cells (Related to figure 2A)

After calling mutations in individual cells, overlapping mutations between cells from the same donor were identified. Only mutations with at least 10X coverage were considered for phylogenetic analyses – we made this decision to reduce the risk of calling a mutation private to one cell when coverage over the mutant site was low in other cells. If any mutations were shared between 2 or more cells, we ran the mpileup function of SAMtools to count the reference/mutant reads for all other mutations in all cells to ensure that they were genuinely private to each of the cells. Only in rare cases, this identified mutations that were in fact shared but had been missed by the mutation calling algorithm because they were just below the MuTect2 detection thresholds. In these rare cases, the mutation was added to any samples for which it was missed.

After compiling a list of shared and private somatic mutations across the cells, a phylogenetic tree was constructed using an R script that employs the dplyr, tidyr, and ggplot2 packages. The script is publicly available for use at [https://github.com/delahny/phylogenetictrees]. It identifies and counts both shared and private mutations to generate datasets that corresponds to branches and trunks, which are plotted by applying hierarchical leveling. The resulting phylogenetic tree was rooted at the germline state, with trunks scaled according to the number of shared mutations and branches scaled based on the number of private mutations. Each cell's identity was labeled at the terminus of its branch, and cells without a vertical branch had no private mutations.

Construction of clonality plots (Related to figures 2A and S5)

After constructing phylogenetic trees, we drew schematic images to depict the clonal architecture of each biopsy. In these plots, individual cells are represented as points, with phylogenetically related cells grouped within circles. Cells sharing pathogenic mutations and their corresponding clones are highlighted in red. In these schematics, the precise spatial localization of a given cell is unknown.

To estimate the total surface area of each biopsy, we calculated the surface area based on the diameters of biopsies. For biopsies with non-circular shapes, measuring scales, present within the biopsy images, were used to calculate the area with ImageJ. For each set of clonally related cells, we estimated

the surface area using the formula: (number of phylogenetically related cells / total cells genotyped from the biopsy) * total area of the biopsy. In samples with subclones, the largest encompassing circle in each clonal relationship represents the trunk in the phylogenetic tree, while smaller concentric circles illustrate subsequent waves of clonal expansion. We adjusted the circle areas based on the ratio of phylogenetically related cells to the total cell population within each circle.

In theory, if all cells in a biopsy exhibit a common lineage, the resultant circle would exceed the boundaries of the square due to their geometrical properties. Thus, in cases where any of the circle exceeded the boundary, we proportionally reduced the dimensions of all the inner circles to fit within the outer square's limits.

Dissection of neoplastic tissues for profiling squamous cell carcinomas in association with actinic keratoses

Twenty cutaneous squamous cell carcinomas with an adjacent actinic keratosis were retrieved from the archive of the Dermatopathology Service at UCSF. H&E scanned images were marked by a pathologist to identify the areas of squamous cell carcinoma, actinic keratosis, or non-neoplastic tissue (used as genetic reference). Consecutive unstained sections (10 slides at 10μm thickness) were dissected with a scalpel under a dissection scope to separate the histopathologically distinct tissues. Genomic DNA was isolated using the QIAamp DNA FFPE Tissue Kit (Qiagen, 56404). In four cases, the DNA yields were insufficient in one of the tissue areas, and these cases were discarded from future analyses. The remaining sixteen cases were retained for sequencing. This study was approved by the UCSF Human Research Protection Program, and all tissues were collected in accordance with the Institutional Review Board.

Immunohistochemistry (IHC) staining

P53 immunohistochemistry (IHC) staining (clone DO-7 mouse, Roche, 1:400 dilution) was performed by the UCSF Dermatopathology Service following a clinically standardized protocol as described previously(58). Phospho-MAPK was detected using the Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) antibody (clone 4370, Cell Signaling, 1:100 dilution) by the UCSF Histology and Biomarkers Core on an automated Ventana BenchMark Ultra platform. The primary signal was visualized using the Discovery Purple detection kit. Images were captured using a Zeiss Axio Scanner.Z1 slide scanner equipped with a 20X Plan-Apochromat objective and processed with Zeiss Zen Lite v3.6 software.

DNA sequencing and somatic alteration calling from neoplastic tissues

DNA sequencing and the initial steps of bioinformatic analyses was performed by the UCSF Clinical Cancer Genomics Laboratory (CCGL) as previously described(59). CCGL is a CLIA-approved laboratory that performs gene-panel sequencing of tumors to help guide targeted treatments and assist in diagnosis. Specifically, 20-250 ng of genomic DNA was prepared for sequencing using the KAPA HyperPrep Kit with Library Amplification (Roche, KK8504). Target enrichment with a customized bait panel targeting 538 cancer-relevant genes (Table S4) was performed using NimbleGen SeqCap EZ Developer library (Roche, Ref: 06471706001). Sequencing was performed on an Illumina HiSeq 2500 instrument. Alignment and grooming were performed with Burrows-Wheeler Aligner (BWA)(49), Genome Analysis Tool-Kit (GATK)(60), and Picard (https://broadinstitute.github.io/picard/). Copy number inference was performed with CNVkit(54,55).

We used two separate approaches to call somatic point mutations. First, we ran MuTect (v4.1.2.0) by comparing the tumor bam files to reference bam files from non-lesional tissue of the same patient. We filtered out all variants with fewer than 4 reads. In a typical tumor/normal sequencing study, MuTect is sufficient to call point mutations. However, in our study, the reference bam often had small numbers of mutant reads. Mutant reads were common in normal-appearing tissue of our cohort because occult fields of keratinocytes clonally related to the neoplasm sometimes extended into the histologically normal skin, as we previously demonstrated to be a feature of keratinocyte cancers(61). MuTect tends to reject variants

with reads in the reference bam, even under permissive parameters. MuTect is also not designed to call indels.

To supplement the MuTect calls, we also called variants against the reference genome using UnifiedGenotyper (v4.1.2.0) and FreeBayes (v1.3.1–19)(62). These variant callers were incorporated to identify any point mutations missed by MuTect. We discarded variants from these lists, which were likely to be germline SNPs. We accomplished this goal by removing variants that resided in known SNP sites (from 1000 genomes) and/or had an allele frequency in the normal tissue of greater than 20%. We also discarded variants that were likely to be sequencing and/or alignment artifacts if they were observed in blacklist of artifacts, previously defined by our cancer center's clinical cancer genomics laboratory from a list of recurring variants in panels of hundreds of normal tissues who had been sequenced through their services. The variants that were not filtered out in the above operations were considered somatic mutations and added to the list of somatic variants called by MuTect.

We called indels using Pindel. Candidate indels with greater than 4 supporting reads in the tumor but not the normal were manually inspected. The final list of somatic point mutations and indels is available in Table S5.

Inference of cancer genome fraction in neoplastic tissues (related to table S3)

Tumor genome fraction was inferred bioinformatically using multiple strategies. A short name for each strategy is listed in columns E through I of Table S3 and described in more detail below.

Allelic Imbalance: Allelic imbalance over stretches of heterozygous SNPs is introduced when copy-number-neutral (CNN) loss-of-heterozygosity (LOH) or a deletion occurs in a tumor cell. In sequencing reads originating from tumor cells, the percentage of reads from either allele shifts to 100/0, but remains 50/50 from sequencing reads originating from stromal cells. For fully clonal LOH, the extent of allelic imbalance is therefore proportional to the tumor genome fraction (see Shain *et al.* NEJM, 2015(59) for the specific formula used).

Average Somatic MAF Autosomes: Somatic mutations can be stratified by their mutant allele frequencies (MAFs), which are dictated by the clonality and the zygosity of the mutation. Here, we used the median MAF of somatic mutations occupying portions of the genome without copy number alterations to infer tumor purity. This approach assumes those mutations are fully clonal and heterozygous. In some cases, there was a bimodal distribution of mutant allele frequencies with one population of mutations having extremely low allele frequencies, likely stemming from subclones (also see the cross contamination note below). When this occurred, we only considered, for this calculation, the population of mutations that we believed to be fully clonal. After calculating the median MAF of clonal mutations, we multiplied this value by 2 to arrive at the tumor genome fraction.

Driver Mutation: Some samples had few mutations, precluding the usage of the 'Median Somatic MAF Autosomes' method of inference described above, but every sample had at least one driver mutation. The mutant allele frequency of the driver mutation was used to estimate tumor purity under the assumption that the mutation was heterozygous and fully clonal – before making these assumptions, we checked for loss-of-heterozygosity or a copy number alteration affecting the locus of the driver mutation. We multiplied the MAF of driver mutations by 2 to arrive at the tumor cellularity estimate.

Median Somatic MAF XY: In a male sample, a somatic mutation on the X or Y chromosome will have a mutant allele frequency 100% from sequencing reads derived from the tumor cells. Sequencing reads from stromal cells will not contribute any mutant reads. The observed mutant allele frequency of the mutation can therefore be used to infer the relative proportions of tumor and stromal cells. This approach assumes these mutations are fully clonal and do not reside in chromosomes with copy number alterations.

Cross-contamination estimates (Related to figure 3B, S7, and table S3)

In column J of Table S3, we include a column with a "cross-contamination note". Some the squamous cell carcinomas show signs of contamination with actinic keratosis cells or vice-versa. Typically, contamination was evident when mutations of the squamous cell carcinoma had trace sequencings reads in the actinic keratosis (or vice-versa). The level of trace sequencing reads was used to infer degrees of cross contamination by doubling their allele frequencies (which assumes that the median contaminant mutation is fully clonal and heterozygous).

Allelic imbalance calculation from tumor sequencing data (related to figure 3E)

To measure allelic imbalance in the sequencing data from squamous cell carcinomas and actinic keratoses, we first identified a set of heterozygous SNPs for each patient as described above. Once a set of high confidence SNPs was derived, we counted the ref and alt reads of each SNP in the sequencing data of the squamous cell carcinoma and actinic keratosis. Next, we calculated the variant allele fraction of the major allele (i.e. the more abundant allele in the sequencing data) and subtracted 0.5 (the expected fraction if the allele were sampled equally in the sequencing data). We plotted allelic imbalance values across the genome for each sample to identify contiguous regions with imbalance relative to the background values.

Annotation of pathogenic mutations in squamous cell carcinomas in association with actinic keratosis

See the section entitled, "*Annotation of pathogenic mutations in individual cells*", above for a description of this process. The full list of mutations is available in Table S5 for interpretation, and our annotation of pathogenic mutations can be found in column W.

Phylogenetic tree construction for neoplastic tissues (Related to figures 3F, 4, S8 and S9)

We constructed phylogenetic trees for the 5 squamous cell carcinomas that evolved from the neighboring actinic keratoses (shown in figures 3F and 4A). Mutations were categorized as shared or private between the dominant clones in each area and respectively placed on the trunk or branch of each tree. Our determinations of trunk vs branch mutations are shown in the color-coded scatterplots of figures 3B and S7, and below we describe how these calls were made.

Classifying mutations as shared or private was challenging, due to low tumor cell content and the presence of cross-contamination between different neoplastic areas. To take these factors into account, a mutation was considered present in the dominant clone of the AK or the SCC when it was more than 50% clonal. The clonality of a point mutation was estimated by its allele frequency relative to tumor purity and after accounting for cross contamination and the copy number/zygosity of the mutation. Despite establishing this cutoff, there were instances in which mutations from unrelated clones of keratinocytes appeared to be incorporated into the phylogenetic trees, requiring further refinement on a sample-by-sample basis. To further refine the mutations assigned to each clone, we generated histograms of mutant allele frequencies for the actinic keratosis and squamous cell carcinoma areas. Histograms tended to be multimodal, and we assumed each peak corresponded to the median mutant allele frequency of mutations in a given clone. We identified breaks between peaks in the histograms to remove clusters of low allele frequency mutations that likely stemmed from unrelated clones of keratinocytes.

Due to the complex clonal structure of the skin samples, we elected not to call branch mutations in actinic keratoses. The tumor purities of actinic keratoses were low, and this made it difficult to distinguish clusters of mutations in the actinic keratosis from those originating from unrelated clones in the area.

We also constructed phylogenetic trees for the cutaneous squamous cell carcinomas and adjacent skin samples from Kim and colleagues(18). Since their gene panel was relatively small, fewer mutations were available to calculate tumor cell content and clonality of mutations. Instead, we generated scatterplots of mutation allele frequencies in for different progression stages (see figure S8 and S9) and manually

categorized mutations as shared or private based on how they clustered in these plots. Our phylogenetic trees are shown side-by-side with the original trees from Kim and colleagues.

Comparing the genomic landscape of keratinocytes inferred from single-cell and bulk sequencing workflow (Related to figure S4)

To compare the mutational landscape inferred through our single-cell workflow with that from standard bulk sequencing, we performed both methods on patient-matched biopsies. These biopsies were collected from a deceased 74-year-old Caucasian male donor (D56) and was part of this study via the UCSF Willed Body program. We collected 5mm and 3.5mm punch biopsies from both the face and shoulder. The 5 mm biopsies were used to establish a bulk primary culture, followed by the subsequent single-cell sequencing workflow. A total of 15 keratinocytes from the face and 13 keratinocytes from shoulder were sequenced following the procedures described earlier (Fig. 1, 2, S2-S5). The 3.5mm biopsies were used for bulk-cell sequencing of epidermis. Specifically, for the 3.5mm biopsies, the epidermis and dermis were separated with dispase treatment (10mg/ml for 16 hours). Bulk DNA was isolated from the epidermis using the prepIT.L2P kit (DNA Genotek, PT-L2P-5). This bulk DNA was directly used for sequencing library preparation. Library preparation and hybridization was performed identical to the corresponding single cell samples, utilizing KAPA HyperExome (Roche, 09062556001) probes for exome enrichment. Sequencing (100 bp, paired-end) was performed on the NovaSeq 6000 platform. Somatic mutations were called using MuTect and pindel, as described above. The mutation burden per cell was estimated from bulk-cell sequencing data following a method similar to that described by Martincorena *et al.* (10). In their approach, mutations in normal skin are parsimoniously assumed to be heterozygous. We also made this assumption for autosomal mutations, but we made a slight modification to accounts for the non-diploid nature of sex chromosomes in male samples. The clone size (ρ) was estimated as:

$\rho = 2*VAF$ (autosomal region) and $\rho = VAF$ (X and Y chromosome)

where VAF is the variant allelic frequency of mutations in those regions.

The mutation burden per cell (β) was calculated as:

$$\beta = \Sigma(\rho)/L_{Mb}$$

Here, $\Sigma(\rho) \approx \Sigma(VAF_{X/Y}) + 2*\Sigma(VAF_{autosomal})$, where $VAF_{X/Y}$ represents VAF of mutations in sex chromosomes and $VAF_{autosomal}$ represents VAF of mutations in autosomal regions. L_{Mb} is the footprint (genomic region with good coverage). The L_{Mb} for KAPA HyperExome has been thoroughly validated as 43.2 Mb.

Mutation signature analysis was performed on an aggregated list of the mutations from the 15 keratinocytes sequenced from the face, the 13 keratinocytes sequenced from the shoulder and mutations inferred in adjoining face and shoulder biopsies. The same single-cell pipeline described earlier was applied for this analysis.

Spatial transcriptomics (Related to figures 5, S10, S11 and S12)

Spatial transcriptomics was performed on five squamous cell carcinomas in association with actinic keratoses. We chose a subset of the cases that underwent DNA-sequencing, as described above, so that matching mutational and spatial transcriptomic information would be available. In four of the five cases, the squamous cell carcinoma was genetically related to the neighboring actinic keratosis (BB05, BB12, BB13 and BB16), and in the remaining case, the squamous cell carcinoma was not genetically related to the neighboring actinic keratosis (BB09). All cases were profiled on a version of the 10X FFPE Visium platform.

One case (BB13) was profiled on a relatively older version of Visium (v1.0). For this case, we cut additional sections of the tissue from its original block and placed them within the fiducial frame on the slide. The slide was prepared according to manufacturer's protocols at UCSF. The remaining cases (BB05, BB09,

BB12 and BB16) were profiled with a relatively newer version of Visium (v2.0) that is compatible with a CytAssist machine (10X Genomics). For these cases, we took an existing H&E slide, removed the coverslip, and situated the tissue within the designated capture area on the CytAssist machine, where probe hybridization occurred. Hybridization and preparation for sequencing was performed according to manufacturer's protocols by an outside company, Abiosciences.

Paired-end sequencing was performed by the Center for Advanced Technology at UCSF on an Illumina instrument (NovaSeq 6000). Read 1 (the barcode read) was sequenced with a read length of 28 bp and read 2 (the probe read) was sequenced with a read length of 90 bp. Sequencing data was processed with the SpaceRanger pipeline (v1.3.0 and v2.0.1) to generate a cloupe file, which was visualized in the Loupe browser (version 7, 10X Genomics). The SpaceRanger workflow can run samples one-by-one or in aggregate mode. The data from the four samples (BB05, BB09, BB12 and BB16) that were run on the relatively newer version of Visium were merged into an aggregate run. The sequencing data from the remaining sample (BB13), which was run on a relatively older version of Visium, could not be merged with the others due to differences in chemistry of the platforms.

For the copy number analyses described in figure S10, we used STmut to infer copy number from individual spots, as previously described(35). Briefly, we ran STmut in grouping mode, which can combine contiguous spots from the same gene expression cluster that have fewer than 1000 genes detected. After combining spots with low coverage, the groups of spots are treated as a single spot. In practice, most spots had more than 1000 genes detected and were not affected by this parameter, but this feature improved the signal to noise in a subset of spots with low sequencing coverage. STmut can also accept lists of known copy number alterations and generate q-values for a given spot, which reflects the likelihood that it matches a known copy number profile. We called arm-level gains and losses from DNA-sequencing data and input these calls into STmut. The histograms and Q-Q plot in figure S10 show the spots with high CNVscores (i.e. high similarity to the known copy number profile), which we considered to be spots overlying squamous cell carcinoma.

For the gene expression analyses described in figure S11B and D, we used the graph-based clusters generated by the SpaceRanger software. The identities of the clusters were manually annotated after examining the histology of the underlying cells and their gene expression patterns. Differentially expressed genes across the spots in these clusters were exported from the Loupe browser. For the gene expression cluster in figure S11C, we manually annotated areas of squamous cell carcinoma or actinic keratosis and performed differential gene expression analyses. To annotate squamous cell carcinoma versus actinic keratosis, we considered copy number data and histology to make the final calls.

Acknowledgements

The study was supported by grants from: NIH NCI Human Tumor Atlas (HTAN) network (U01 CA294536), NIH NCI (R01 CA265786), NIH NIAMS (AR080626), Department of Defense Melanoma Research Program (ME210014), Melanoma Research Alliance (Team Science Award and Dermatology Fellows Award), the LEO Foundation Region Americas Award, a private donation by Tracy and Guy Jaquier, and the UCSF Department of Dermatology.

References

1. Nehal KS, Bichakjian CK. Update on Keratinocyte Carcinomas. N Engl J Med. 2018;379:363–74.

2. Mansouri B, Housewright CD. The Treatment of Actinic Keratoses-The Rule Rather Than the Exception. JAMA Dermatol. 2017;153:1200.

3. Karia PS, Han J, Schmults CD. Cutaneous squamous cell carcinoma: estimated incidence of disease, nodal metastasis, and deaths from disease in the United States, 2012. J Am Acad Dermatol. 2013;68:957–66.

4. Wu W, Weinstock MA. Trends of keratinocyte carcinoma mortality rates in the United States as reported on death certificates, 1999 through 2010. Dermatol Surg. 2014;40:1395–401.

5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA: A Cancer Journal for Clinicians. 2016;66:7–30.

6. Chang D, Shain AH. The landscape of driver mutations in cutaneous squamous cell carcinoma. NPJ Genom Med. 2021;6:61.

7. Nakazawa H, English D, Randell PL, Nakazawa K, Martel N, Armstrong BK, et al. UV and skin cancer: specific p53 gene mutation in normal skin as a biologically relevant exposure measurement. Proc Natl Acad Sci USA. 1994;91:360–4.

8. Jonason AS, Kunala S, Price GJ, Restifo RJ, Spinelli HM, Persing JA, et al. Frequent clones of p53-mutated keratinocytes in normal human skin. Proc Natl Acad Sci USA. 1996;93:14025–9.

9. Ziegler A, Jonason AS, Leffell DJ, Simon JA, Sharma HW, Kimmelman J, et al. Sunburn and p53 in the onset of skin cancer. Nature. 1994;372:773–6.

10. Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science. 2015;348:880–6.

11. Fowler JC, King C, Bryant C, Hall MWJ, Sood R, Ong SH, et al. Selection of Oncogenic Mutant Clones in Normal Human Skin Varies with Body Site. Cancer Discov. 2021;11:340–61.

12. King C, Fowler JC, Abnizova I, Sood RK, Hall MWJ, Szeverényi I, et al. Somatic mutations in facial skin from countries of contrasting skin cancer risk. Nat Genet. 2023;

13. Wei L, Christensen SR, Fitzgerald ME, Graham J, Hutson ND, Zhang C, et al. Ultradeep sequencing differentiates patterns of skin clonal mutations associated with sun-exposure status and skin cancer burden. Sci Adv. 2021;7.

14. Kim Y-S, Bang CH, Chung Y-J. Mutational Landscape of Normal Human Skin: Clues to Understanding Early-Stage Carcinogenesis in Keratinocyte Neoplasia. J Invest Dermatol. 2023;143:1187-1196.e9.

15. Wang NJ, Sanborn Z, Arnett KL, Bayston LJ, Liao W, Proby CM, et al. Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma. Proc Natl Acad Sci USA. 2011;108:17761–6.

16. Chitsazzadeh V, Coarfa C, Drummond JA, Nguyen T, Joseph A, Chilukuri S, et al. Cross-species identification of genomic drivers of squamous cell carcinoma development across preneoplastic intermediates. Nat Commun. 2016;7:12601.

17. Thomson J, Bewicke-Copley F, Anene CA, Gulati A, Nagano A, Purdie K, et al. The Genomic Landscape of Actinic Keratosis. J Invest Dermatol. 2021;141:1664-1674.e7.

18. Kim Y-S, Shin S, Jung S-H, Park YM, Park GS, Lee SH, et al. Genomic Progression of Precancerous Actinic Keratosis to Squamous Cell Carcinoma. J Invest Dermatol. 2022;142:528-538.e8.

19. Kennedy SR, Zhang Y, Risques RA. Cancer-Associated Mutations but No Cancer: Insights into the Early Steps of Carcinogenesis and Implications for Early Cancer Detection. Trends Cancer. 2019;5:531–40.

20. Fowler JC, Jones PH. Somatic mutation: What shapes the mutational landscape of normal epithelia? Cancer Discov. 2022;12:1642–55.

21. Kakiuchi N, Ogawa S. Clonal expansion in non-cancer tissues. Nat Rev Cancer. 2021;21:239–56.

22. Menon V, Brash DE. Next-generation sequencing methodologies to detect low-frequency mutations: “Catch me if you can.” Mutat Res Rev Mutat Res. 2023;792:108471.

23. Tang J, Fewings E, Chang D, Zeng H, Liu S, Jorapur A, et al. The genomic landscapes of individual melanocytes from human skin. Nature. 2020;586:600–5.

24. Yoshida K, Gowers KHC, Lee-Six H, Chandrasekharan DP, Coorens T, Maughan EF, et al. Tobacco smoking and somatic mutations in human bronchial epithelium. Nature. 2020;578:266–72.

25. Machado HE, Mitchell E, Øbro NF, Kübler K, Davies M, Leongamornlert D, et al. Diverse mutational landscapes in human lymphocytes. Nature. 2022;608:724–32.

26. Kucab JE, Zou X, Morganella S, Joel M, Nanda AS, Nagy E, et al. A Compendium of Mutational Signatures of Environmental Agents. Cell. 2019;177:821-836.e16.

27. Petljak M, Alexandrov LB, Brammeld JS, Price S, Wedge DC, Grossmann S, et al. Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. Cell. 2019;176:1282-1294.e20.

28. Willis A, Jung EJ, Wakefield T, Chen X. Mutant p53 exerts a dominant negative effect by preventing wild-type p53 from binding to the promoter of its target genes. Oncogene. 2004;23:2330–8.

29. Gilchrest BA, Eller MS, Geller AC, Yaar M. The pathogenesis of melanoma induced by ultraviolet radiation. N Engl J Med. 1999;340:1341–8.

30. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578:94–101.

31. Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Turki SA, et al. Timing, rates and spectra of human germline mutation. Nat Genet. 2016;48:126–33.

32. Abby E, Dentro SC, Hall MWJ, Fowler JC, Ong SH, Sood R, et al. Notch1 mutations drive clonal expansion in normal esophageal epithelium but impair tumor growth. Nat Genet. 2023;55:232–45.

33. Yokoyama A, Kakiuchi N, Yoshizato T, Nannya Y, Suzuki H, Takeuchi Y, et al. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. Nature. 2019;565:312–7.

34. Bailey P, Ridgway RA, Cammareri P, Treanor-Taylor M, Bailey U-M, Schoenherr C, et al. Driver gene combinations dictate cutaneous squamous cell carcinoma disease continuum progression. Nat Commun. 2023;14:5211.

35. Chen L, Chang D, Tandukar B, Deivendran D, Pozniak J, Cruz-Pacheco N, et al. STmut: a framework for visualizing somatic alterations in spatial transcriptomics data of cancer. Genome Biol. 2023;24:273.

36. Ji AL, Rubin AJ, Thrane K, Jiang S, Reynolds DL, Meyers RM, et al. Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. Cell. 2020;182:1661–2.

37. Okuyama R, Tagami H, Aiba S. Notch signaling: its role in epidermal homeostasis and in the pathogenesis of skin diseases. J Dermatol Sci. 2008;49:187–94.

38. Murai K, Dentro S, Ong SH, Sood R, Fernandez-Antoran D, Herms A, et al. p53 mutation in normal esophagus promotes multiple stages of carcinogenesis but is constrained by clonal competition. Nat Commun. 2022;13:6206.

39. Zhang W, Remenyik E, Zelterman 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. Proc Natl Acad Sci U S A. 2001;98:13948–53.

40. Zhou B, Lin W, Long Y, Yang Y, Zhang H, Wu K, et al. Notch signaling pathway: architecture, disease, and therapeutics. Signal Transduct Target Ther. 2022;7:95.

41. Reeves MQ, Kandyba E, Harris S, Del Rosario R, Balmain A. Multicolour lineage tracing reveals clonal dynamics of squamous carcinoma evolution from initiation to metastasis. Nat Cell Biol. 2018;20:699–709.

42. Reeves MQ, Balmain A. Mutations, Bottlenecks, and Clonal Sweeps: How Environmental Carcinogens and Genomic Changes Shape Clonal Evolution during Tumor Progression. Cold Spring Harb Perspect Med. 2024;14:a041388.

43. Chaudhri A, Lizee G, Hwu P, Rai K. Chromatin Remodelers Are Regulators of the Tumor Immune Microenvironment. Cancer Res. 2024;84:965–76.

44. Campbell JD, Mazzilli SA, Reid ME, Dhillon SS, Platero S, Beane J, et al. The Case for a Pre-Cancer Genome Atlas (PCGA). Cancer Prev Res (Phila). 2016;9:119–24.

45. Yang HW, Chung M, Kudo T, Meyer T. Competing memories of mitogen and p53 signalling control cell-cycle entry. Nature. 2017;549:404–8.

46. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods. 2015;12:519–22.

47. Macaulay IC, Teng MJ, Haerty W, Kumar P, Ponting CP, Voet T. Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. Nat Protoc. 2016;11:2081–103.

48. Gonzalez-Pena V, Natarajan S, Xia Y, Klein D, Carter R, Pang Y, et al. Accurate genomic variant detection in single cells with primary template-directed amplification. Proc Natl Acad Sci U S A. 2021;118:e2024176118.

49. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.

50. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.

51. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.

52. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

53. Bohrson CL, Barton AR, Lodato MA, Rodin RE, Luquette LJ, Viswanadham VV, et al. Linked-read analysis identifies mutations in single-cell DNA-sequencing data. Nat Genet. 2019;51:749–54.

54. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput Biol. 2016;12:e1004873.

55. Talevich E, Shain AH. CNVkit-RNA: Copy number inference from RNA-Sequencing data. bioRxiv. 2018;408534.

56. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:31.

57. Islam SMA, Díaz-Gay M, Wu Y, Barnes M, Vangara R, Bergstrom EN, et al. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. Cell Genom. 2022;2:None.

58. Fernandez-Pol S, Ma L, Ohgami RS, Arber DA. Immunohistochemistry for p53 is a useful tool to identify cases of acute myeloid leukemia with myelodysplasia-related changes that are TP53 mutated, have complex karyotype, and have poor prognosis. Mod Pathol. 2017;30:382–92.

59. Shain AH, Yeh I, Kovalyshyn I, Sriharan A, Talevich E, Gagnon A, et al. The Genetic Evolution of Melanoma from Precursor Lesions. N Engl J Med. 2015;373:1926–36.

60. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8.

61. Chen L, Chang D, Tandukar B, Deivendran D, Cho R, Cheng J, et al. Visualizing somatic alterations in spatial transcriptomics data of skin cancer [Internet]. bioRxiv; 2022 [cited 2023 Apr 10]. page 2022.12.05.519162. Available from: https://www.biorxiv.org/content/10.1101/2022.12.05.519162v1

62. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing [Internet]. arXiv; 2012 [cited 2024 May 26]. Available from: http://arxiv.org/abs/1207.3907

Figure Legends

Figure 1. Keratinocytes have distinct mutational landscapes compared to other cell types. A. Mutation burdens (mutations/megabase) of individual keratinocytes (Ker.) compared to melanocytes (Mel.) and fibroblasts (Fib.) **B.** Mutation burden, driver mutations, and mutational signatures for 137 keratinocytes with each column of the three stacked panels representing an individual cell. Top panel: mutation burden of keratinocytes in descending order. Red bars indicate cells harboring one or more pathogenic mutations. Middle panel: tiling plot of pathogenic mutations (rows). Bottom panel: the fractions of different mutational signatures for each cell. White bars indicate keratinocytes with too few mutations to perform signature analysis. **C.** Mutation burdens of keratinocytes with and without pathogenic mutations. **D**. Left panel: fraction of mutations with UV signatures (SBS7a) in keratinocytes, melanocytes, or fibroblasts. Right panel: fraction of cells with detectable SBS7a in keratinocytes, melanocytes, or fibroblasts. **E**. The data is plotted as in panel D but for SBS87. For all plots, an asterisk (*) or a hash (#) denotes $p<0.05$ using the Wilcoxon rank-sum test (cell to cell comparisons) or the Poisson test (proportion comparisons) respectively. Horizontal bars show the median (panels A and C) or mean (panels D and E). Error bars in panels D and E show 95% confidence intervals (Poisson test).

Figure 2. Clonal architecture of keratinocytes in human skin. A. Clonal structure of keratinocytes from four representative skin biopsies (see Fig. S5 for all biopsies). The surface area of each biopsy is drawn to scale, as indicated, with dots representing the cells genotyped from each biopsy. The circles group phylogenetically related cells, with pathogenic mutations labeled in red. To the right of each schema, the corresponding phylogenetic trees, rooted in the germline state, are shown for all cells from that biopsy. **B.** The area occupied by individual clones was calculated from the size of each biopsy and the proportion of cells attributed to each clone. Clone areas are shown for keratinocytes and melanocytes with clones harboring pathogenic mutations indicated in red. **C-D.** Fraction of biopsies with a detectable clone (panel C) and fraction of clones with an underlying pathogenic mutation (panel D), separately plotted for keratinocytes (Ker.) and melanocytes (Mel.). * denotes $p<0.05$ (Poisson test).

Figure 3. The genetic evolution of a cutaneous squamous cell carcinoma from an actinic keratosis. A. H&E-stained section of a skin biopsy with adjacent areas of squamous cell carcinoma and actinic keratosis dissected, as indicated by the dashed lines. **B.** Scatter plot of mutant allele fractions in the squamous cell carcinoma and actinic keratosis reveal three clusters of mutations. **C.** The same scatterplot as shown in panel B with pathogenic mutations annotated. **D.** Copy number alterations were inferred over bins of the genome (columns) for each histologic area (rows) and are shown as a heatmap (red = gain, blue = loss, white = no change). No somatic gains or losses were observed. **E.** Major allele frequency – 0.5 (y-axis) for heterozygous SNPs across the genome (x-axis) show loss of heterozygosity over chromosome 9p. **F.** Phylogenetic tree rooted at the germline state. **G** and **H.** Immunostaining for p53 (panel G, brown stain) and phospho-MAPK (panel H, purple stain), show keratinocytes overexpressing p53 in both regions with increased phospho-MAPK in the squamous cell carcinoma.

Figure 4. The sequential order of genetic alterations during progression from actinic keratosis to squamous cell carcinoma. A. Phylogenetic trees, rooted in the germline state, summarize the evolution of four squamous cell carcinomas that evolved from actinic keratoses. See figure S7 for further details on these four cases and figure 3 for a summary of the example case. **B.** Eight squamous cell carcinomas that evolved from neighboring precursor lesions were identified as described. The stacked bar plot (top panel) indicates the proportion of mutations, recurrently mutated in these eight cases, in the trunk versus branch of phylogenetic trees. The bar plot (lower panel) indicates the number of cases with a mutation in each pathway. Mutations in the p53, Notch, TERT, and Rb pathways tended to occur early, contributing to the formation of actinic keratoses. Mutations affecting the SWI/SNF chromatin remodeling complex or activating the MAPK/PI3K pathways tended to occur later, driving the transition to squamous cell carcinoma. **C.** The frequency of mutations in select driver genes in normal skin biopsies versus squamous cell carcinoma. Error bars show 95% confidence intervals (Poisson test) with a y=x line included for orientation.

Figure 5. Spatial heterogeneity in gene expression of immune cells at the interface of squamous cell carcinoma versus actinic keratosis. Each column of images shows a different view of spatial transcriptomic data from case BB05, including: an H&E overview, annotated spots, and gene expression of immune checkpoints and their ligands. See figure S12 for an overview of other cases. Gene expression intensities represent the combined expression of the checkpoint or ligand genes listed. Zoomed insets show the interface of tumor epithelia and immune cells, illustrating different levels of checkpoint and ligand expression in squamous cell carcinoma versus actinic keratosis. Dotted lines indicate the tumor/immune boundary.

Figure 6. Summary of key events that occur during the evolution of cutaneous squamous cell carcinoma. After continual exposures to UV radiation, fibroblasts modestly increase their mutation burdens, melanocytes sharply increase their mutation burdens, and keratinocytes have a mixed response. Most keratinocytes accumulate little mutational damage, but a subset with pathogenic mutations build up mutations more rapidly than other skin cells. UV radiation induces expansion of independent clones of keratinocytes, often in close proximity and admixed, resulting in a complex clonal structure whereby adjacent lesions are not necessarily related. Driver mutations undergo selection in a stereotypical order, linked to histologic and genetic changes that occur during tumor evolution. An immune response builds during progression, but activity is blunted via engagement of immune checkpoints in squamous cell carcinoma.

Figure S1. An approach to measure the mutational landscapes of individual skin cells, and a summary of cells genotyped in this study. A. An overview of our single-cell genotyping workflow. **B.** Summary of donors, biopsies, and cell types included in this study. **C.** The lineage of each cell was confirmed from gene expression data of each colony. A t-SNE plot where each data point corresponds to a cell. Cells with similar gene expression profiles clustered together and are colored based on their morphological features (appearing as either keratinocytes, melanocytes, or fibroblasts) with a representative image from each group shown above the plot. **D.** Differential gene expression analysis was performed, comparing the three groups of cells in panel C. The top 15 upregulated genes in each group are shown in the heatmap. Each column corresponds to a cell and each row to a gene, as indicated, with red/blue tiles depicting genes that are relatively up/down-regulated. The top genes in each group are consistent with genes known to be expressed in either keratinocytes, melanocytes, or fibroblasts.

Figure S2. Melanocytes and fibroblasts have distinct mutational landscapes from keratinocytes. A. Mutation burdens (mutations/megabase) of individual keratinocytes, melanocytes, or fibroblasts, stratified by their anatomic location. Black bars denote median values. **B.** Top panel: mutation burden of melanocytes (left panel) and fibroblasts (right panel). Each column represents an individual cell, arranged in descending order of mutation burdens. Red bars indicate the presence of a pathogenic mutation(s) in that cell. Middle panel: tiling plot of pathogenic mutations (rows) in each cell (columns) with red tiles for gain of function mutation(s), light blue tiles for missense mutation(s), dark blue tiles for truncating (nonsense/splice-site/frameshift/indel) mutation(s) and a white circle for cells with multiple hits in the same gene. Bottom panel: the fractions of different mutational signatures attributed to each cell. White bars indicate cells with too few mutations to perform signature analysis.

Figure S3. The tri-nucleotide context of single nucleotide point mutations in skin cells. 96-barplots show the frequency of the six potential somatic point mutations in all possible tri-nucleotide contexts. Panel A includes signature 7 as a reference as well as the cumulative frequencies of mutations, aggregated from individual cells, of different skin cell types. Panel B shows the types of mutations in keratinocytes with and without underlying pathogenic mutations, aggregated from individual cells, and the differences between these groups.

Figure S4. Mutational landscape of individual keratinocytes compared to adjoining skin biopsy. A. The schematic illustrates the workflow for comparing the mutational profiles of keratinocytes derived from single-cell (Fig. S1A) and bulk sequencing approaches, using adjacent skin biopsies from the face and shoulder of the same donor (D56). **B.** Mutation burdens (mutations per megabase) of individual

keratinocytes from the face and shoulder are compared with those of adjacent skin biopsies. The mutation burden for skin biopsies is inferred from bulk sequencing data, as described by Martincorena *et al.* (Science, 2015). Error bars for skin biopsy mutation burdens represent 95% confidence intervals (Poisson test). For keratinocytes, each dot represents a single cell, and the horizontal bar indicates the mean mutation burden. **C.** Stacked bar plots compare the mutation signatures of individual keratinocytes from face (n = 15) with the adjoining face skin biopsy and shoulder keratinocytes (n = 13) with the adjoining shoulder skin biopsy. Mutation signatures shown for individual keratinocytes are inferred from a compilation of all mutations in keratinocytes sequenced from the face and shoulder, respectively.

Figure S5. Clonal architecture of keratinocytes in human skin- the full sample set. A-D. In figure 2A, we show the clonal structure of keratinocytes from a representative set of four biopsies. Here, we show all biopsies with more than 1 keratinocyte sampled at greater than 0.05 cells per square millimeter. Each schematic is plotted as described in figure 2A. Biopsies in each panel are grouped by their sampling densities, as indicated (most dense in panel A to least dense in panel D), with appropriate scale bars for each group.

Figure S6. Examples of squamous cell carcinomas that did not evolve from the neighboring actinic keratoses. On the lefthand side of each panel, we show overview images and zoomed insets of H&E stained biopsies with histologically distinct areas. From each biopsy, we macrodissected areas with actinic keratosis histology or squamous cell carcinoma histology, as indicated. On the righthand side of each panel, point mutations are stratified by their mutant allele fraction in each area with pathogenic mutations labeled. **A.** A squamous cell carcinoma that does not share mutations with the neighboring actinic keratosis. This pattern suggests that the two neoplasms are not related. **B.** A squamous cell carcinoma that shares mutations with adjacent tissue, but the squamous cell carcinoma lacks private mutations. This pattern implies that the adjacent tissue is a continuation of the squamous cell carcinoma, despite differences in their histologic appearance. **C.** A complex case with multiple clones of keratinocytes whose phylogenetic relationship cannot be resolved.

Figure S7. Examples of squamous cell carcinomas that evolved from neighboring actinic keratoses. On the lefthand side of each panel, we show overview images of H&E stained biopsies with histologically distinct areas, macrodissected as indicated. On the righthand side of each panel, point mutations are stratified by their mutant allele fraction in each area. In the first version of the scatterplot, mutations were classified as “shared” or “exclusive to the SCC” (see methods). A second version of the scatterplot is shown with pathogenic mutations labeled.

Figure S8. Examples of squamous cell carcinomas, from publicly available data, that do not share mutations with neighboring tissue. Kim *et al. JID*, 2022 sequenced squamous cell carcinomas (SCCs) and adjacent tissue. The adjacent tissue was classified as one of the following: actinic keratosis (AK), squamous cell carcinoma in situ (SCCIS), or sun-exposed skin (SE). Mutations were plotted by their allele frequencies in the adjacent tissue (x-axis) versus the squamous cell carcinoma (y-axis). Phylogenetic trees from the original study were reinterpreted by our group (see methods). In this series of cases, we do not believe that the dominant clone in the adjacent skin (x-axis) shares mutations with the dominant clone in the squamous cell carcinoma (y-axis). The mutations that were assigned to the trunks of trees in the original study tended to have high allele frequencies in one tissue but only trace sequencing reads in the other tissue, which likely stemmed from minor levels of contamination. Taken together, we conclude that the dominant clones in the squamous cell carcinomas were unrelated to the dominant clones in the neighboring tissue, and therefore, these clones arose independently, despite their proximity.

Figure S9. Additional examples of squamous cell carcinomas from publicly available data, some of which evolved from neighboring precursor lesions. Kim *et al. JID*, 2022 sequenced squamous cell carcinomas (SCCs) and adjacent tissue. The adjacent tissue was classified as one of the following: actinic keratosis (AK), squamous cell carcinoma in situ (SCCIS), or sun-exposed skin (SE). Mutations were plotted by their allele frequencies in the adjacent tissue (x-axis) versus the squamous cell carcinoma (y-axis). Phylogenetic trees from the original study were reinterpreted by our group (see methods). **A.** Cases where

the SCC are mutationally indistinguishable from neighboring tissue. In these cases, we did not find compelling evidence for branch mutations, private to either the precursor or descendent lesions. While there are mutations exclusively found in the SCCIS or SCC in cases 3 and 4, their allele frequencies were much lower than the mutations in the dominant clones in each tissue; therefore, we could not rule out the possibility that they came from an unrelated clone of cells, contaminating the macrodissected tissue. Taken together, we conclude the SCC to be mutationally indistinguishable from the dominant clone in the adjacent skin for these three cases. **B.** Cases where the SCC evolved from the neighboring tissue. In cases 8, 9 and 10, there were shared mutations between the two tissues, suggesting that the squamous cell carcinomas were related to the neighboring tissues. There were also mutations exclusive to the squamous cell carcinomas, suggesting that they underwent an additional wave of clonal expansion. Taken together, we generally agree with the authors original interpretation for these cases that these squamous cell carcinomas evolved from neighboring precursor lesions.

Figure S10. Copy number alterations are detectable in spatial transcriptomic data. A. Copy number alterations (CNAs) were inferred from DNA-Sequencing data (top heatmap) and from RNA-sequencing data of individual spots (lower heatmap). Spots (rows in the lower heatmap) are ranked ordered by the similarity of their copy number profiles to the DNA copy number alterations. **B.** For each spot in panel A, we calculated a CNVscore (see Chen *et al. Genome Biology*, 2023) to capture the similarity of the spot's copy number profile to the copy number inferred from DNA-seq data. Histogram of CNVscores from permuted and observed data are shown. A quantile-quantile plot was generated, comparing observed CNVscores to permuted scores, from which we established q-values to determine which spots were "Likely", "Possibly", or "Not" tumor, as indicated. **C.** The spatial distribution of spots as defined in panel B.

Figure S11. Spatial heterogeneity in gene expression during the evolution of squamous cell carcinoma from actinic keratosis. A. A t-SNE plot of all spots from spatial transcriptomics data colored by sample ID. Spots clustered primarily by cell type or cell state, suggesting that batch effects were minimal. We performed spatial transcriptomics on a fifth case (BB13), but it was not included in this plot because it was profiled with an older version of Visium, whose design did not permit aggregate analyses with samples profiled on newer versions of the platform. **B-D.** t-SNE plots of spots, aggregated from spatial transcriptomic data of four squamous cell carcinomas in association with actinic keratosis. Clusters of main cell types are annotated, as described, with the most highly expressed genes in each group highlighted in heatmaps to the right. **E.** Zoomed view of the junction between squamous cell carcinoma and actinic keratosis in BB05. H&E, copy number, and gene expression clusters are shown, as indicated. Note that keratinocytes in both the squamous cell carcinoma and actinic keratosis occupy a spectrum of differentiation states.

Figure S12. Spatial heterogeneity in gene expression of immune cells at the interface of squamous cell carcinoma versus actinic keratosis -- all cases. Each column of images shows a different view of spatial transcriptomic data, including: annotated spots and gene expression of immune checkpoints and their ligands. See figure 5 for an overview of the BB05. Gene expression intensities represent the combined expression of the checkpoint or ligand genes listed.

Supplementary information

Table S1. A summary of clinical, cytologic, genetic, and quality control metrics for individual skin cells sequenced in this study.

Table S2. List of point mutations in individual skin cells genotyped in this study.

Table S3. A summary of clinical, histopathologic, genetic, and quality control metrics for each area of the cutaneous squamous cell carcinomas, in association with actinic keratoses, that were sequenced in this study.

Table S4. Genes and bait intervals targeted for capture when sequencing cutaneous squamous cell carcinomas, in association with actinic keratoses.

Table S5. List of point mutations detected in cutaneous squamous cell carcinomas, in association with actinic keratoses, that were genotyped in this study.

Table S6. List of point mutations detected in cutaneous squamous cell carcinomas and adjacent skin that were genotyped by Kim *et al*.(18) and reanalyzed here.

Figure 1.

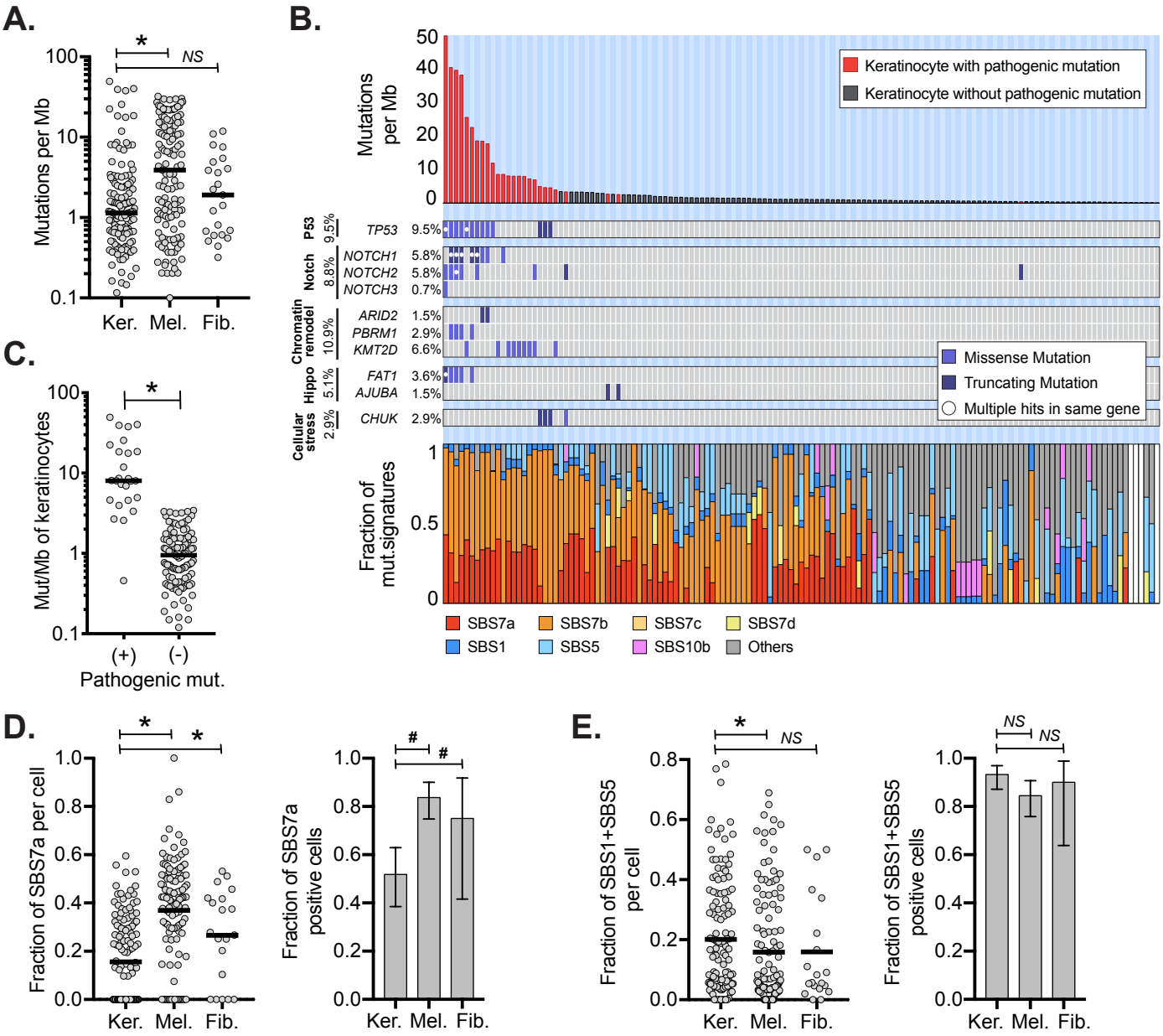


Figure 1. Keratinocytes have distinct mutational landscapes compared to other cell types. A. Mutation burdens (mutations/megabase) of individual keratinocytes (Ker.) compared to melanocytes (Mel.) and fibroblasts (Fib.) **B.** Mutation burden, driver mutations, and mutational signatures for 137 keratinocytes with each column of the three stacked panels representing an individual cell. Top panel: mutation burden of keratinocytes in descending order. Red bars indicate cells harboring one or more pathogenic mutations. Middle panel: tiling plot of pathogenic mutations (rows). Bottom panel: the fractions of different mutational signatures for each cell. White bars indicate keratinocytes with too few mutations to perform signature analysis. **C.** Mutation burdens of keratinocytes with and without pathogenic mutations. **D.** Left panel: fraction of mutations with UV signatures (SBS7a) in keratinocytes, melanocytes, or fibroblasts. Right panel: fraction of cells with detectable SBS7a in keratinocytes, melanocytes, or fibroblasts. **E.** The data is plotted as in panel D but for clock-like signatures (SBS1 and SBS5). For all plots, an asterisk (*) or a hash (#) denotes $p<0.05$ using the Wilcoxon rank-sum test (cell to cell comparisons) or the Poisson test (proportion comparisons) respectively. Horizontal bars show the median (panels A and C) or mean (panels D and E). Error bars in panels D and E show 95% confidence intervals (Poisson test).

Figure 2.

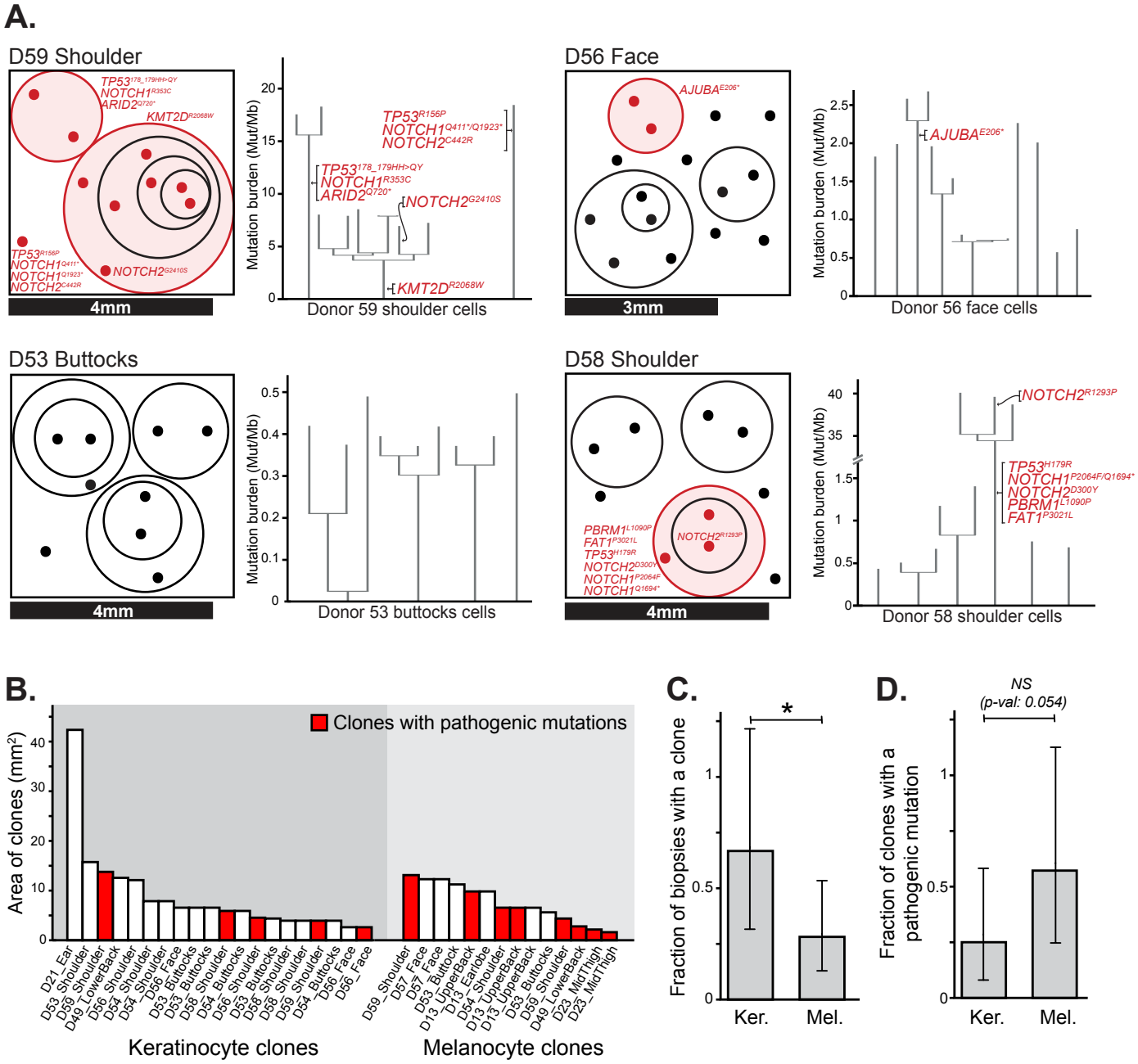


Figure 2. Clonal architecture of keratinocytes in human skin. A. Clonal structure of keratinocytes from four representative skin biopsies (see fig. S5 for all biopsies). The surface area of each biopsy is drawn to scale, as indicated, with dots representing the cells genotyped from each biopsy. The circles group phylogenetically related cells, with pathogenic mutations labeled in red. To the right of each schema, the corresponding phylogenetic trees, rooted in the germline state, are shown for all cells from that biopsy. **B.** The area occupied by individual clones was calculated from the size of each biopsy and the proportion of cells attributed to each clone. Clone areas are shown for keratinocytes and melanocytes with clones harboring pathogenic mutations indicated in red. **C-D.** Fraction of biopsies with a detectable clone (panel C) and fraction of clones with an underlying pathogenic mutation (panel D), separately plotted for keratinocytes (Ker.) and melanocytes (Mel.). * denotes $p<0.05$ (Poisson test).

Figure 3.

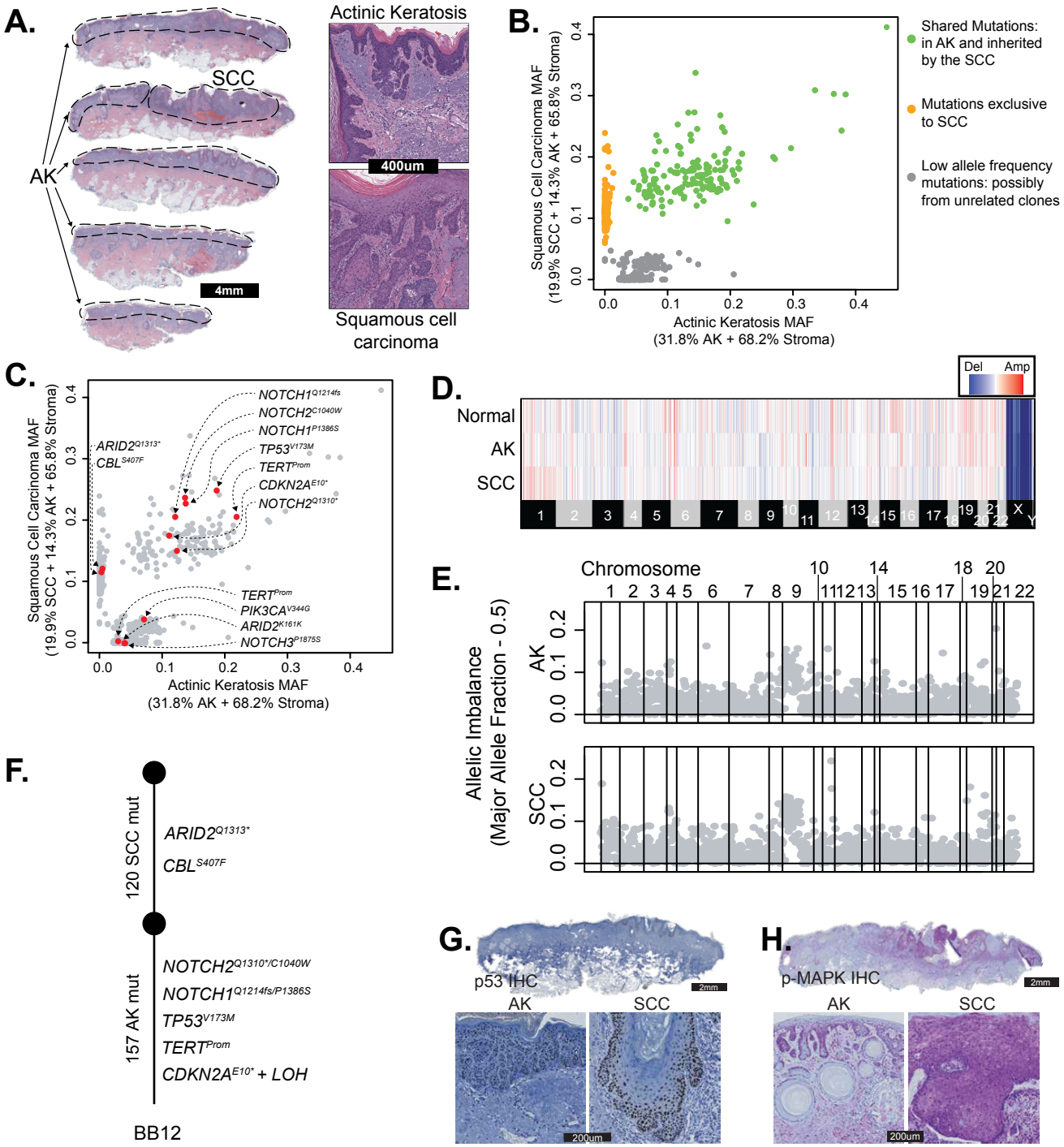


Figure 3. The genetic evolution of a cutaneous squamous cell carcinoma from an actinic keratosis. A. H&E-stained section of a skin biopsy with adjacent areas of squamous cell carcinoma and actinic keratosis dissected, as indicated by the dashed lines. **B.** Scatter plot of mutant allele fractions in the squamous cell carcinoma and actinic keratosis reveal three clusters of mutations. **C.** The same scatterplot as shown in panel B with pathogenic mutations annotated. **D.** Copy number alterations were inferred over bins of the genome (columns) for each histologic area (rows) and are shown as a heatmap (red = gain, blue = loss, white = no change). No somatic gains or losses were observed. **E.** Major allele frequency – 0.5 (y-axis) for heterozygous SNPs across the genome (x-axis) show loss of heterozygosity over chromosome 9p. **F.** Phylogenetic tree rooted at the germline state. **G** and **H.** Immunostaining for p53 (panel G, brown stain) and phospho-MAPK (panel H, purple stain), show keratinocytes overexpressing p53 in both regions with increased phospho-MAPK in the squamous cell carcinoma.

Figure 4.

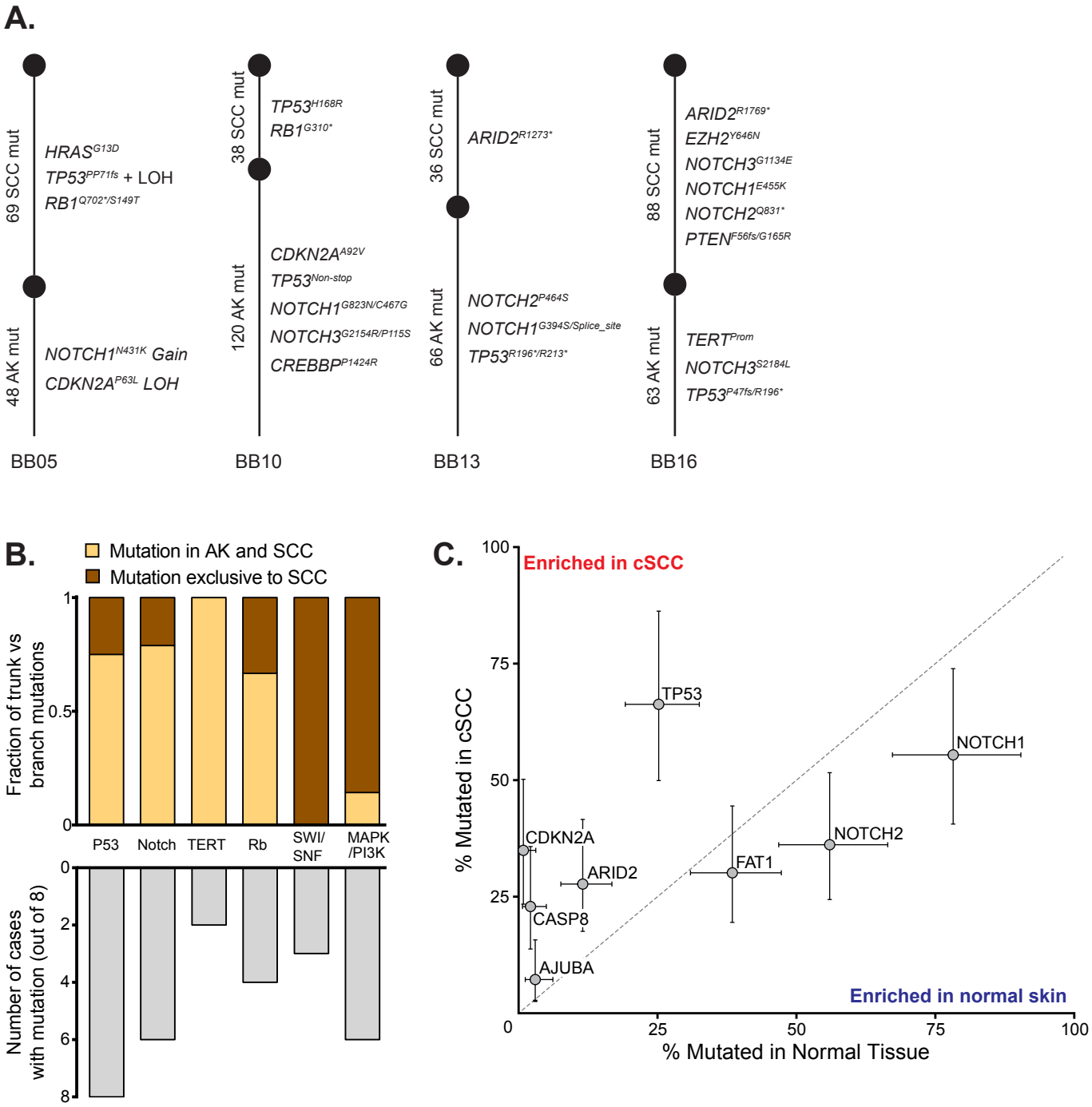


Figure 4. The sequential order of genetic alterations during progression from actinic keratosis to squamous cell carcinoma. A. Phylogenetic trees, rooted in the germline state, summarize the evolution of four squamous cell carcinomas that evolved from actinic keratoses. See figure S7 for further details on these four cases and figure 3 for a summary of the example case. **B.** Eight squamous cell carcinomas that evolved from neighboring precursor lesions were identified as described. The stacked bar plot (top panel) indicates the proportion of mutations, recurrently mutated in these eight cases, in the trunk versus branch of phylogenetic trees. The bar plot (lower panel) indicates the number of cases with a mutation in each pathway. Mutations in the p53, Notch, TERT, and Rb pathways tended to occur early, contributing to the formation of actinic keratoses. Mutations affecting the SWI/SNF chromatin remodeling complex or activating the MAPK/PI3K pathways tended to occur later, driving the transition to squamous cell carcinoma. **C.** The frequency of mutations in select driver genes in normal skin biopsies versus squamous cell carcinoma. Error bars show 95% confidence intervals (Poisson test) with a y=x line included for orientation.

Figure 5.

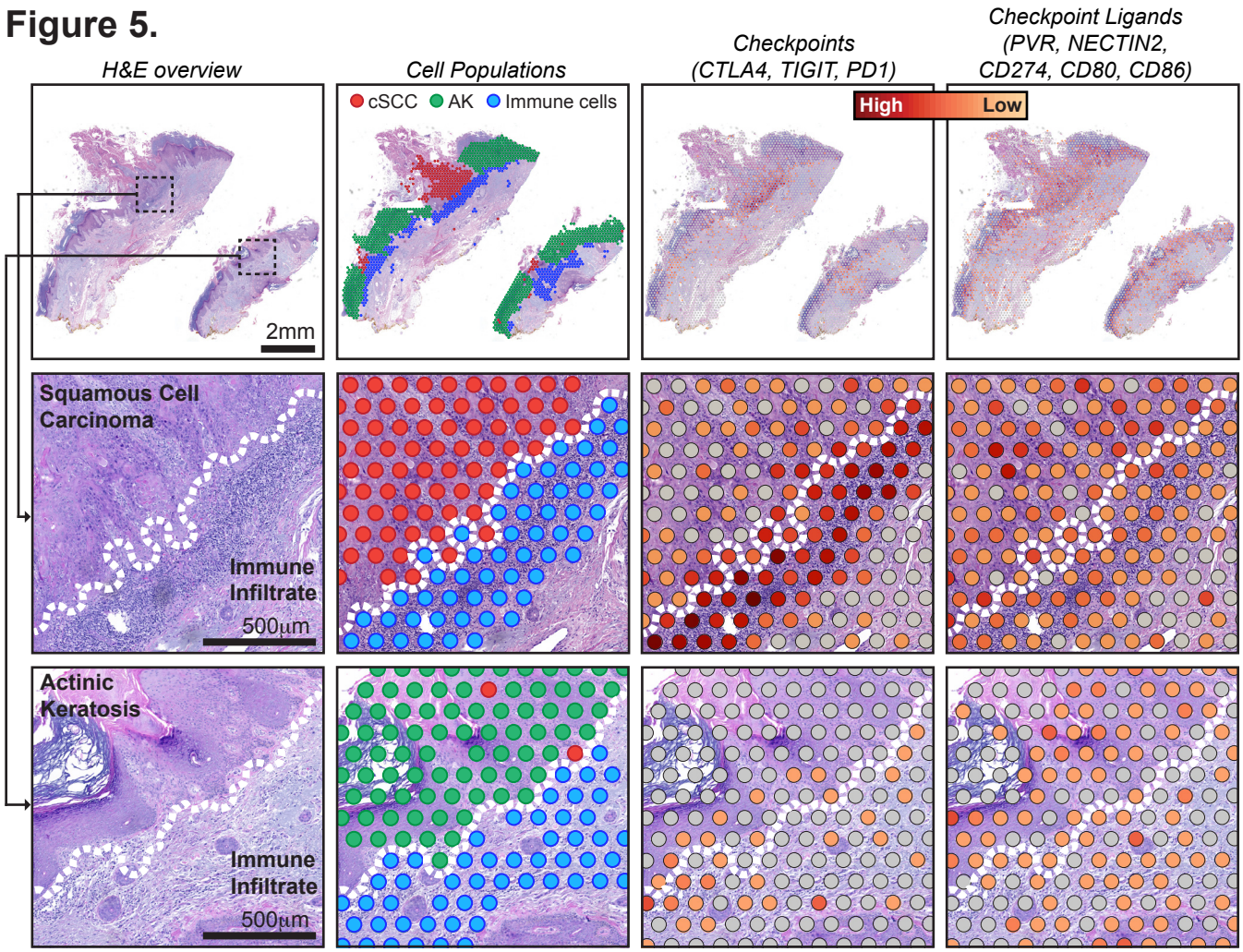


Figure 5. Spatial heterogeneity in gene expression of immune cells at the interface of squamous cell carcinoma versus actinic keratosis. Each column of images shows a different view of spatial transcriptomic data from case BB05, including: an H&E overview, annotated spots, and gene expression of immune checkpoints and their ligands. See figure S12 for an overview of other cases. Gene expression intensities represent the combined expression of the checkpoint or ligand genes listed. Zoomed insets show the interface of tumor epithelia and immune cells, illustrating different levels of checkpoint and ligand expression in squamous cell carcinoma versus actinic keratosis. Dotted lines indicate the tumor/immune boundary.

Figure 6.

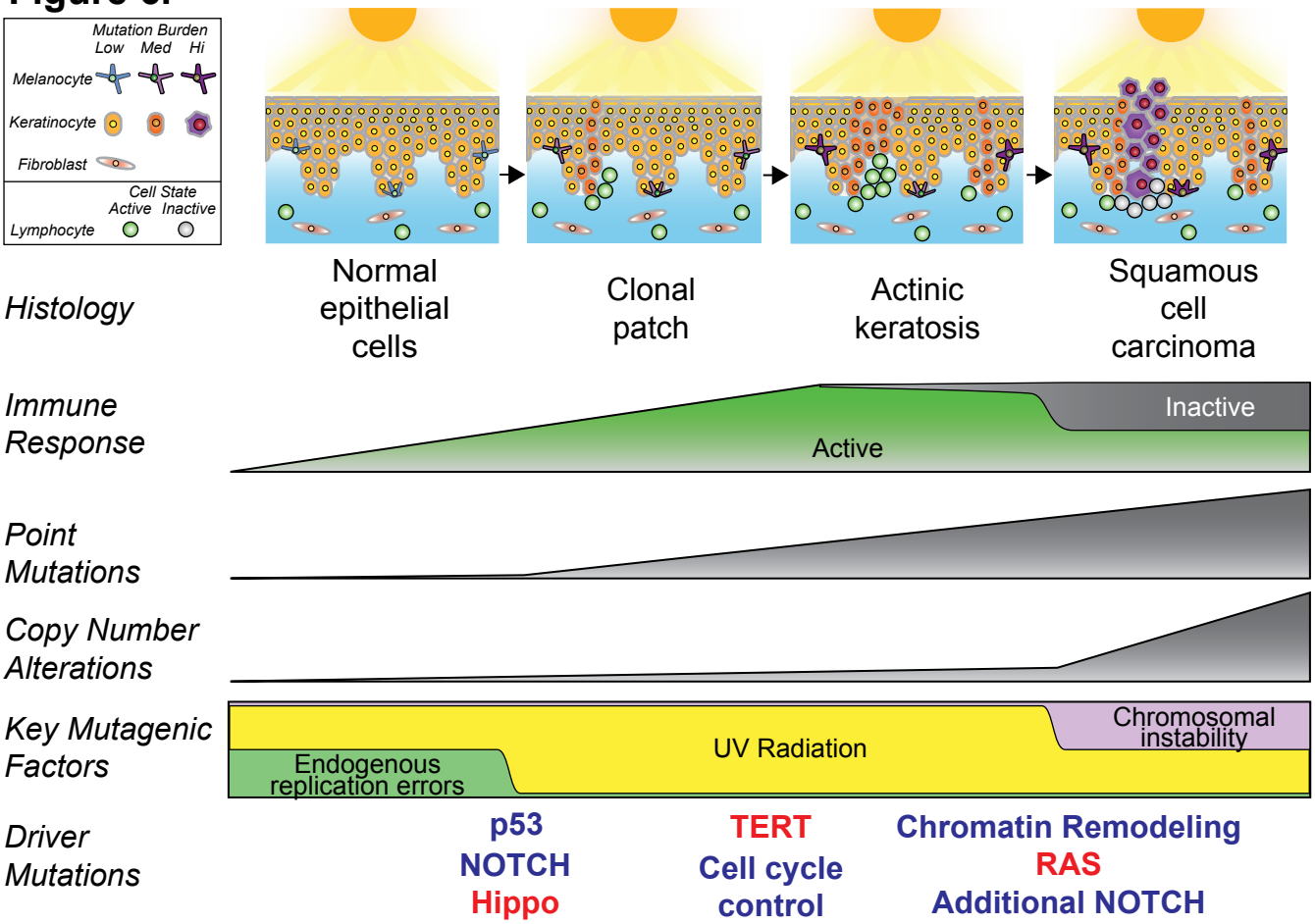
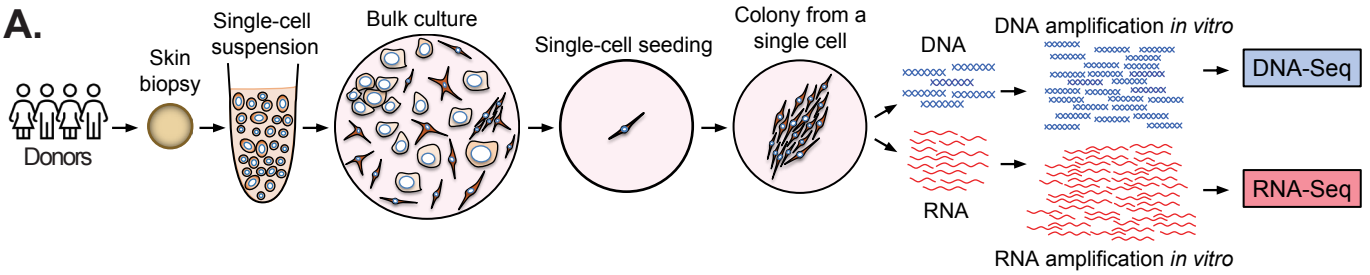


Figure 6. Summary of key events that occur during the evolution of cutaneous squamous cell carcinoma. After continual exposures to UV radiation, fibroblasts modestly increase their mutation burdens, melanocytes sharply increase their mutation burdens, and keratinocytes have a mixed response. Most keratinocytes accumulate little mutational damage, but a subset with pathogenic mutations build up mutations more rapidly than other skin cells. UV radiation induces expansion of independent clones of keratinocytes, often in close proximity and admixed, resulting in a complex clonal structure whereby adjacent lesions are not necessarily related. Driver mutations undergo selection in a stereotypical order, linked to histologic and genetic changes that occur during tumor evolution. An immune response builds during progression, but activity is blunted via engagement of immune checkpoints in squamous cell carcinoma.

Figure S1.

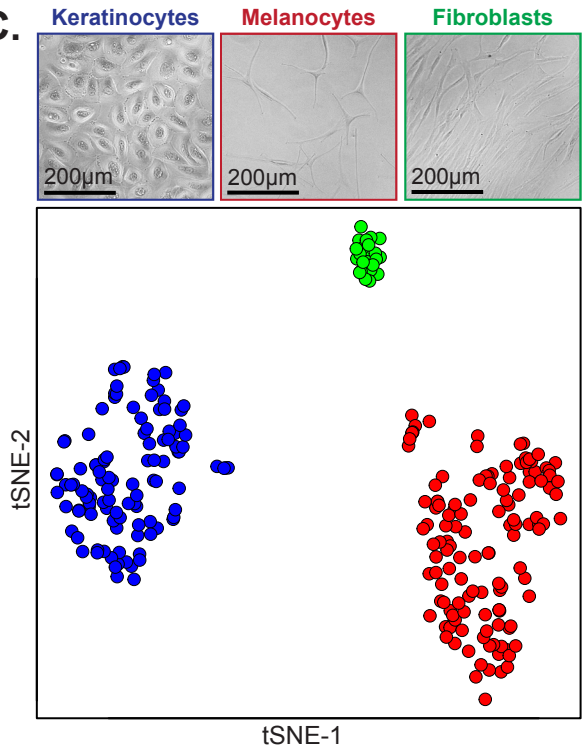
A.



B.

Donor 6, Male, 64 yrs *L. Upper Back, R. Back* - 3 keratinocytes - 29 melanocytes - 2 fibroblasts	**Donor 21, Male, 57 yrs** *L. Ear* - 4 keratinocytes - 5 melanocytes - 0 fibroblasts	**Donor 49, Female, 52 yrs** *R. Lower Back* - 8 keratinocytes - 9 melanocytes - 2 fibroblasts	**Donor 54, Male, 42 yrs** *Shoulder, Buttocks* - 20 keratinocytes - 9 melanocytes - 3 fibroblasts	**Donor 57, Female, 70 yrs** *Face* - 5 keratinocytes - 10 melanocytes - 3 fibroblasts
Donor 13, Male, 74 yrs *Scalp, R. Earlobe, Upper Back* - 4 keratinocytes - 30 melanocytes - 0 fibroblasts	**Donor 37, Female, 43 yrs** *R. Lower Back* - 4 keratinocytes - 2 melanocytes - 0 fibroblasts	**Donor 51, Female, 95 yrs** *Scalp* - 2 keratinocytes - 0 melanocytes - 0 fibroblasts	**Donor 55, Male, 70 yrs** *Shoulder, Buttocks* - 14 keratinocytes - 7 melanocytes - 3 fibroblasts	**Donor 58, Female, 75 yrs** *Shoulder* - 10 keratinocytes - 3 melanocytes - 2 fibroblasts
Donor 14, Female, 35 yrs *Abdomen* - 2 keratinocytes - 2 melanocytes - 0 fibroblasts	**Donor 38, Female, 40 yrs** *R. Lower Back* - 4 keratinocytes - 6 melanocytes - 1 fibroblast	**Donor 53, Female, 50 yrs** *Shoulder, Buttocks* - 19 keratinocytes - 7 melanocytes - 3 fibroblasts	**Donor 56, Male, 74 yrs** *Face, Shoulder* - 28 keratinocytes - 3 melanocytes - 3 fibroblasts	**Donor 59, Female, 56 yrs** *Shoulder* - 10 keratinocytes - 9 melanocytes - 1 fibroblasts

C.



D.

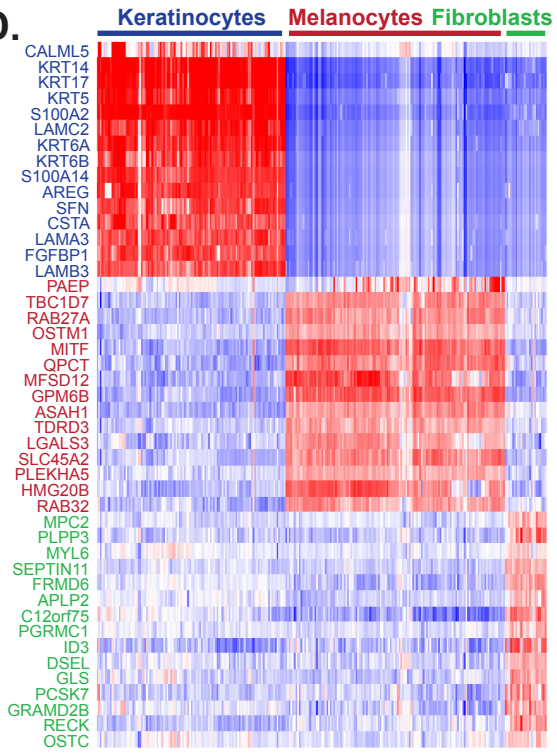


Figure S1. An approach to measure the mutational landscapes of individual skin cells, and a summary of cells genotyped in this study. A. An overview of our single-cell genotyping workflow. **B.** Summary of donors, biopsies, and cell types included in this study. **C.** The lineage of each cell was confirmed from gene expression data of each colony. A t-SNE plot where each data point corresponds to a cell. Cells with similar gene expression profiles clustered together and are colored based on their morphological features (appearing as either keratinocytes, melanocytes, or fibroblasts) with a representative image from each group shown above the plot. **D.** Differential gene expression analysis was performed, comparing the three groups of cells in panel C. The top 15 upregulated genes in each group are shown in the heatmap. Each column corresponds to a cell and each row to a gene, as indicated, with red/blue tiles depicting genes that are relatively up/down-regulated. The top genes in each group are consistent with genes known to be expressed in either keratinocytes, melanocytes, or fibroblasts.

Figure S2.

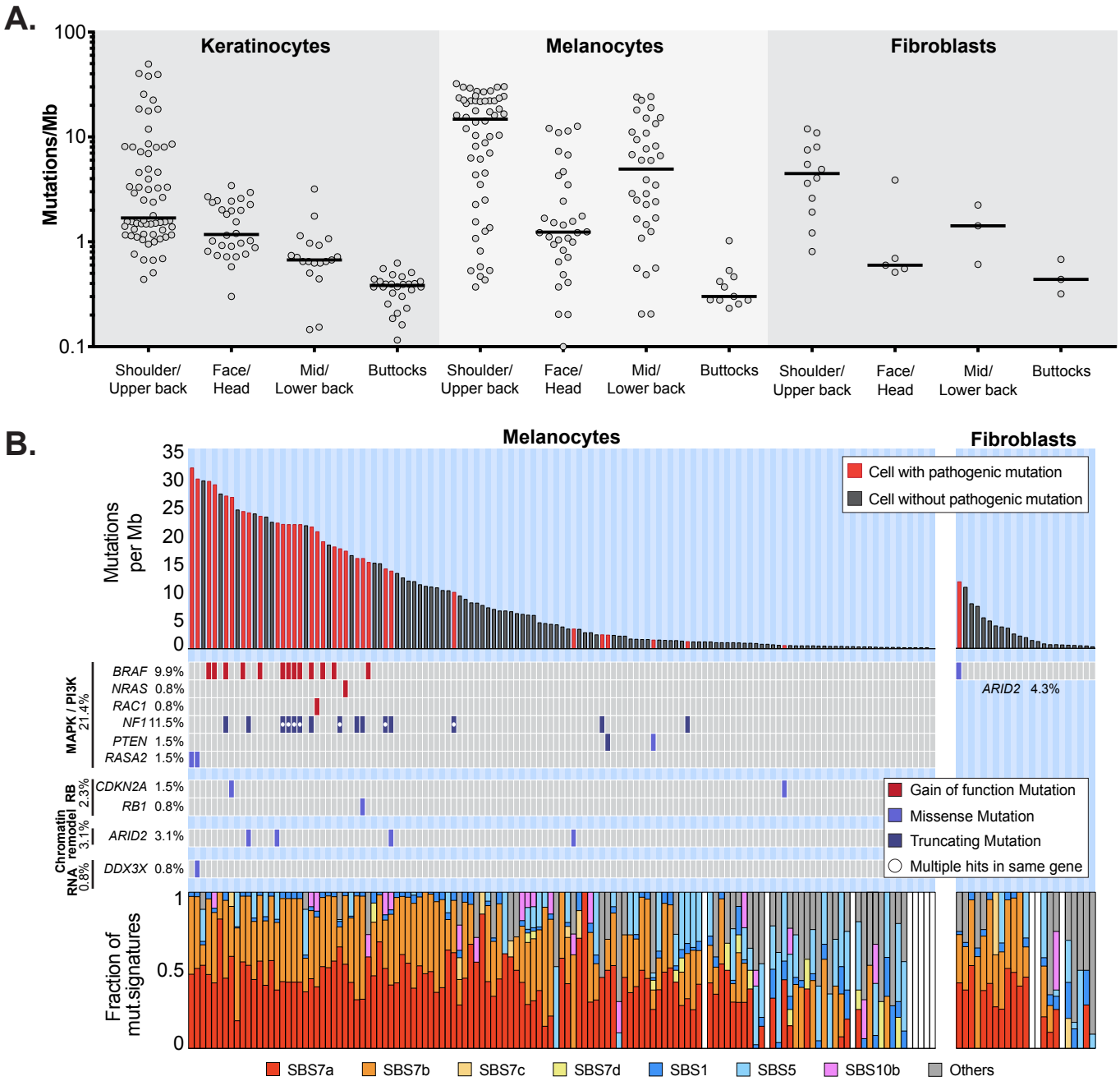


Figure S2. Melanocytes and fibroblasts have distinct mutational landscapes from keratinocytes. A. Mutation burdens (mutations/megabase) of individual keratinocytes, melanocytes, or fibroblasts, stratified by their anatomic location. Black bars denote median values. **B.** Top panel: mutation burden of melanocytes (left panel) and fibroblasts (right panel). Each column represents an individual cell, arranged in descending order of mutation burdens. Red bars indicate the presence of a pathogenic mutation(s) in that cell. Middle panel: tiling plot of pathogenic mutations (rows) in each cell (columns) with red tiles for gain of function mutation(s), light blue tiles for missense mutation(s), dark blue tiles for truncating (nonsense/splice-site/frameshift/indel) mutation(s) and a white circle for cells with multiple hits in the same gene. Bottom panel: the fractions of different mutational signatures attributed to each cell. White bars indicate cells with too few mutations to perform signature analysis.

Figure S3.

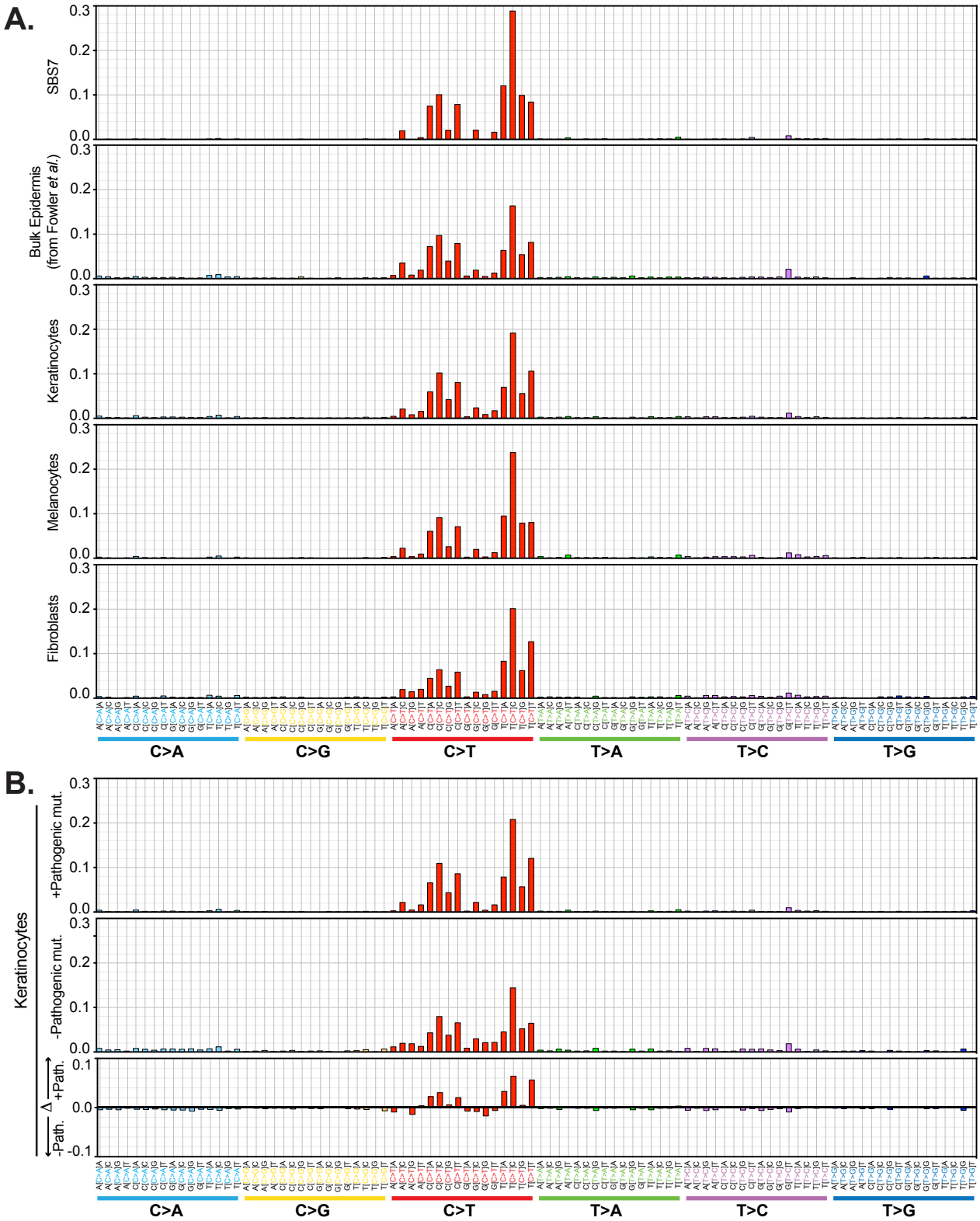


Figure S3. The tri-nucleotide context of single nucleotide point mutations in skin cells. 96-barplots show the frequency of the six potential somatic point mutations in all possible tri-nucleotide contexts. Panel A includes signature 7 as a reference as well as the cumulative frequencies of mutations, aggregated from normal skin biopsies (Fowler *et al.* 2021), and individual cells, of different skin cell types. Panel B shows the types of mutations in keratinocytes with and without underlying pathogenic mutations, aggregated from individual cells, and the differences between these groups.

Figure S4.

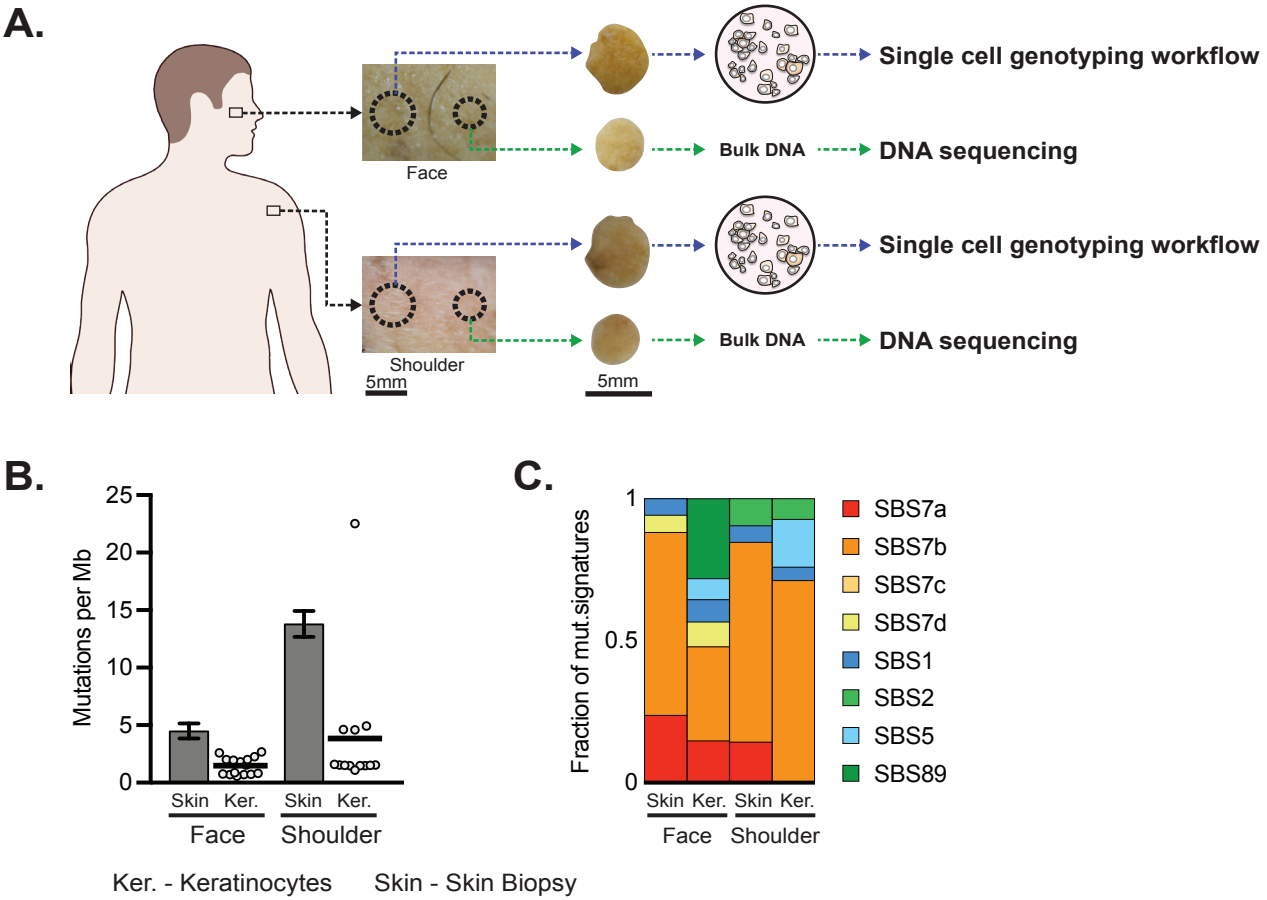


Figure S4. Mutation landscape of individual keratinocytes compared to adjoining skin biopsy. **A.** The schematic illustrates the workflow for comparing the mutational profiles of keratinocytes derived from single-cell (Fig. S1A) and bulk sequencing approaches, using adjacent skin biopsies from the face and shoulder of the same donor (D56). **B.** Mutation burdens (mutations per megabase) of individual keratinocytes from the face and shoulder are compared with those of adjacent skin biopsies. The mutation burden for skin biopsies is inferred from bulk sequencing data, as described by Martincorena *et al.* (Science, 2015). Error bars for skin biopsy mutation burdens represent 95% confidence intervals (Poisson test). For keratinocytes, each dot represents a single cell, and the horizontal bar indicates the mean mutation burden. **C.** Stacked bar plots compare the mutation signatures of individual keratinocytes from face (n = 15) with the adjoining face skin biopsy and shoulder keratinocytes (n = 13) with the adjoining shoulder skin biopsy. Mutation signatures shown for individual keratinocytes are inferred from a compilation of all mutations in keratinocytes sequenced from the face and shoulder, respectively.

Figure S5.

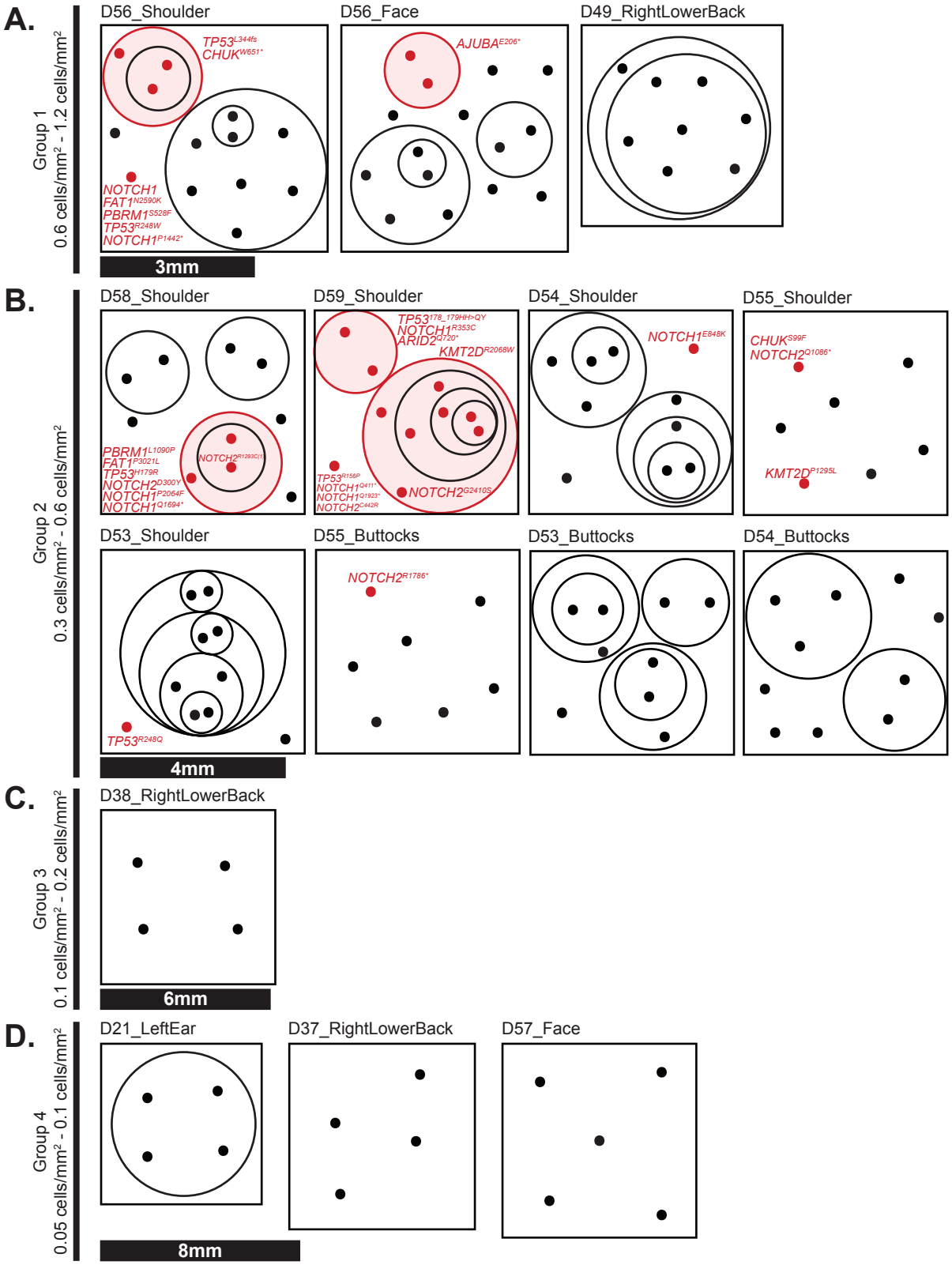


Figure S5. Clonal architecture of keratinocytes in human skin – the full sample set. A-D. In figure 2A, we show the clonal structure of keratinocytes from a representative set of four biopsies. Here, we show all biopsies with more than 1 keratinocyte sampled at greater than 0.05 cells per square millimeter. Each schematic is plotted as described in figure 2A. Biopsies in each panel are grouped by their sampling densities, as indicated (most dense in panel A to least dense in panel D), with appropriate scale bars for each group.

Figure S6.

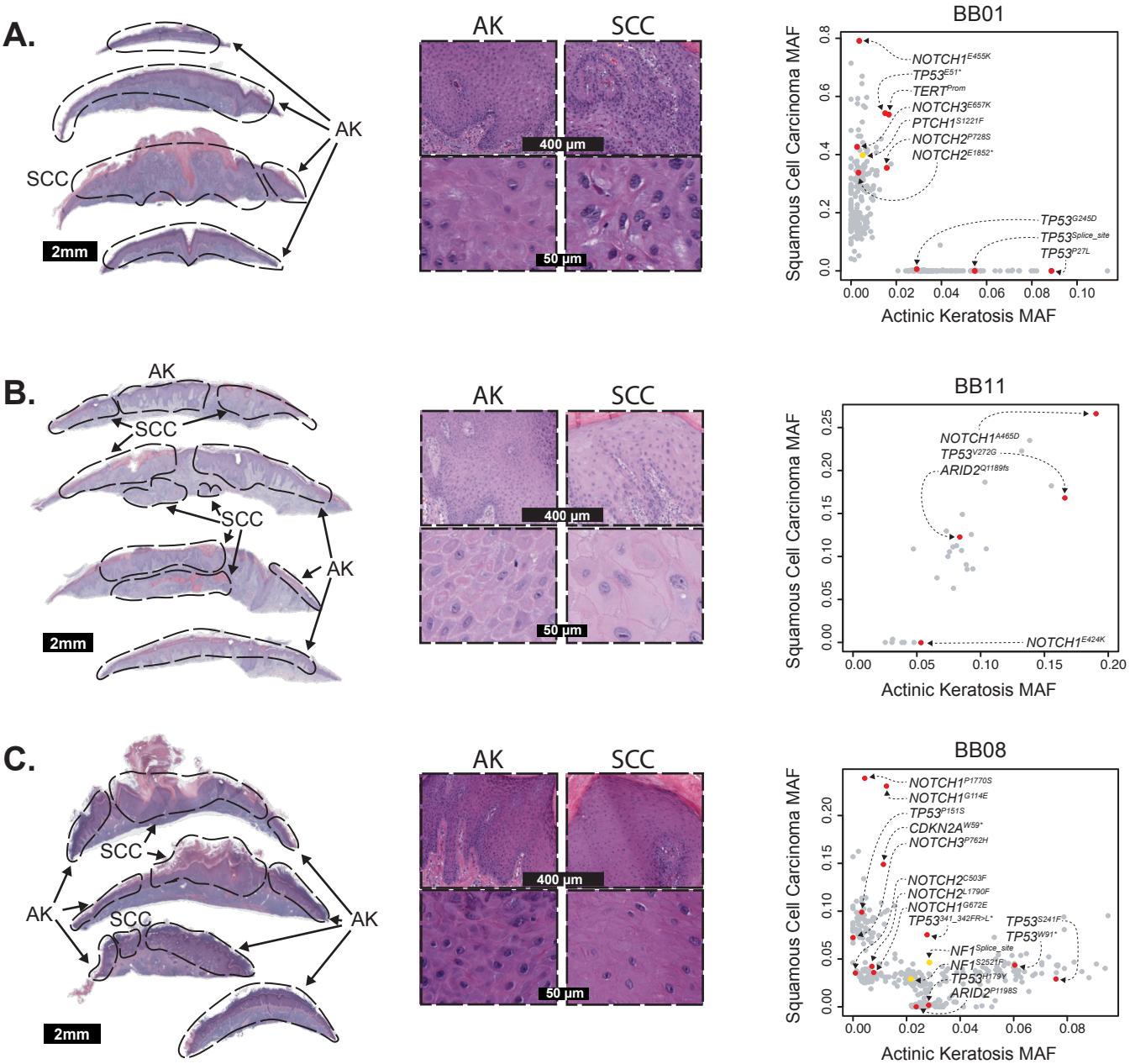


Figure S6. Examples of squamous cell carcinomas that did not evolve from the neighboring actinic keratoses. On the lefthand side of each panel, we show overview images and zoomed insets of H&E stained biopsies with histologically distinct areas. From each biopsy, we macrodissected areas with actinic keratosis histology or squamous cell carcinoma histology, as indicated. On the righthand side of each panel, point mutations are stratified by their mutant allele fraction in each area with pathogenic mutations labeled. **A.** A squamous cell carcinoma that does not share mutations with the neighboring actinic keratosis. This pattern suggests that the two neoplasms are not related. **B.** A squamous cell carcinoma that shares mutations with adjacent tissue, but the squamous cell carcinoma lacks private mutations. This pattern implies that the adjacent tissue is a continuation of the squamous cell carcinoma, despite differences in their histologic appearance. **C.** A complex case with multiple clones of keratinocytes whose phylogenetic relationship cannot be resolved.

Figure S7.

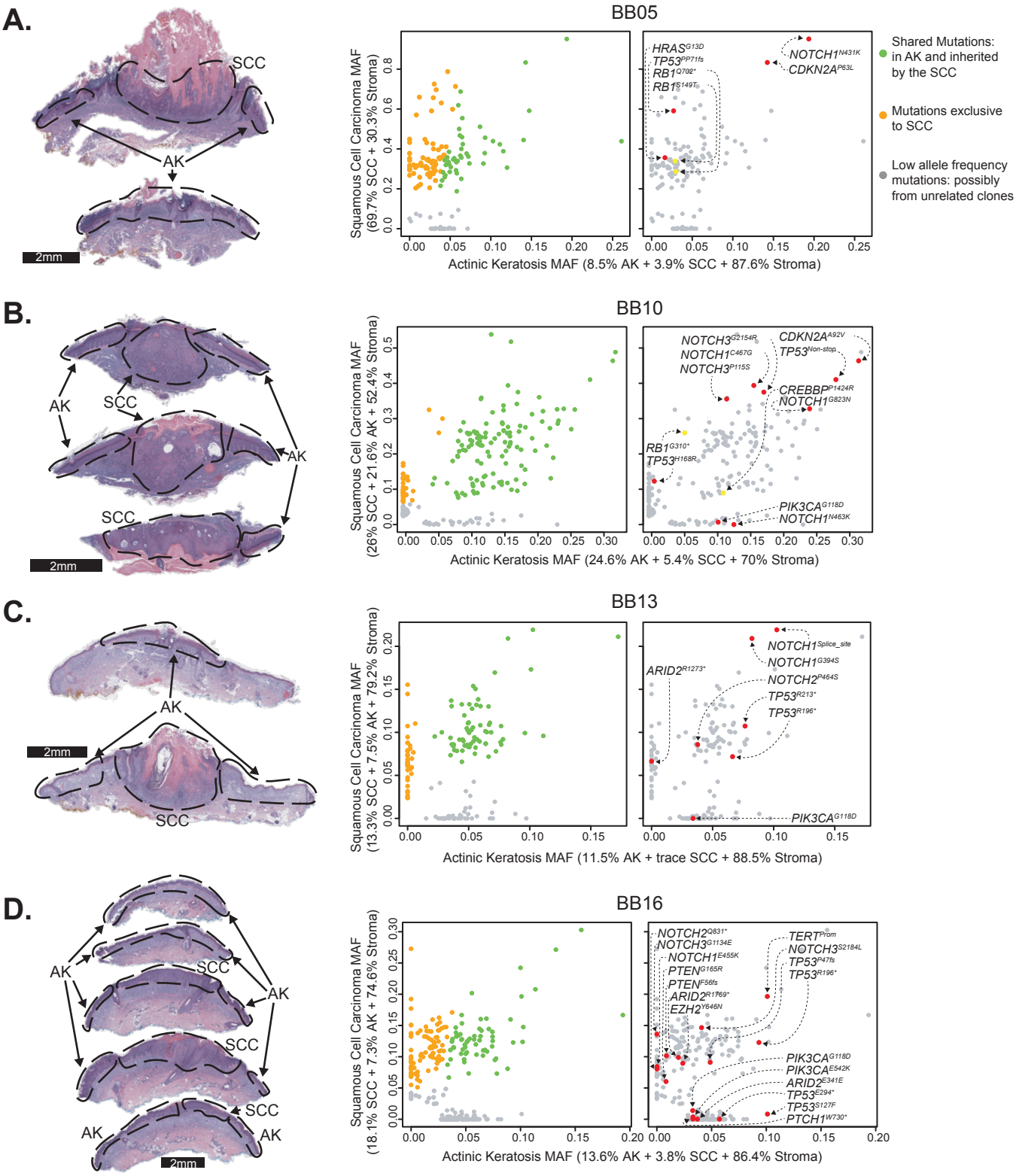


Figure S7. Examples of squamous cell carcinomas that evolved from neighboring actinic keratoses. On the lefthand side of each panel, we show overview images of H&E stained biopsies with histologically distinct areas, macrodissected as indicated. On the righthand side of each panel, point mutations are stratified by their mutant allele fraction in each area. In the first version of the scatterplot, mutations were classified as “shared” or “exclusive to the SCC” (see methods). A second version of the scatterplot is shown with pathogenic mutations labeled.

Figure S8.

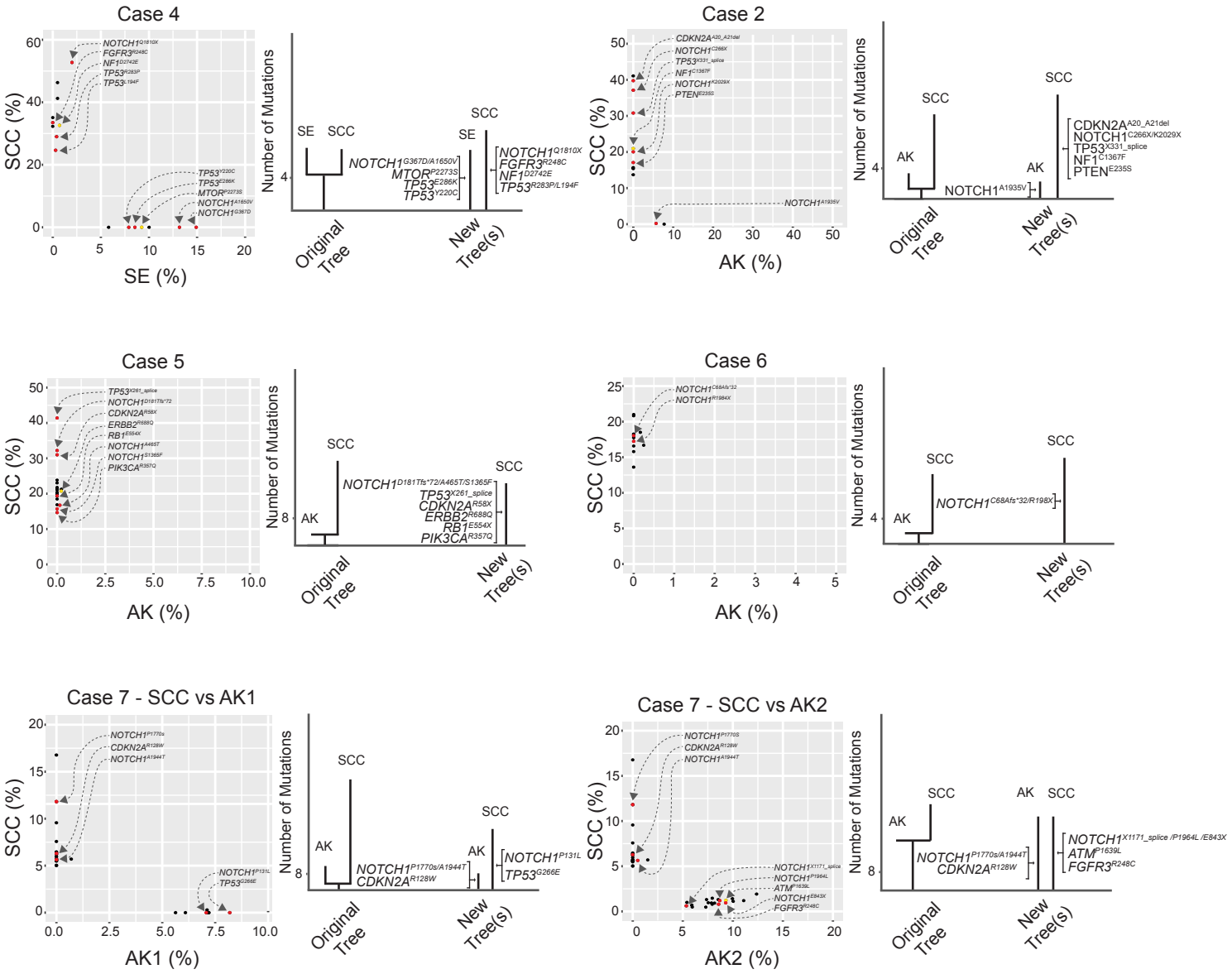


Figure S8. Examples of squamous cell carcinomas, from publicly available data, that do not share mutations with neighboring tissue. Kim *et al. JID*, 2022 sequenced squamous cell carcinomas (SCCs) and adjacent tissue. The adjacent tissue was classified as one of the following: actinic keratosis (AK), squamous cell carcinoma in situ (SCCIS), or sun-exposed skin (SE). Mutations were plotted by their allele frequencies in the adjacent tissue (x-axis) versus the squamous cell carcinoma (y-axis). Phylogenetic trees from the original study were reinterpreted by our group (see methods). In this series of cases, we do not believe that the dominant clone in the adjacent skin (x-axis) shares mutations with the dominant clone in the squamous cell carcinoma (y-axis). The mutations that were assigned to the trunks of trees in the original study tended to have high allele frequencies in one tissue but only trace sequencing reads in the other tissue, which likely stemmed from minor levels of contamination. Taken together, we conclude that the dominant clones in the squamous cell carcinomas were unrelated to the dominant clones in the neighboring tissue, and therefore, these clones arose independently, despite their proximity.

Figure S9.

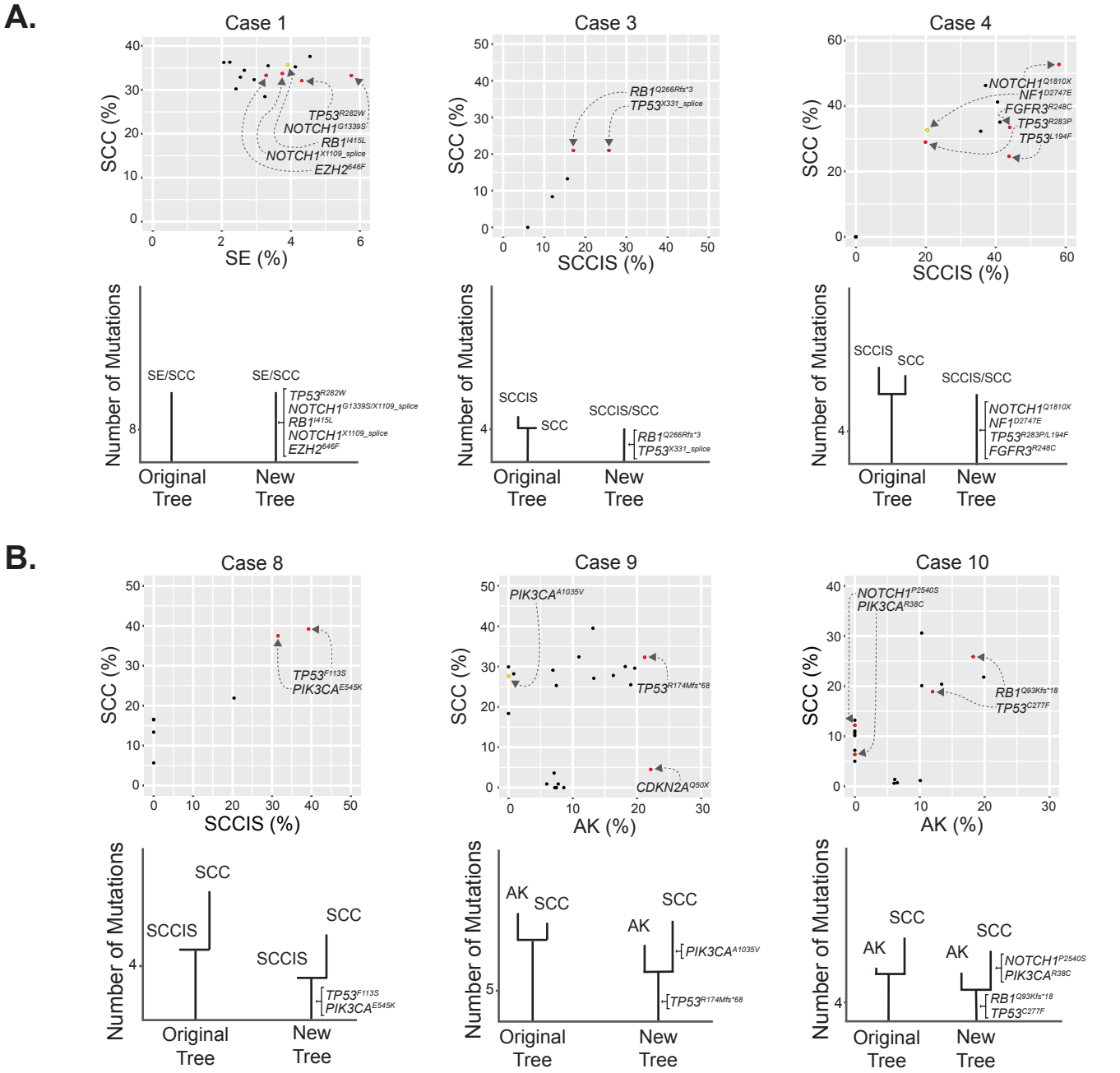


Figure S9. Additional examples of squamous cell carcinomas from publicly available data, some of which evolved from neighboring precursor lesions. Kim *et al. JID*, 2022 sequenced squamous cell carcinomas (SCCs) and adjacent tissue. The adjacent tissue was classified as one of the following: actinic keratosis (AK), squamous cell carcinoma in situ (SCCIS), or sun-exposed skin (SE). Mutations were plotted by their allele frequencies in the adjacent tissue (x-axis) versus the squamous cell carcinoma (y-axis). Phylogenetic trees from the original study were reinterpreted by our group (see methods). **A.** Cases where the SCC are mutationally indistinguishable from neighboring tissue. In these cases, we did not find compelling evidence for branch mutations, private to either the precursor or descendent lesions. While there are mutations exclusively found in the SCCIS or SCC in cases 3 and 4, their allele frequencies were much lower than the mutations in the dominant clones in each tissue; therefore, we could not rule out the possibility that they came from an unrelated clone of cells, contaminating the macrodissected tissue. Taken together, we conclude the SCC to be mutationally indistinguishable from the dominant clone in the adjacent skin for these three cases. **B.** Cases where the SCC evolved from the neighboring tissue. In cases 8, 9 and 10, there were shared mutations between the two tissues, suggesting that the squamous cell carcinomas were related to the neighboring tissues. There were also mutations exclusive to the squamous cell carcinomas, suggesting that they underwent an additional wave of clonal expansion. Taken together, we generally agree with the authors original interpretation for these cases that these squamous cell carcinomas evolved from neighboring precursor lesions.

Figure S10.

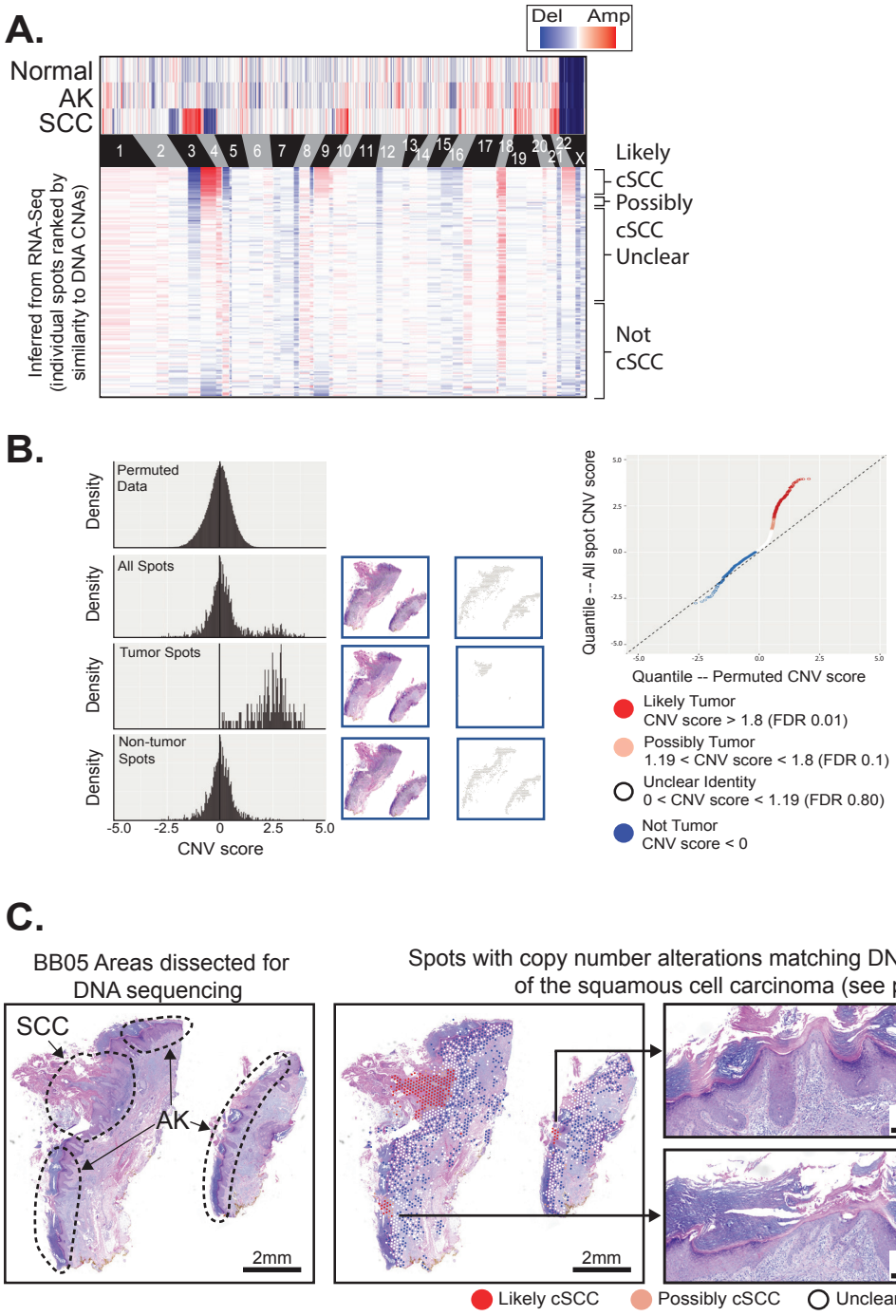


Figure S10. Copy number alterations are detectable in spatial transcriptomic data. A. Copy number alterations (CNAs) were inferred from DNA-Sequencing data (top heatmap) and from RNA-sequencing data of individual spots (lower heatmap). Spots (rows in the lower heatmap) are ranked ordered by the similarity of their copy number profiles to the DNA copy number alterations. **B.** For each spot in panel A, we calculated a CNVscore (see Chen *et al. Genome Biology*, 2023) to capture the similarity of the spot's copy number profile to the copy number inferred from DNA-seq data. Histogram of CNVscores from permuted and observed data are shown. A quantile-quantile plot was generated, comparing observed CNVscores to permuted scores, from which we established q-values to determine which spots were "Likely", "Possibly", or "Not" tumor, as indicated. **C.** The spatial distribution of spots as defined in panel B.

Figure S11.

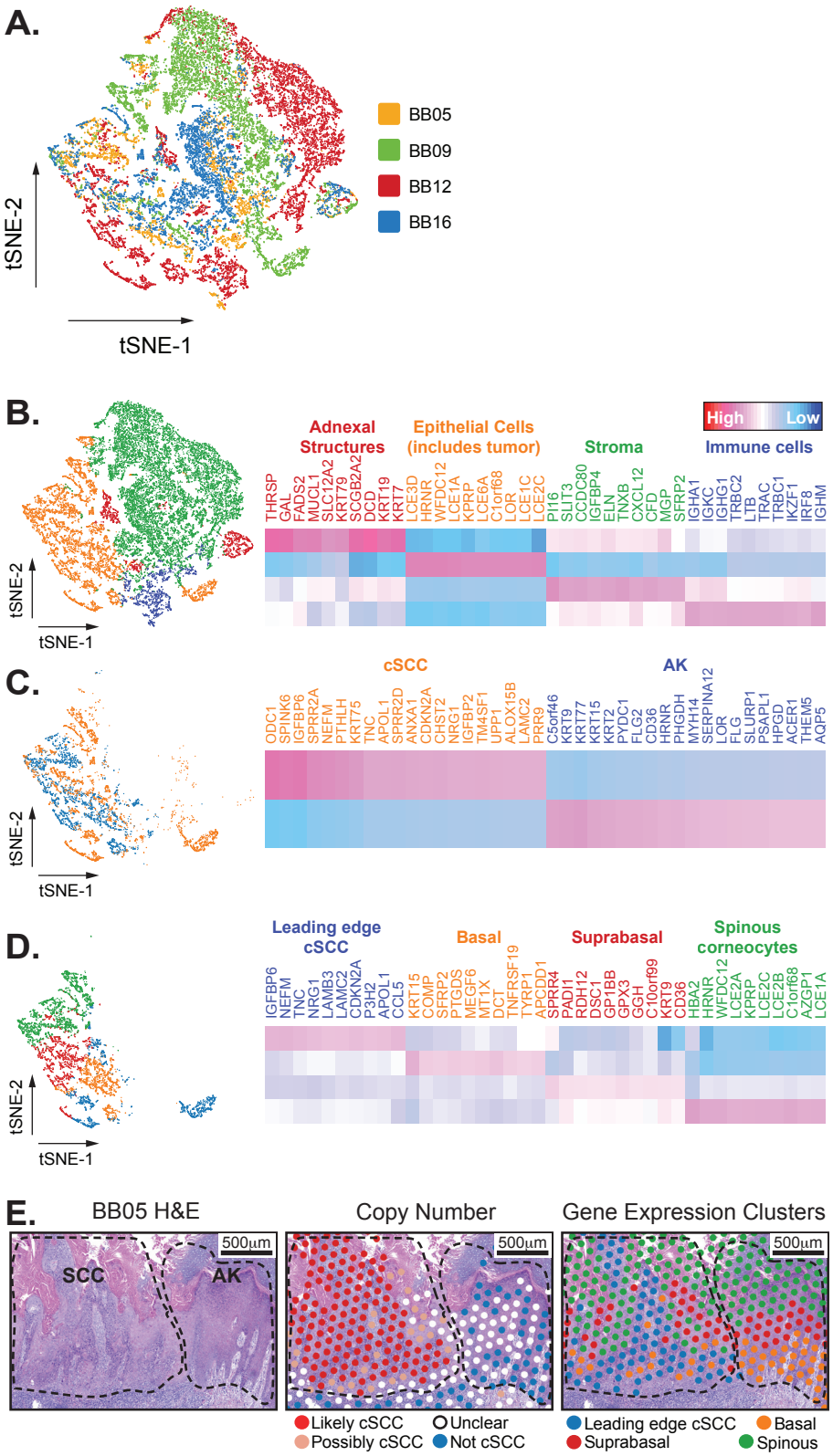


Figure S11. Spatial heterogeneity in gene expression during the evolution of squamous cell carcinoma from actinic keratosis. A. A t-SNE plot of all spots from spatial transcriptomics data colored by sample ID. Spots clustered primarily by cell type or cell state, suggesting that batch effects were minimal. We performed spatial transcriptomics on a fifth case (BB13), but it was not included in this plot because it was profiled with an older version of Visium, whose design did not permit aggregate analyses with samples profiled on newer versions of the platform. **B-D.** t-SNE plots of spots, aggregated from spatial transcriptomic data of four squamous cell carcinomas in association with actinic keratosis. Clusters of main cell types are annotated, as described, with the most highly expressed genes in each group highlighted in heatmaps to the right. **E.** Zoomed view of the junction between squamous cell carcinoma and actinic keratosis in BB05. H&E, copy number, and gene expression clusters are shown, as indicated. Note that keratinocytes in both the squamous cell carcinoma and actinic keratosis occupy a spectrum of differentiation states.

Figure S12.

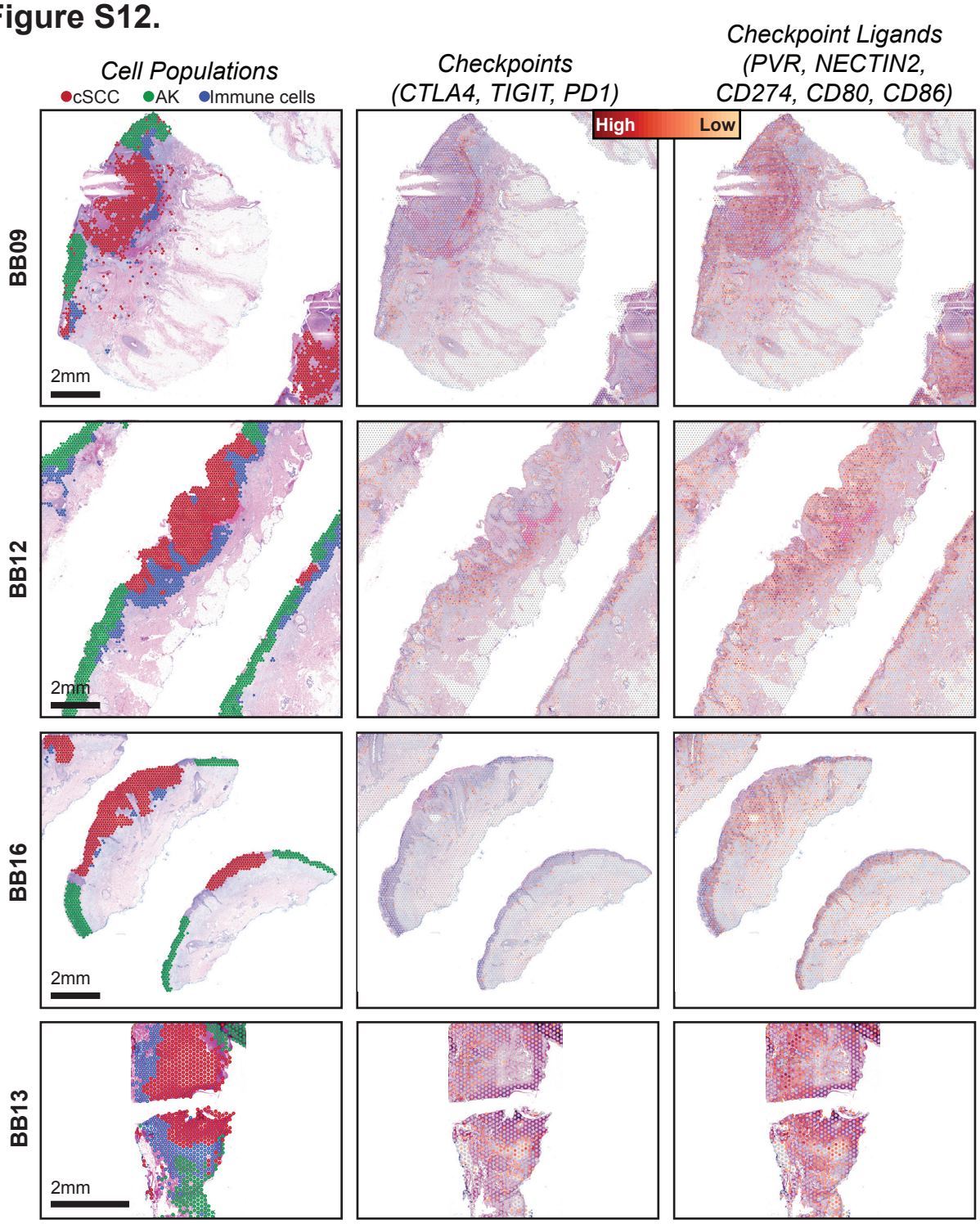


Figure S12. Spatial heterogeneity in gene expression of immune cells at the interface of squamous cell carcinoma versus actinic keratosis -- all cases. Each column of images shows a different view of spatial transcriptomic data, including: annotated spots and gene expression of immune checkpoints and their ligands. See figure 5 for an overview of the BB05. Gene expression intensities represent the combined expression of the checkpoint or ligand genes listed.