**SPECIFIC AIMS**

Cancers have readily-defined characteristics often referred to as “hallmarks”. Nevertheless, the question of how the sequence of cancer development progresses -- from normal tissue to carcinogen-damaged tissue to precancerous lesion and finally to malignant tumors – remains unanswered. Classically, these steps are attributed to the sequential acquisition of discrete genetic events such as driver mutations. However, in humans, the clonal dynamics governing cancer development happen over years, remain largely invisible even in model systems, and have been difficult to link to specific molecular changes. This rubric fails to account for clonal dynamics in the context of tissue architecture and fails to explain the consequences of large numbers of mutations present in normal tissue. Our ***long-term goal*** is to apply ecological and evolutionary principles to **cancer initiation and development** in order to test whether the hallmarks of cancer are acquired in three distinct phases each with distinct **selective pressures** and manifestations of **cell competition and cooperation**.

Nowhere is this more accessible to investigation than in skin. For skin carcinomas, the most important carcinogen is ultraviolet radiation. Cutaneous squamous cell carcinoma (cuSCC) has the most tractable and clinically well-characterized progression sequence of any human cancer, from normal tissue, to a distinct precancerous lesion (the actinic keratosis), to invasive carcinoma. *Therefore, it is ideal for establishing an eco-evolutionary paradigm of cancer initiation and development with respect to modelling clonal dynamics, genetic composition and the dynamics of molecular traits.*

Our **central hypothesis** is that cancer initiation and development occurs in three phases, each with distinctive clonal dynamics. In the first phase, tissue disruption from UV exposure provides a permissive environment where extrinsically-driven mechanisms allow for some clones to have unusually long runs of cell division and turnover. This will increase the variance among clone sizes with larger clones accumulating greater heritable variation. The second phase sees the emergence of intrinsic mechanisms where mutations that confer a competitive advantage allow for clonal selection with directed expansion of some clones at the expense of others. In the third phase, one or several clones escape local tissue control, acquire a distinct fitness function, and form tumors. Within the emerging tumor microenvironments, selection pressures will promote ecological and molecular diversification of the malignant clade(s). Our approach uses novel combinations of serial *in-vivo* quantitative imaging, mathematical modeling, and deep single-cell molecular interrogation to discern the ecological and molecular drivers of clonal dynamics, cell-to-cell competition and cooperation, and clonal evolution, producing a fundamentally unprecedented view of cancer initiation in the following three Aims.

**Aim 1: Characterize the effects of UV-mediated tissue disruption on clonal dynamics**

We hypothesize that chronic UV exposure (phase 1) will increase the coefficient of variation in clone sizes, and increase genetic variability principally between clades rather than within clades as assessed by transcriptional or mutational heterogeneity. To maintain tissue integrity under stress, cooperative dynamics between epithelial clades will necessitate some clades proliferating for more cell divisions than under normal conditions.

**Aim 2: Characterize the clonal dynamics of cancer initiation**

We hypothesize that as UV-exposure continues and is eventually stopped (phase 2), clonal dynamics will shift from extrinsically-mediated damage to intrinsic inter-clade competition and selection where larger clades manifest increased molecular and genetic variability, prior to the emergence of lesions. We posit that suitably expanded clones can now be under selection for the presence of genomic drivers and pathogenic mutations.

**Aim 3: Identify mechanisms of multiclonal eco-evolutionary dynamics in cancer development**

We hypothesize that tumor subpopulations from established lesions (phase 3) can be extracted and used to measure competition and cooperation *in-vivo*. Multiple lines will be derived, cultured, transplanted in combinations, followed over time as distinct clones, and their relative fitness linked to specific molecular and genetic traits. We propose that specific molecular interactions responsible for cell-cell cooperation and competition can be identified using scRNAseq analyses of receptor-ligand and homotypic interactions and validated *in-vivo*.

This proposal establishes a novel framework of cancer development combining ecological and evolutionary principles with deep molecular interrogation to formulate a novel understanding of carcinogenesis in the presence of the selective pressure of UV exposure. The multidisciplinary investigative team at Moffitt Cancer Center is ideally suited to pursue this work and comprises individuals with expertise in skin cancer (Tsai), mathematical oncology and ecology (Brown), clonal dynamics (Andor), and single-cell sequencing (Chen).

**BACKGROUND AND SIGNIFICANCE**

Perhaps the dominant paradigm of cancer initiation states that cancers arise from sequential accumulations of largely cell-intrinsic alterations, often genetic in nature, over a period of years to decades1. The identification of driver mutations or inactivated tumor suppressor genes is consistent with the hypothesis. This orderly progression has been most classically elucidated for colon adenocarcinoma2. While the critical role of the tumor microenvironment and immune system have come to the fore more recently, this dominant paradigm remains the focus of most approaches to understanding cancer etiology, progression, and therapy3,4.

Nevertheless, this pathway to tumorigenesis seems incomplete and violated in skin cancers. Skin is a highly ordered structure. The squamous epithelia of the epidermis undergo constant self-renewal, repopulated by progenitor and stem cell populations which can be differentially engaged during normal homeostasis or under conditions of stress or wounding5. Despite this backdrop of highly-ordered regeneration, skin cancers (carcinomas in particular) are by far the most common of all human malignancies6. Evidence shows that this is due to exposure to UV radiation. This is further evidenced by the finding of a high mutational burden in UV-exposed skin dominated by C🡪T transitions7,8. At the whole exome level, this represents a mutational burden of 5 mutations per Mb, exceeding the median of many human cancers8-10. Furthermore, sun-exposed vs. sun-protected areas of human skin have vastly different mutational profiles11. *Yet the vast majority of chronically exposed areas of skin do not develop cancer.* For example, introduction of mutant *Trp53* in mouse epidermis fails accelerate initial tumorigenesis following chronic UV exposure or following chemical carcinogenesis. A short term proliferative advantage is quickly quashed. After the UV exposure, these clones are eventually replaced by other spontaneously emergent clones12,13. While mutant *Trp53* does allow for cells to persist in the face of UV exposure, it is also clear that these clones are constrained from expanding indefinitely. Recent data suggest that the outcome of clonal competition in the esophagus is dictated by fitness differences of adjacent clones, though no molecular mechanisms have been identified14. Keratinocyte clones with presumably mutant *TP53* can readily be found in sun-exposed human skin15. These data collectively suggest that clones contributing to cancers must acquire an optimal combination of mutations at the right time.

**Mutant clones competition**

**Tumor**

growth of larger clones

constrained

Clones grow to fill vacated neighborhoods;

DNA damage

Evolution and diversification independently of the host

**Homeostasis**

chronic UV compromises spatial constraints

Clones probing (epi)genetics to increase fitness (intra & inter lineage competition)

**(1) Disruption Phase**

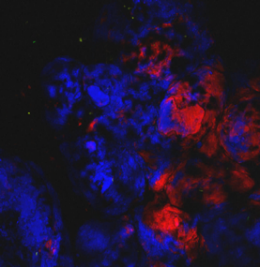
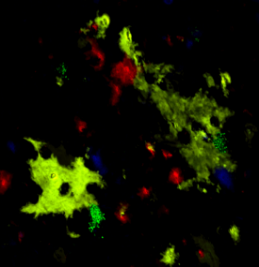
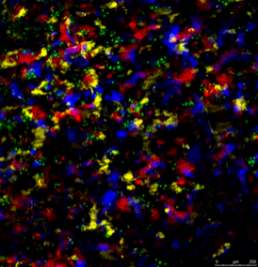
host control in place over expanded clones

**(2) Clonal Selection Phase**

cells developing self-defined fitness

**(3) Lesion phase**

breakout of the fittest



**Coefficient of variance ≈ 1**

**Coefficient of variance ˃˃ 1**

**Tumor**

**Figure 1. Cancer development as a process in three phases.** In the course of extrinsic damage, a wide variance in clone sizes begins to develop, but tissue control predominates (phase 1) until cell intrinsic alterations drive competition (phase 2) eventuating in tumors which have independent fitness functions (phase 3).

Most approaches that examine clonal dynamics do so indirectly and focus on hi-depth bulk sequencing to discern clonal structure7,8. Approaches such as focusing on the ratio of non-synonymous to synonymous mutations in cancer-related genes assume that key driver events are mutational7,16. The role of the spatio-temporal context in which the mutations arise go unmeasured and unknown. Our own data suggest that the vast majority of clones detected at the whole exome level in the skin of UV-exposed mice have mutations in genes that are not cancer-related and are not likely to be expressed in skin11. Such mutations may remain unrepaired because of non-uniform DNA repair mechanisms across the genome17. Therefore, looking only at the coding regions of “cancer-related” genes misses a large part of the clonal dynamics which likely govern initial cancer development.

In many complex tissue types such as skin epidermis, differentiation is an orderly process of maturation across ascending layers of keratinocytes with complete turnover of the epidermis requiring approximately 2 weeks in mice18,19. This likely provides the context for suppressing tumor initiation even in the presence of driver mutations.

Our **central hypothesis** is that cancer initiation and development occurs in three phases, each with specific and recognizable clonal dynamics (**Fig. 1**)3. Prior models have not discerned these three phases. We propose to use a model system for a disease process which we think is ideal for identifying and testing for these distinct phases. Cutaneous squamous cell carcinoma (cuSCC) *has the most accessible and clinically well-characterized progression sequence of any human cancer*, from normal tissue, to a distinct precancerous lesion (the actinic keratosis), to invasive carcinoma. *Therefore, it is an ideal model for establishing an eco-evolutionary paradigm of cancer initiation and progression with respect to modelling clonal dynamics, genetic composition and molecular traits.* Stratified squamous epithelia form environmental barriers in the airways, gastrointestinal / genitourinary tracts, and skin. The majority of SCCs driven by carcinogenic exposures such as tobacco and solar radiation. Previous results and our own data show that SCCs from diverse sites share molecular commonalities including alterations in global gene expression and in *TP53*, *TP63*, *NOTCH*, and *SOX2* signaling20-26. We surmise that the lessons derived from this work has wide applicability across all types of SCC. We propose to use a novel set of tools, analyses, and conceptual framework to examine the entire sequence of carcinogenesis.

The first phase (“disruption”) requires that basal cells (or progenitor cells) experience unusually long runs of cell division and cell turnover as would occur following perturbations (e.g., UV radiation to the skin), wounding or chronic inflammation12,27-30. This ultimately permits existing somatic mutations within that cell to accumulate additional epigenetic and/or genetic changes. Here, interventions that mitigate UV-induced inflammatory effects, might balance cell divisions more equitably among progenitor cell units. As an example, prostaglandin PGE2 is a well-established mediator of UV-induced inflammation and the proliferative adaptive response in epidermis which follows UV-mediated injury31-34. Mice lacking PGE2 receptors are relatively resistant to UV-driven cuSCC formation31, which can be mimicked by topical diclofenac, a COX2 inhibitor35. Alternatively, those stimuli which disrupt tissue structure directly such as wounding, at this stage, would be expected to have maximal effect in promoting carcinogenesis. In this phase the variability in the size and proliferation rates of different clades arises exogeneously from tissue control and the need for clonal cooperation to maintain cell number and epidermal integrity. Some clades are filling the space left by damaged or lost cells, or gaps created by injured tissue.

The second phase (“clonal selection”) involves the larger clones retaining and even expanding their space at the expense of other clones. Continued cell turnover by virtue of filling a larger space permits the accumulation of further somatic mutations that favor clonal expansion, selection and evolution7,8,36. In this phase, we expect to see the first evidence of oncogenic genes or gene expression permitting clonal selection. The endogenous properties of the clone become salient. Cooperation turns to competition with adjacent clades. During this phase, we predict that the introduction of specific oncogenic genetic elements or the introduction of targeted agents that shut down proliferation pathways will have strong effects in promoting or restraining carcinogenesis, respectively.

The third phase (“lesion”) involves a clade that has advanced to be cancerous. It is now subjected to natural selection, and we expect evolutionary triage37. Adaptive traits favoring survival and proliferation should replace less successful ones. An ecological and evolutionary bottleneck ensues as the cancerous cells forming the lesion evolve from and branch off of the large clade that had emerged during the previous phase. What emerges is a lesion and progressive disease. In this 3rd phase, the tumor itself may begin either from one clone (that outcompetes all others) or from several clones (that mutually benefit each other)38 As the founder cell(s) proliferate and evolve adaptations to avoid hazards, to exploit opportunities, and in response to each other’s traits, we expect them to diversify to fill ecological opportunities and specialize on different tumor microenvironments. Following the genetic bottleneck there should be a rapid accumulation of molecular and genetic diversity that is distinct from that seen within and among the clades of Phase 2. Such diversity will be in response to the necessities and opportunities for cell-cell competition and cooperation.

Cell-cell competition and cooperation has a rich history particularly in developmental biology39. Recently, it has been shown that even in the context of normal skin homeostasis, basal layer progenitor cells appear to compete for sites on the basement membrane40. Manipulation of c-Myc expression can drive fitness differences among epidermal keratinocytes41. Human Flower isoforms can also confer differential fitness on human tumor cells42. How these mechanisms occur for UV-driven skin carcinogenesis, however, remains an open question.

While various components of our central hypothesis have been noted and even tested in the literature, they have never been examined as a continuum that moves seamlessly from one phase to the next. Furthermore, we have extensive experience with models of carcinogenesis in the skin that lend themselves to incorporating these processes, which can then be expanded to incorporate the eco-evolutionary dynamics, as well as the effects of perturbations hypothesized in this proposal3,43-46. *Our proposal aims to anticipate the key ecological and evolutionary drivers of each phase, testing their relevance experimentally by validating or refuting specific predictions that our model makes for each phase of tumor development.*

**PRELIMINARY DATA**

**Figure 2. Dynamical quantification of clades across time following UV-exposure.** The violin plots show the distribution of clade numbers (top), log-transformed clade volumes (middle row), and the coefficient of variation of clade sizes (bottom row) for each sample (exposed vs. non-exposed; left column) and in aggregate (exposed vs. non-exposed; right column). The total number of clades declines significantly from month 1 to subsequent months regardless of exposure. Mean clade sizes remain constant with time with UV exceeding non-UV in months 1 & 2 and the opposite in months 3 & 4. The CV his higher for UV than non-UV and this difference increases substantially in months 3 & 4.



Unexposed (NON)

Exposed (UV)

We have extensive experience with the UV-driven model of cuSCC in Hairless mice to elucidate the genomics of UV-induced cuSCC development8. To enable *in-vivo* serial assessment of clonal dynamics, we generated mice harboring the ROSA26-BRainbow2.1 cassette47 in combination with K14-ERT2Cre48 to obtain K14-CreERT2 Confetti mice. We then bred the Confetti mice with SKH1 hairless mice (hereafter referred to as K14-Confetti mice), which are immunocompetent and susceptible to developing cuSCC upon UV irradiation. Upon exposure to topical tamoxifen, one of four fluorophores is permanently expressed by K14-expressing keratinocytes at that time and that expression is subsequently inherited by all daughter cells. Therefore, a group of cells expressing a single fluorophore and related by descent.

Two weeks following the topical tamoxifen, the mice were UV-irradiated in evenly divided doses three times a week for 3 months for a total UVB exposure of 175 kJ/m2 (Newport Solar Simulators), at a standard erythemal dose of 3.48,49,50. During UV exposure, half of the mouse’s back was covered with fabric for a within-individual unexposed control (**Fig. 6**). Intravital laser scanning confocal imaging (Leica SP5) was performed on exposed (UV) and unexposed (non-UV) for within-mouse controls. At monthly intervals (0, 1, 2, 3, and 4 months after the start of irradiation) and for each mouse, we generated from 4 to 5 spatially-distributed images per UV and non-UV areas through the entire course of radiation and for additional months out to 6 months. Each image measures 1.55 mm x 1.55 mm x 60-100 microns deep. Once lesions were observed (around month 5), focused imaging on those areas and on the lesions (in this proposal we will extend the months of imaging and starting with the 5 month sampling we will include lesions as well as skin areas without detectable lesions). Argon (488 and 514 nm) and helium-neon (543 and 633 nm) lasers were deployed at 20% power to excite the fluorescent reporters. Imaging areas were sectioned optically producing a z-stack with 3 micron spacing starting from the basal layer up through the superficial epidermis. A scanner frequency of 400 Hz with 1024x1024 pixel resolution provides the optimal balance between image quality and time spent restraining the mouse. Images were deconvoluted to ensure accurate reads from each of the four fluorophores. A set of filtering operations was applied to the images to accurately recognize adjacent clades. The contrast was adjusted using the Percentage Linear Contrast Stretch method, where the percentage range was set to 95%51. To obtain 3-D volumetric data, images were grouped and stacked together, and segmented into positively stained regions that correspond to the space occupied by a given clade using the Otsu thresholding algorithm52. The volume of each clade was calculated by summing the number of pixels across the 3-D image53,54. For the proposed research ecological data from each sample will include i) the mouse, the month, ii) the UV treatment, iii) the color of a given clade, iv) the number of clades per color, v) clade size, vi) variance in clade sizes per sample (using coefficient of variation = (std. dev.)/(mean)), vii) clade compaction (density of cells per unit volume), and viii) interdigitation of the clades (co-localization of cells from adjacent clades). Here, we can report on preliminary data relating to i-vi.

The ecological data lend themselves to multi-way, fully-crossed ANOVAs where Color and Mouse represent random effects variables and month and UV-treatment represent fixed effects (Table 1 provides an example with CV being tested for the main effects and two-way interaction effects). For **clade numbers** there is a significant effect of month. The first month had c. 30% more clades than later months. UV treatment did not significantly influence the number of clades. But, a significant interaction effect demonstrated that in the first two months there were fewer clades under UV exposure and the opposite in months 3 and 4. The distribution of clade volumes per sample are highly skewed towards very large clades (mean > median). The distribution of clone sizes from the 456 total samples (6 mice x 4 months x 2 colors x 2 UV treatments x 4-5 samples) closely fit a lognormal distribution. Hence, we can fully characterize the within sample distribution of clade volumes as the mean and variance of the natural-log transformed volume data. Average **clade size** does not vary in any predictable way with month. Average clade size declines significantly with UV exposure. A significant interaction effect shows that in months 1 and 2 clade sizes do not differ between UV and non-UV (in fact the trend is UV > non-UV), but that clade sizes under no-UV are significantly larger than UV in months 3 and 4. As predicted, the **coefficient of variation** (CV) of clade sizes within a sample increase significantly with month, and with UV. A significant interaction effect shows that the effect of UV on increasing the variance in clone sizes becomes most pronounced in months 3 and 4. Using an adjusted CV (based on mean CV’s for a given mouse during a given month), three of the mice showed a striking increase in month 3 and three not until month 4 (**Fig. 3**). By all three ecological metrics, we see phase shifts, and these phase shifts primarily distinguish months 1 & 2 from months 3 & 4.



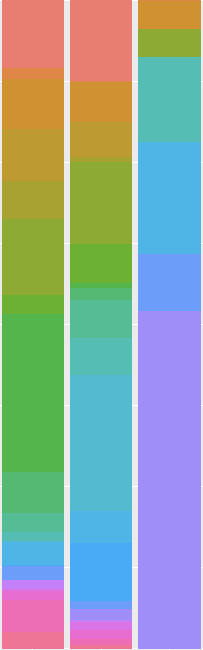
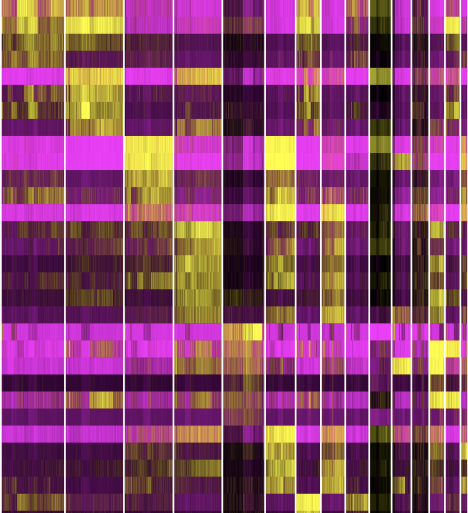
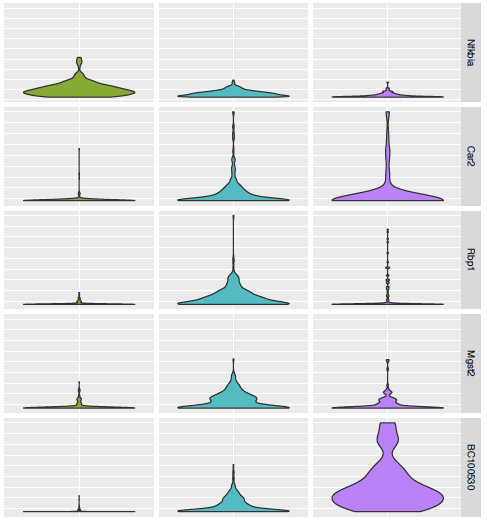
**Figure 3. Changes in the coefficient of variation in response to UV for six mice over a period of 4 months where UV ceased at month 3.** The y-axis is an adjusted coefficient of variation (CV Adjusted) calculated as the ratio of mean CV between UV and non-UV exposed skin for a given mouse on a given month (6 mice x 4 months = 24 measures). While UV exposure increases CV (all but one point is >1) this increase remains flat between months 1 and 2; between 2 and 3 some mice increase a lot while others not so much, and clear increases for all mice by month 4.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source** | **Type III SS** | **df** | **F-Ratio** | **p-Value** |
| Mouse | 48.85 | 5 | 16.49 | **0.000** |
| Month | 29.74 | 3 | 16.73 | **0.000** |
| UV Treatment | 61.34 | 1 | 103.53 | **0.000** |
| Color | 34.73 | 1 | 58.61 | **0.000** |
| Month\*Mouse | 24.21 | 15 | 2.72 | **0.001** |
| UV Treatment\*Mouse | 5.94 | 5 | 2.00 | 0.077 |
| Color\*Mouse | 6.41 | 5 | 2.16 | 0.057 |
| UV Treatment\*Month | 14.07 | 3 | 7.91 | **0.000** |
| Color\*Month | 4.11 | 3 | 2.31 | 0.076 |
| UV Treatment\*Color | 2.14 | 1 | 3.62 | 0.058 |
| Error | 244.11 | 412 |  |  |

The ANOVA results (Table) also indicate that mice vary significantly from each other, and that the color of the clones (in this case yellow versus red) vary significantly. Red significantly exceeds Yellow for clade numbers and CV, Yellow significantly exceeds Red for average clade size. These color effects may reflect differential recombination efficiencies previously reported in Confetti models55. For our purposes, we can factor out the effects of individual mouse and color, and in no cases did mouse or color effects reverse the results pertaining to effects of month and UV.

We probed the molecular genetic structure of these clades using scRNAseq (10X Genomics). We took 2 mm – 4 mm punch biopsies, dissociated them into single-cell suspensions, performed encapsulation, and sequenced on the Illumina platform. Differential gene expression analysis and unbiased Louvain clustering of cells across all epidermal samples (non-UV and UV exposed) produced 16 different clusters (**Fig. 4**), the majority of which could be mapped to previously defined keratinocyte populations including multiple subtypes of interfollicular keratinocytes, suprabasilar keratinocytes, basal keratinocytes and infundibular keratinocytes56-58. scRNAseq of samples following UV exposure or tumor development revealed differential representation of these clusters (**Fig. 4, middle**). These clusters in UV areas were associated with expression of cystatins (Scfa 3, BC100530), and alarmins / proliferative keratins, such as Krt16 and Krt6a, which have been associated with skin injury. By contrast, clusters expressing genes related to keratinocyte differentiation, such as filaggrin, Crct1, and Krtdap occurred at reduced frequency. Clusters expressing cystatins and alarmins also increased in tumors.56-58 Flr-expressing keratinocytes harvested from large clones in skin that had seen 3 months UV exposure exhibited altered keratinocyte differentiation (downregulation of Krt77, Loricrin and Nfkbia, upregulation of cystatin), inflammation (downregulation of Nfkbia), and upregulation of metabolic regulators (carbonic anhydrase II and retinol transport (Rbp1)) (**Fig. 4**). These changes were maintained during tumor formation, suggesting that they are likely to be important in carcinogenesis.

**Figure 4. Single-cell RNAseq reveals defined clusters of epidermal keratinocytes and differentially represented clusters and genes.** Unbiased clustering reveals 16 subsets of keratinocytes (left) partially defined by keratin and differentiation marker expression. These clusters, represented as individual colors (middle), are plotted as proportions of total sequenced cells grouped in each cluster. There is a progressive enrichment for the blue / cyan clusters in exposed skin and then the purple cluster in the tumor samples. are differentially represented in unexposed (NON) vs. exposed skin (EXP) vs. tumors (TUM) (middle). Genes differentially expressed in exposed skin and retained in tumors may be required for carcinogenesis; those differentially expressed only in exposed skin may only be required for adaptive responses to UV exposure.



NON

EXP

TUM

Nfkbia

Car2

Rbp1

Mgst2

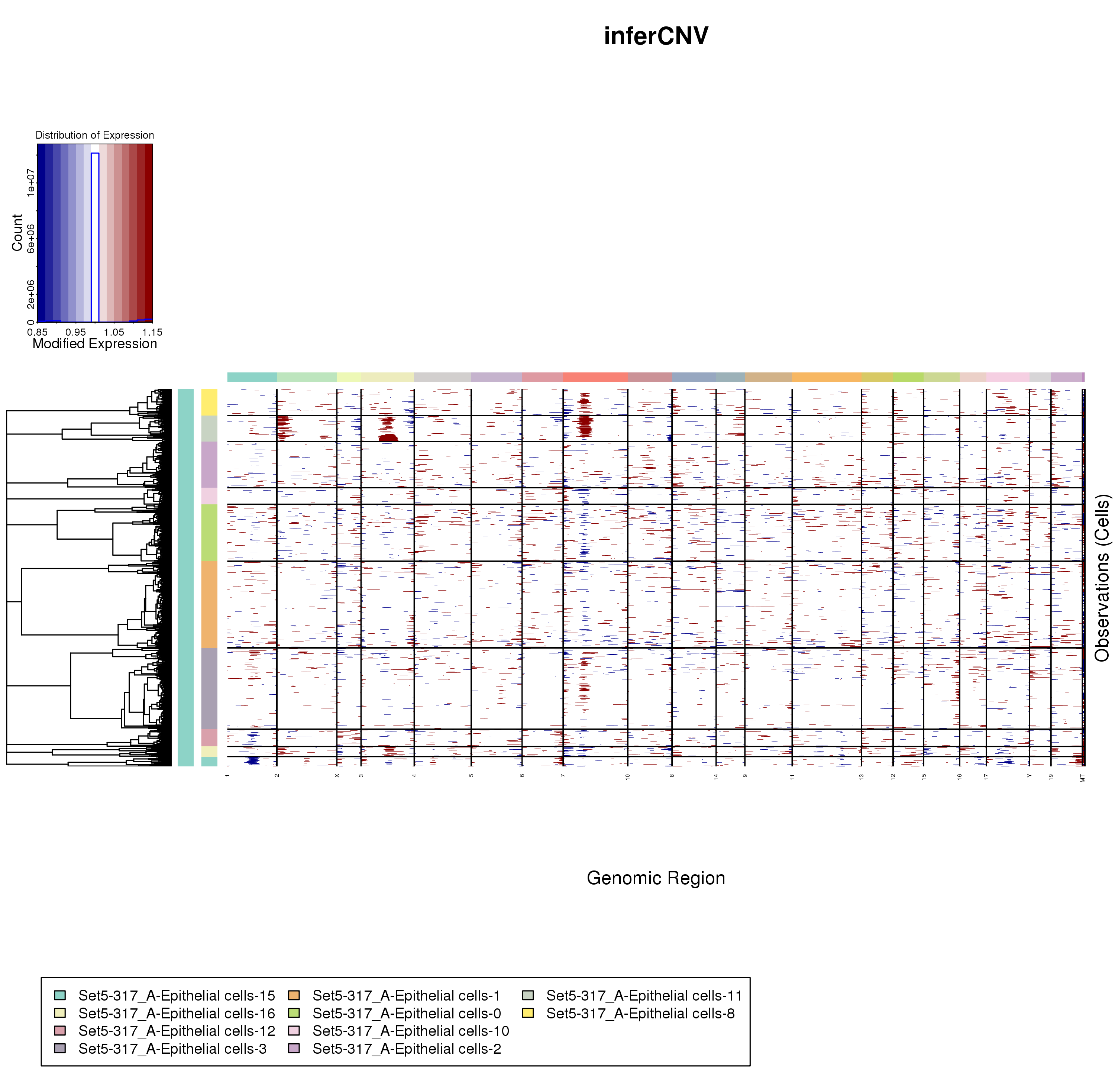
BC100530

NON

EXP

TUM

Clonal expansion in UV-exposed areas was associated with upregulation of oxidative phosphorylation, , adipogenesis, MYC targets, mTORC1 signaling, and glycolysis. UV-exposed clades showed downregulation of IFN-and TNF- signaling, myogenesis, epithelial mesenchymal transition (EMT), apoptosis, and p53 related pathways as compared to non UV-exposed ones. It is known that c-MYC mediates cell-cell competition in a variety of contexts39,41,59, properties which we will exploit experimentally as part of Aim 3. Some of these pathways continued to be downregulated during tumor development, implying a role in driving tumorigenesis. By contrast, upregulation of mTORC1 signaling and metabolic pathways were not observed in tumor cells, suggesting a more transient role in acute adaptations to UV exposure. Therefore, scRNAseq profiles of keratinocytes revealed higher genomic transcriptional diversity in UV vs. non-UV exposed skin. We then asked whether this diversification was accompanied by genetic diversification. Employed InferCNV to assess copy number variations, UV-exposed clades showed some recurrent amplifications in chr 3 and 7. In tumor samples, we also identified recurrent amplifications in chr 3 and chr 7. Interestingly, we see evidence for genetic subpopulations exhibiting different CNV (**Fig. 5**) in tumors suggesting that the diversification expected during Phase 3 had already begun. *Therefore, cells that had multiplied from a founder subclade can occupy distinct transcriptional states and diversify into different genetic subpopulations.*



1

3

7

A

A

B

B

**EXP**

**TUM**

**Figure 5. InferCNV Identifies Recurrent Copy Number Variations (CNV) in EXP and Tumor samples.** Two samples processed for InferCNV are shown: an EXP sample following 3-month exposure (top row) and a tumor sample (bottom row) highlighting two scRNAseq-based transcriptional clusters each (A, B) within each sample. Each row containing red (gain) or blue (loss) lines is one sequenced cell. On the bottom, chromosomes 1, 3, 7 are highlighted showing areas with the most recurrent CNVs. Whereas the EXP sample shows some gains and losses in Chr 7, the tumor sample appears to show recurrent gains in Chr 1, 3, 7. Within cluster B of the tumor sample, the dotted blue line delineates distinct genetic subpopulations bearing distinct sets of CNVs. Above the blue line are cells enriched for gains in Chr 7; below the line show cells enriched for gains in Chr 3, but not Chr 7.

For our purposes, a clade is a structural entity composed of identically-labeled cells related by descent. *Our data also suggest that cells within a clade may also exhibit multiple transcriptional states (****Fig. 4****) as well as multiple genetic subpopulations (****Fig. 5****), two additional sources of heterogeneity.* This key issue distinguishes our series of experiments from those previously reported. Within individual clades there are multiple transcriptional states (perhaps related to differentiation, adaptation to UV exposure, or cancer-related pathways) and multiple genetic subpopulations. We can resolve these multiple levels of cell-to-cell relatedness and cell-to-cell communication so critically important for the understanding of progression to carcinoma.

**INNOVATION** The fundamental questions of cancer initiation addressed here can only be answered by model systems that permit real-time assessment of clonal dynamics, single cell interrogation, and in-vivo manipulation of clonal competition. Therefore, we propose that there are points of both conceptual and experimental novelty.

The key points of novelty are:

1. Use of long-term serial *in-vivo* imaging over months in a physically accessible model of skin carcinogenesis.
2. Use of an etiologically-relevant complete carcinogen (solar simulated UV exposure) in a genomically-characterized faithful model of human cuSCC.
3. Deep single-cell characterization of transcriptional diversification and pathway analysis of specific labeled clades.
4. Adaptation of single-cell sequencing to discern clonal and subclonal genetic structure (CNV, mutations) within clades.
5. Demonstration and exploration of the significance of polyclonality and the ability to directly test molecular drivers of cooperativity and competition among clones of the same lineage.
6. The ability to make specific predictions about the significant and salient features of each phase of tumor development and how specific perturbations affect the process only in those specific phases. This key difference from most modeling approaches is that our conceptual framework makes predictions about how specific perturbations affect each phase.

We feel that the combination of conceptual novelty and unique set of tools position us to probe the ecological and evolutionary underpinnings of carcinogenesis with unprecedented resolution in a living organism over time.

**RESEARCH STRATEGY**

***Four Key Experiments with Confetti Mice***:

We know that: 1) tamoxifen successfully labels clades of keratinocytes within confetti mice that can be quantified, 2) three months of UV can be administered experimentally in a manner where exposure is limited to just half of the mouse’s back, and 3) approximately two months after this UV exposure no lesions will form on the unexposed skin while multiple lesions will form on the exposed skin. This protocol of administering tamoxifen two weeks prior to time 0 (start of UV), followed by 3 months of UV exposure to half the skin surface area, and then a cessation of UV for the subsequent months of study provided the preliminary data. It will also provide our Standard Procedure and the reference treatment for four distinct and inter-related experiments: **Timing of clade labelling, Dose spacing, Chemical disruption, Genetic disruption.** Each of these experiments will generate the necessary data and source material for **Aims 1, 2 & 3**. Our approach allows us to capitalize on the temporal phases of the model and utilize the power of both within and across mouse comparisons in time (across months) and space (UV versus non-UV exposed skin). A key advantage of our approach is the ability to correlate imaging findings to deep single cell interrogation.

***Ecological Properties of Clades***: For each mouse, the z-stack confocal imaging will be performed once a month starting with month 0 and continuing through to month 7 (8 temporal sampling periods). By month 4 essentially none of the reference treatment mice will exhibit lesions, by month 5 most will, and by 6 essentially all will8. Each month, 4 randomly selected areas (1.55 mm x 1.55 mm) will be imaged from the non-UV exposed skin and 4 from the UV exposed areas (note that these 8 sampling areas will not correspond to the areas either from prior or subsequent months). Starting with month 5, there will be 8 imaged areas free of lesions, and up to four additional imaged areas will be added where the center of the area includes a lesion. During months 6, 7, and 8 the areas with lesions will follow the same lesion through time (unlike the randomly placed imaged areas). For each mouse this will yield 64 images = 4 areas x UV vs. non-UV x 8 imaging months. If there are 4 lesions by month five, this will also yield 12 lesion-specific images = 4 lesions x 3 months as repeated measures of the same lesion. From each of these image sets we will obtain information on 1) the **number** of clades, 2) the **volume** of each clade, 3) the **co-localization** of clades (degree of overlap and inter-digitization of calls from adjacent clades, and 4) **compaction** (cell number per unit volume) from up to 10 clades per sample where clades will represent a full spread from large to small clades.

***Molecular and Genetic Properties****:* The molecular and genetic interrogation of clades as identified by imaging will be based on deep single cell interrogation using scRNAseq. The serial imaging serves as the basis for selecting the lesions for sampling. Once the distributions of clade volume, co-localization, and compaction are established to some degree, we will perform 1 mm – 2 mm punch biopsies to sample areas with 1-3 easily identified clades. This sampling is destructive and we will be limited to eight samples per mouse total. Therefore, we will require additional mice to enable the sampling of 9 clades at 4 time points (0, 2, 4, 6 months). These nine lesions will be composed of three sets of three lesions at the c. 15th, 50th, >85th (or near largest of that image area) percentile of clade volumes at that time point (as averaged over the 8 samples from the mouse) such that small and large clades are sampled.

The sampled lesions will be dissociated and scRNAseq performed. First, we will classify each sequenced cell to previously defined keratinocyte populations, including subtypes of interfollicular keratinocytes, suprabasilar keratinocytes, basal keratinocytes and infundibular keratinocytes57,58,60. Next, we will assign individual cells to their cell cycle state as previously described61-63. We will then delineate transcriptional states and genetic subpopulations among cells in the G0G1 state only. InferCNV (<https://github.com/broadinstitute/inferCNV>) will be used to identify genetically distinct subpopulations defined by large scale chromosomal changes. The distribution of clusters (**Fig. 4**) will be assessed across all samples. Transcriptional diversity will be measured using SinCHet64, a novel algorithm developed by Ann Chen, PhD (co-I) which has been successfully used to measure heterogeneity in cell lines. We will adapt its use here to distinguish transcriptional states within and across clones. CellPhoneDB65 will be applied to fluorescing cells and any adjacent cells as identified by imaging to identify co-regulated or co-expressed receptor ligand interactions. TRANSFAC motif-based analysis will also be performed to identify upstream transcription factors driving these interactions8. Finally, to quantify how cells differ in their pathway activity we will use GSVA66 whereby the REACTOME database will define the pathway search space67 and will include pathways involved in differentiation and contact inhibition. Taken together this integrated analysis will inform: (i) the **number** of CNV-defined clones and how it relates to the **number** of clades inferred from imaging; (ii) what percentage of cells from each CNV-defined clone are in **S-phase** – this statistic can serve as a surrogate of clone growth rate61 and will be compared to clade **volume**; (iii) receptor-ligand interactions between co-existing clones and (iv) genes, transcription factors and pathways differentially expressed between CNV-defined clones – for example, HIPPO pathway activity will be tested as a proxy of **contact inhibition** and will be compared to **compaction** as inferred from imaging.

***Mathematical Modelling***: Integrated Mathematical Oncology at Moffitt has a portfolio of wound healing and clonal evolution models of the skin to build upon25,68,69. Other teams have developed models applicable to epithelial tissues that distinguish between neutral evolution (as we predict in Phase 1) and positive selection resulting in clonal competition (as we predict in Phase 2)70. Such models also can be tested and fit against data on the distribution of clade sizes in samples such as ours71. In addition to using these, we will develop a quasi-agent based model where basal cells populate, in a semi-uniform pattern, a 2-D landscape representing the skin from above (the height from the basal cells to the differentiated keratinocytes, and other cell types will remain implicit). The homeostatic function of each basal cell is to populate the tessera (irregular convex polygon) defined by all surface points nearest to it. The rate at which the basal cell needs to divide depends upon the skin area it services, the rate of surface cell turnover, and the number of subsequent cell divisions accruing to a daughter cell that goes on to eventually differentiate. We assume that basal cells primarily divide asymmetrically. Thus, maintaining but rarely propagating themselves. Mutations can accrue to the basal cell at a very low rate based upon each cell division. In addition to neutral or deleterious mutations, mutations may 1) increase the probability of a symmetric cell division of the basal cell, 2) increase the area serviced by the basal cell at the expense of neighbors (a boundary pressure trait such as contact inhibition), or 3) reduce the background death rate of a basal cell. These mutations can be associated with known oncogenes or traits revealed in the preliminary results. We can then model UV perturbation (or other insults) as raising the mortality rate of basal cells (thus increasing the tessera size of remaining basal cells) and differentiated cells (further increasing the proliferation rates of surviving basal cells), and acting to increase the mutation rate. Based on model output we can parameterize the model to reflect the time series data of clone sizes, distributions, compaction, and mutational states. The mice experiencing the Standard Procedure provide a near zero likelihood of cancer in the non-UV areas and the near certainty in the UV-exposed areas following 5 months. The model can be parameterized and fit to Standard Procedure mice and then applied to predict outcomes for the mice undergoing different treatments from the 4 experiments.

**Figure 6. Experimental Plan for Clone Identification.** The standard irradiation scheme of exposure across the first three months is illustrated using three yellow blocks, followed by two black ones. The backs of the mice are divided evenly into UV-exposed (yellow) and UV-protected (black) areas. This maximizes the ability to use within-mouse controls for sample comparisons.

UV-exposed

**Clonal ID**

0

1

2

3

4

5

6

7

Tam

Tam

Tam

Tam

**Months**

**UV Dose**

**Regimen**

UV-protected

***Experiment 1: Timing of clade labelling*:** Our first experiment aims to reset the clock on clade identification by adding the tamoxifen at different time points (**Fig. 6**). This will involve 20 mice divided into 4 treatment groups. All mice will receive the Standard Procedure with the following adjustments. The four groups of five mice will vary in when tamoxifen is applied. Tamoxifen will be applied two weeks prior to either 0 (Standard Procedure), 2, 4, or 6 months relative to the application of UV. We presume that the volume of cells represented by a specific color and contiguous space represents a clade of cells that originated from a single progenitor/basal cell. The clades over time may see heritable changes (epigenetically or mutationally) that create distinct branch points and subclades that may now have different ecological and evolutionary properties. This should especially be true for UV exposed skin during Phase 2 and, in particular, during Phase 3. When tamoxifen is applied at time 0, any branching of a clade into subclades will be obscured as they will all retain the same color and space. By starting tamoxifen at later time points (of course there is no need to image samples prior to applying tamoxifen) one resets the membership within clades and less or more successful emerging subclades will now manifest with their own colors, volumes and compaction. Thus, we expect no effects of delaying tamoxifen on non-UV exposed skin other than to create a time delay in observed clade structures. But, differences in clade metrics (adjusted for the time lag) at 2, 4 and 6 months will reveal and test for Phase 1 (for **Aim 1**), Phase 2 (for **Aim 2**) and/or Phase 3 (for **Aim 3**) effects, respectively.

***Experiment 2: Dose Spacing*:**Our second experiment tests for the effects of duration and timing of UV exposure relative to the Standard Procedure (**Fig. 7**). This will involve 25 mice divided into 5 treatment groups. For exposure duration, sets of mice will receive 1, 2 or 3 consecutive months of UV, respectively. For the timing of dosing, one treatment will receive two 1-month “holidays” from UV in between three 1-month duration UV exposures. As an additional control to the Standard Procedure, we add a fifth treatment where the mice receive no UV for the first 2-months of sampling and then UV for three consecutive months. Three month exposure results in multiple cancerous lesions with near certainty. We suspect that just one month will result in no lesions and two month may lean either way. For Aims 1 and 2, one and two month exposure treatment will assist in testing whether clade architectures develop gradually or more abruptly transitioning between cooperativity between clades and competition between clades. The spacing of UV holidays between single months of UV will test the role of total dose versus the relentlessness of dosing. This will feed directly into **Aims 1 & 2**, and into **Aim 3** should lesions develop.

**Figure 7. Experimental Plan to Test Effects of Dose Spacing.** Three irradiation regimens are shown – the standard (top – lined in red), one in which the equivalent UV dose is divided (middle) and a control for aging performed by replicating the standard regimen with a delayed start (bottom). The backs of the mice are divided evenly into UV-exposed (yellow) and UV-protected (black) areas. This maximizes the ability to use within-mouse controls for sample comparisons.

**UV Dose**

**Regimens**

**Phase**

I

II

III

UV-protected

UV-exposed

0

1

2

3

4

5

6

7

**Months**

***Experiment 3: Chemical Disruption*:** Our third experiment tests for the effects of chemical interventions that may slow or eliminate the eco-evolutionary dynamics driving Phase 1 and/or Phase 2. This will involve 25 mice divided into 5 treatment groups. All mice will receive the Standard Procedure with the following adjustments. One treatment will be the Standard Procedure. The four experimental treatments will involve all combinations of administering one of two drugs for either the first two months (during UV exposure) or during months 4 & 5 (during the period following UV exposure). Drugs will be applied topically to one half of either the no-UV or UV half of the mouse’s skin. In this way, the back of the mouse is now divided into four quadrats representing all combinations of the presence and absence of UV, and the presence and absence of the topical agent **(Fig. 8)**. We will continue to collect 8 images per mouse per month, and now the images will be distributed to insure 2 images per quadrat. One drug will be the anti-inflammatory diclofenac, a COX2 inhibitor35. Diclofenac suppresses the synthesis of prostaglandin PGE2, a well-established mediator of UV-induced inflammation, the proliferative adaptive response in epidermis, and driver of cuSCC formation31-34. The second drug will be a chemotherapy agent, the novel MEKi NFX179 for which we have access NFlection Therapeutics (see Letter). It targets proliferating cells with high ERK activity8,72. We hypothesize that diclofenac will be more effective at suppressing the emergence of very large clades during Phase 1 (exogenous control of clade expansions) and less so during Phase 2 (endogenous control of clade expansions). We expect the opposite for the MEKi. These data will test predictions or Aims 1 & 2 and identify any distinctive features of Phase 1 and Phase 2 dynamics and molecular & genetic features. As lesions form, this experiment will provide unique cancerous cell lines (Phase 3) with potentially unique properties where we can contrast and compete the cancer lines emerging from the non-perturbed and perturbed quadrats of a mouse.

**Figure 8. Experimental Plan To Test Effects of Specific Perturbations in Phase 1 and 2 on Carcinogenesis**. The standard UV-regimen is depicted here, as before, aligned to predicted phases. The perturbations hypothesized to have the phase-specific positive or negative effects on carcinogenesis will be applied in phase 1 or phase 2. The backs of the mice are now quadrisected to preserve UV-exposed vs. UV-protected internal controls but also areas for control vs. active perturbations to be introduced in the skin at the various phases. This maximizes the ability to use within-mouse controls.

**UV Dose**

**Regimen**

**Phase**

I

II

III

UV-protected

UV-exposed

0

1

2

3

4

5

6

7

**Months**

**0** = untreated

**+** = treated

**Perturbations**

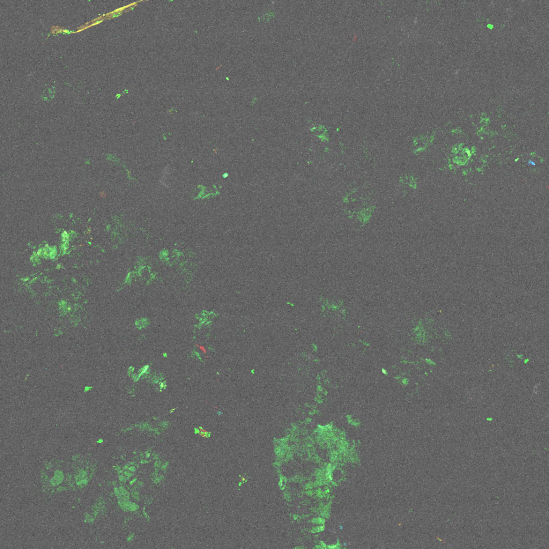
0

0

+

+

***Experiment 4: Genetic Disruption*:** Our fourth experiment tests for the effects of CRISPR interventions that may accelerate the eco-evolutionary dynamics by introducing well-studied driver mutations. The experimental design parallels that of **Experiment 3**. This will involve 25 mice divided into 5 treatment groups. One treatment will be the Standard Procedure. The four experimental treatments involve all combinations of intradermal injection of one of two CRISPR/Cas9 constructs administered either at 0 months (beginning of UV exposure) or at 3 months (end of UV exposure). The two constructs will inactivate either *Trp53* or *Notch1*. These are the two most frequently inactivated tumor suppressors in cuSCC8,22,73-75. These vectors have already been constructed and validated (pLV[CRISPR]-hCas9:T2A:EGFP). We have previously engineered CRISPR/Cas9 constructs to inactivate *Trp53* using sub-lethal (immunosuppressive) gamma irradiation followed by intradermal injection of 100,000 viral particles (**Fig. 9**). A unique opportunity of using intradermal injections is that we can apply the constructs to just one half of either the no-UV or UV half of the mouse’s skin. In this way, the back of the mouse is divided into four quadrats representing all combinations of the presence and absence of UV, and the presence and absence of the CRISPR construct **(Fig. 8)**. We will collect 8 images per mouse per month (2 images per quadrat). We hypothesize that the *Notch1* inactivation (by virtue of disrupting differentiation) will be more effective at accelerating the emergence of very large clades during Phase 1 (breaking exogenous controls of clade expansions) and less so during Phase 2 (endogenous controls already in place). We expect the opposite for *Trp53* inactivation, which we hypothesize is more relevant for facilitating acquisition of genetic changes in phase 2. This coincides with data showing that *NOTCH1* inactivation is likely an early event in human cuSCC pathogenesis with mutations in *TP53* occurring slightly later22,23. These data will test predictions of Aims 1 & 2 and identify distinctive features of Phase 1 & 2 dynamics and molecular & genetic features. As lesions form, this experiment will provide unique cancerous cell lines (Phase 3) exhibiting the direct and indirect properties conferred by inactivation. We can contrast and compete cancer lines emerging from the mouse’s non-perturbed and perturbed quadrats.



**Figure 9. Intradermal CRISPR Injection Can Effectively Localize Tumor Formation in Mice.** On the left, the CRISPR construct targeting *Trp53* labels infected cells (keratinocytes) green. Following UV-exposure, tumors localize in areas previously injected with CRISPR.

**Aim 1: Characterize the effects of UV-mediated tissue disruption on clonal dynamics**

*Rationale:* UV exposure disrupts skin as a mutagen and by killing or damaging cells. Ecologically, we hypothesize that restoring and maintaining homeostasis will require exogenous tissue signaling for clades to cooperatively fill space left from lost cells and supply cells for skin that would otherwise be vacant from the loss of progenitor or basal cells. Some clades will need to proliferate more and form larger volumes. For the molecular genetic, we hypothesize that the scRNAseq data, as analyzed to reveal cellular (differentiation, proliferative) states, transcriptional states, and genetic subpopulations (CNV) will reveal some diversity of transcriptional states and genetic subpopulations, but that the largest clades will fail to show decreased differentiation, oncogenic and proliferative transcriptional programs and enrichment for mutations in genes related to cuSCC pathogenesis.

*Approach*: Based on these hypotheses, we expect images in UV areas to show an initial decline in clade numbers, an increase in average clone size and an increase in the coefficient of variation in clade sizes, relative to non-UV areas. There should be no change in compaction, and little to no increase in inter-digitation of adjacent clones. All treatments in Experiment 2 and the early application of tamoxifen in Experiment 1 should show these ecological effects in months 1 and/or 2.

In Experiment 3, we expect the anti-inflammatory diclofenac to be more effective than MEKi in dampening or preventing the ecological effects of UV in months 1 & 2, and to cause a decrease in the number of lesions observed. We expect to see a balancing of clade sizes (with decreased variance) but importantly a more equitable distribution of cells undergoing apoptosis and proliferation (as assessed by cleaved caspase 3 and BrdU incorporation on sections) across multiple clades. In contrast to control conditions, we expect diclofenac-treated epidermis and clades to exhibit transcriptional and mutational heterogeneity closer to baseline unexposed samples as compared to control-treated UV-exposed samples.

In Experiment 4, we expect *Notch1* inactivation to accelerate or amplify the effects of UV more so than *Trp53* inactivation. In contrast to control conditions, we expect *Notch1*-inactivated clades to exhibit transcriptional and mutational heterogeneity, including enrichment for driver mutations or oncogenic signatures that should correlate with clade size. Within-mouse comparisons (**Fig. 8**) of different clade sizes permits a definitive test.

**Aim 2: Characterize the clonal dynamics of cancer initiation**

*Rationale*: A key novelty in our model of cancer initiation is the interplay between the ecological and evolutionary processes. In the first two months of UV, it is the ecological dynamics and not the evolutionary dynamics that propel the next phase of cancer initiation. Evolutionarily, we hypothesize that 3 months of UV exposure will create diverse and oncogenic-specific mutations that will begin to concentrate in the larger clades. Within clade heterogeneity will begin to supersede between clade heterogeneity. Driver mutations or oncogenic signatures will correlate with clade sizes. This heterogeneity should be substant by month 4. In phase 2, we hypothesize a feedback between clade size and genetic variability. Those experiments that increase CV, cell compaction, and inter-digitation should also promote genetic variability and the concentration of pathogenic mutations within larger clades. In molecular genetic terms, we hypothesize that the scRNAseq data, as analyzed to reveal cellular states, transcriptional states, and genetic subpopulations (CNV) will reveal a greater diversity of states and genetic subpopulations. Further, the largest clades will show decreased differentiation, more oncogenic and proliferative transcriptional programs, and enrichment for mutations in genes related to cuSCC pathogenesis.

In phase 2 some clades of progenitor and basal cells now service a larger volume of subcutaneous cells and a larger surface area of epidermis. They may also disproportionately contribute to additional progenitor or basal cells. In this way, these clades may shift from acting cooperatively with adjacent clades to competing with these clades resulting in some clades “stealing” and occupying space. The higher rates of cell turnover and proliferation in the larger clades permits selection for inter-clade competition and space pre-emption. The ecological and evolutionary dynamics become endogenously driven independent of and even in the absence of continued UV exposure. What began as varying ecological dynamics between clades can now drive evolutionary dynamics and selection for genomic drivers and pathogenic mutations. The transcriptional and mutational heterogeneity (relative to non-UV samples) as seen in the single cell sequencing will include driver mutations or oncogenic signatures, and these should correlate with clade sizes.

*Approach and Expected Results:* Ecologically, relative to months 1 & 2, we expect UV-exposed images in months 3, 4 & 5 of the Standard Procedure to show large increases in the coefficient of variation in clade sizes. There should now be slight to moderate increases in compaction by larger clades, and in the degree of inter-digitation of adjacent clades (clades are no longer necessarily good neighbors).

Experiment 1 sets the clock on tagging clades at different points during UV exposure. Consider the ecological features of clades revealed in UV samples at 3 months from mice receiving tamoxifen at two weeks before two months (call this A) relative to UV samples from mice at 1 month (B) or at 3 months (C) that received their tamoxifen two weeks before 0 months. In the absence of phase shifts, A = B. A change in phase would in terms of CV show C > B > A, in terms of the number of clones and inter-digitation A > B > C as branching points in larger clades, and in terms of compaction A = C > B.

In Experiment 2 just one month of UV should fail to establish the necessary amount of cell proliferation to propel inter-clad competition and selection. The CV induced by the UV should not increase in subsequent months. Two months may induce some inter-clade competition and CV may increase in subsequent months but less than if there had been 3 consecutive months of UV. Perhaps some mice from this treatment will go on to develop lesions. Alternating one month on with one month off to give 3 months of UV over 5 months should slow or prevent the onset of Phase 2, and the ecological patterns at 5 months should show little to no transition from the patterns expected from 1 or 2 months of UV exposure. In this setting, we expect that the largest clades even at 3-5 months to fail to reveal evidence for diversification, or enrichment for oncogenic programs or mutations.

In Experiment 3, we expect MEKi to be more effective than diclofenac in dampening or preventing the ecological effects of UV when applied during the 3rd and 4th months and to ultimately decrease the number of tumors. In contrast to control conditions, we expect MEKi-treated clades to exhibit transcriptional and mutational heterogeneity closer to baseline non-UV samples as compared to control-treated UV samples. We also expect larger clades to exhibit much less enrichment for oncogenic transcriptional programs or mutations (**Fig. 3**).

In Experiment 4, we expect CRISPR-mediated *Trp53* inactivation to accelerate or amplify the effects of UV more so than *Notch1* inactivation when these are introduced at month 2, and to cause an increase in the number of tumors observed. In contrast to MEKi-treated epidermis, we expect *Notch1*-inactivated epidermis and clades to exhibit transcriptional and mutational heterogeneity, including enrichment for driver mutations or oncogenic signatures, to be more correlated with clade size. Within-mouse comparisons (**Fig. 8**) will enable sampling of multiple clade sizes to test this definitively.

*Alternative approaches for Aims 1, 2*: We do not anticipate any technical challenges with this work, as we have been able to successfully interrogate the model system using all of the proposed methods. Furthermore, our experimental design including substantial within-mouse UV exposure and perturbation controls (Fig. 8), gives us considerable statistical power for distinguishing even small differences (10% or more) in the ecological (note the degree of resolution shown in Table 1 for results in Fig. 2) and evolutionary properties of clades including intra-clade and inter-clade molecular genetic heterogeneity. It is possible that the inactivation of a single gene (*Trp53* or *Notch1*) will fail to produce the necessary effects or to show clearly discernable effects in comparing those perturbations. To address this, we have already constructed point mutant *Trp53* alleles as well as a bi-specific *Trp53* / *Notch1* targeting CRISPR constructs designed to functionally inactivate both genes within the same cells. Finally, it is also possible that four colors limit the ability to distinguish clades at the desired resolution. An alternative strategy is to carry the Brainbow cassette in the homozygous state allowing for detection of 8 colors76.

**Aim 3: Identify mechanisms of multi-clonal eco-evolutionary dynamics in cancer development**

*Rationale*: Our preliminary data confirm that the majority of tumors we have imaged express one or two fluorophores. We hypothesize that tumor subpopulations from established lesions (phase 3) can be extracted and used to measure competition and cooperation *in-vivo* and *in-vitro*. Multiple lines will be derived, cultured, transplanted in combinations, followed over time as distinct clones, and their relative fitness linked to specific molecular and genetic traits. We can validate traits that may explain cell-cell cooperation and/or competition.

*Approach*: We will follow mice established for studies in Aims 1 & 2 for at least 2-4 months following the cessation of UV irradiation. We have previously extracted tumors of up to 1 cm in diameter using survival surgeries (permitted by our IACUC for up to three times). In the week prior to harvest, tamoxifen will be administered either topically (or systemically, required) to label tumor subpopulations. Tumors will be bisected in a way that permits maximal representation of all subpopulations. One half will be directly transplanted to NSG (NOD.Cg-*Prkdcscid Il2rgtm1Wjl*/SzJ) mice as per xenografting protocols for expansion and propagation. The other half will be used for baseline single cell RNAseq (10X) and low-input whole exome sequencing to identify mutations.

The extracted tumors will be maintained and expanded as xenografts in NSG mice. Multiclonal mixtures from tumor extracts will be grown in 3-D cultures formed by the admixture of fibroblasts within a collagen matrix following seeding of tumor cells. We have previously been successful in establishing cultures of SCC cell lines from SKH1-E Hairless mice. Initially, we will take xenografted tumors and administer tamoxifen to label as many clones as possible. To begin, we will culture 10 differentially fluorescing clades through serial culture in-vitro and attempt to reconstruct 1-3 subpopulation-containing tumors.

Each clade’s proliferation rate during exponential expansion can be measured in mono-cladal 3D cultures and xenografts. Bi-cladal tumors (Fig. 1, tumor) will be most straightforward, however, we will endeavor to extend this to at least tri-cladal tumors in order to test as many potential interactions in a tractable manner as possible. Fitness of overall 3D cultures containing varying proportions of 2 and 3 clades can be plotted against those of individual clades. Cultures will be grown until either one clade outcompetes or until strong evidence for coexistence. Further, some mixtures may grow faster than others. Once we have identified mixtures representing differentially fit tumors, we will profile them through serial imaging (Incucyte) and confocal microscopy (Leica SP5) to perform cell counts of each clade (e.g. each fluorophore) over time, and use EdU labeling in-situ as well as cleaved caspase 3 staining to measure proliferation and death rates per clone per tumor over time. These measurement are important as they may also suggest mechanisms of competition for example by induction of apoptosis in less fit clades and / or subsequent phagocytosis41. Indeed, some mixtures may ultimately prove unstable, leading to a dominant monocladal tumor (deterministically or stochastically) or elimination of one of three clades over time, for example.

Single cell sequencing data will focus on three specific approaches. First, receptor-ligand interaction analysis (CellPhoneDB65) applied to fluorescing cells and any adjacent cells as identified by imaging to identify co-regulated or co-expressed receptor ligand interaction. Specific molecular hypotheses to address would be to see if expression of homotypic molecules which mediate keratinocyte-keratinocyte adhesion are upregulated more within larger clades vs. between clades. Previous reports have implicated competition among basal layer clones by virtue of stochastic expression changes or decreases in expression of collagen VII induced by exogenous stressors40. So-called “loser” basal cells express less collagen VII, presumably enabling “winner” clones to establish better adhesion to the basement membrane at hemidesmosomes. In our case, interactions between keratinocytes within clades and between clades are expected to be vitally important. These interactions are mediated by desmosomes, highly specialized membrane domains that mediate keratinocyte-keratinocyte contact and adhesion77. These include desmogleins 1-4, desmocollins 1-3, plakoglobin, plakophilins 1-3, desmoplakin and E-cadherin / beta-catenin as the major protein components of the desmosome77. Finally, transcription factor based analysis on TRANSFAC motif-based analysis performed in conjunction with GSEA analysis from scRNAseq profiles can be used to identify upstream transcription factors driving these subclonal interactions (**Fig. 4**)8. We have already identified increased c-MYC activity in UV-exposed keratinocytes suggesting a potential driver of competition can be present by phase 2 (**Fig. 4**).

Importantly, the single cell data also allows us to discern differentiation and transcriptional states, as well as genetic subpopulations (CNV, mutations), enabling dissection of isolated cultured clades into distinct genetic subclones or distinct transcriptional states. If cells with specific states or CNV / mutations can then be subcloned, then we can additionally test intra-cladal dynamics *in-vitro* and *in-vivo*. These subclones can be mixed at will from distinct tumors from the same mouse or from distinct clades. In this way, the presence of specific CNV, mutations, or transcriptional signatures may also be directly linked to fitness individually and in combination with other subclones as artificially constructed tumors. It will of particular interest to see whether subclones from lesions derived from mice treated with chemical or genetic accelerants or suppressors of carcinogenesis retain their advantages or disadvantages when mixed with subclones of differing origins and life histories.

Although we have proposed here to deconstruct and reconstruct tumors *from the same stage of development* from mice, our system permits a novel approach whereby we may also establish multi-clonal tumors constructed from clones *isolated from different phases of development*. We expect that clades from established tumors will generally outcompete phase 2 clades. No Phase 1 clades should be able to persist. Nevertheless, we can test whether these phase 1 or phase 2 clades exhibit cooperativity or competition with established lesional clades or with fellow phase 1 and phase 2 clades as evidenced by coexistence and fitness facilitation.

*Validation of molecular drivers of cooperation and competition:* Once specific receptor-ligand pairs, transcription factors such as MYC or desmosomal proteins have been identified as over- or under-represented in dominant clades, we can test hypotheses regarding drivers of elevated or depressed fitness in the context of multiclonal tumors. As an example, lentiviral doxycycline-inducible overexpression or shRNA-mediated suppression of gene expression could be engineered into appropriate clones, remixed into the co-culture as above and fitness directly tested in-vivo and in-vitro upon doxycycline induction. Examples of signaling pathways previously implicated include Hippo/Yap, BMP / TGFbeta, WNT signaling have transcriptional signatures we have previously identified in skin SCC carcinogenesis8 and we will be able to implicate mechanical sensing through our examination of desmosome components. *In-vivo*, this could also be ultimately validating by using Cre-recombinase-driven null or activating alleles combined with spatiotemporal activation as in our Confetti model.

**RIGOR AND REPRODUCIBILITY.** As shown in table 1 (4-way ANOVA), 6 mice over 4 months generated 455 imaging samples that could successfully test for the effects of month and UV treatment on clone number, clone sizes and CV. The experimental design insures rigor and reproducibility. Experiments 1-4 allow for replicate intra-mouse and inter-mouse comparisons enabled by shielding half the backs from UV exposure and applying perturbations in only localized fashion (Fig. 8). In all cases, no-treatment controls are built into every experiment, thus ensuring that even as mice are culled as a result of having to biopsy them, we will get more than sufficient imaging and genomics data for statistical rigor, model fitting, and as importantly, have sufficient substrate to generate the multi-cladal tumors needed for Aim 3.

**SEX AS A BIOLOGICAL VARIABLE.** Our predictions are independent of sex; however, some sex effects have been reported for the SKH-1E background for cuSCC development and men have worse disease outcomes in cuSCC than women78. Our experimental design will explicitly test for potential sex effects by ensuring that all experiments include animals of both sexes in each treatment group.

**SUMMARY**. Our proposal addresses the key question in PQ3 by utilizing a combination of novel approaches in a highly validated UV-driven model of skin carcinogenesis. We leverage this to deeply characterize the dynamical changes in and interactions between clade size and composition, transcriptional states, and genetic subpopulations in the face of UV exposure while addressing the novel hypothesis that different clonal dynamics operate at different phases of cancer initiation. Furthermore, molecular determinants driving cooperation and competition between clades and between cells will be elucidated and validated. The work is being conducted by a multi-disciplinary team of investigators with appropriate and complementary expertise.

**REFERENCES**

1. Vogelstein, B. & Kinzler, K.W. The multistep nature of cancer. *Trends Genet* **9**, 138-41 (1993).

2. Vogelstein, B. *et al.* Genetic alterations during colorectal-tumor development. *N Engl J Med* **319**, 525-32 (1988).

3. Gatenby, R.A., Avdieiev, S., Tsai, K.Y. & Brown, J.S. Integrating genetic and non-genetic drivers of somatic evolution during carcinogenesis: the biplane model. . *Evolutionary Applications* **in press**(2020).

4. Gatenby, R.A. Is the Genetic Paradigm of Cancer Complete? *Radiology* **284**, 1-3 (2017).

5. Fuchs, E. Epidermal differentiation and keratin gene expression. *J Cell Sci Suppl* **17**, 197-208 (1993).

6. Rogers, H.W. *et al.* Incidence Estimate of Nonmelanoma Skin Cancer in the United States, 2006. *Arch Dermatol* **146**, 283-287 (2010).

7. Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880-6 (2015).

8. Chitsazzadeh, V. *et al.* Cross-species identification of genomic drivers of squamous cell carcinoma development across preneoplastic intermediates. *Nat Commun* **7**, 12601 (2016).

9. Alexandrov, L.B. & Stratton, M.R. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr Opin Genet Dev* **24**, 52-60 (2014).

10. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-21 (2013).

11. Muradova, E. *et al.* Non-invasive assessment of epidermal genomic markers of UV exposure in skin. *J Invest Dermatol* **in press**(2020).

12. Murai, K. *et al.* Epidermal Tissue Adapts to Restrain Progenitors Carrying Clonal p53 Mutations. *Cell Stem Cell* **23**, 687-699 e8 (2018).

13. Kemp, C.J., Donehower, L.A., Bradley, A. & Balmain, A. Reduction of p53 gene dosage does not increase initiation or promotion but enhances malignant progression of chemically induced skin tumors. *Cell* **74**, 813-22 (1993).

14. Colom, B. *et al.* Spatial competition shapes the dynamic mutational landscape of normal esophageal epithelium. *Nat Genet* (2020).

15. Jonason, A.S. *et al.* Frequent clones of p53-mutated keratinocytes in normal human skin. *Proc Natl Acad Sci U S A* **93**, 14025-9 (1996).

16. Williams, M.J. *et al.* Measuring the distribution of fitness effects in somatic evolution by combining clonal dynamics with dN/dS ratios. *Elife* **9**(2020).

17. Premi, S. *et al.* Genomic sites hypersensitive to ultraviolet radiation. *Proc Natl Acad Sci U S A* (2019).

18. Potten, C.S., Saffhill, R. & Maibach, H.I. Measurement of the transit time for cells through the epidermis and stratum corneum of the mouse and guinea-pig. *Cell Tissue Kinet* **20**, 461-72 (1987).

19. Hsieh, E.A., Chai, C.M., de Lumen, B.O., Neese, R.A. & Hellerstein, M.K. Dynamics of keratinocytes in vivo using HO labeling: a sensitive marker of epidermal proliferation state. *J Invest Dermatol* **123**, 530-6 (2004).

20. Agrawal, N. *et al.* Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* **333**, 1154-7 (2011).

21. Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**, 1157-60 (2011).

22. Wang, N.J. *et al.* Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma. *Proc Natl Acad Sci U S A* **108**, 17761-6 (2011).

23. Durinck, S. *et al.* Temporal dissection of tumorigenesis in primary cancers. *Cancer Discov* **1**, 137-43 (2011).

24. Cancer Genome Atlas Research, N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519-25 (2012).

25. Bass, A.J. *et al.* SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat Genet* **41**, 1238-42 (2009).

26. Cancer Genome Atlas, N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576-82 (2015).

27. Klein, A.M., Brash, D.E., Jones, P.H. & Simons, B.D. Stochastic fate of p53-mutant epidermal progenitor cells is tilted toward proliferation by UV B during preneoplasia. *Proc Natl Acad Sci U S A* **107**, 270-5 (2010).

28. Zhang, W. *et al.* UVB-induced apoptosis drives clonal expansion during skin tumor development. *Carcinogenesis* **26**, 249-57 (2005).

29. Zhang, W., Remenyik, E., Zelterman, D., Brash, D.E. & Wikonkal, N.M. 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* **98**, 13948-53 (2001).

30. Chao, D.L., Eck, J.T., Brash, D.E., Maley, C.C. & Luebeck, E.G. Preneoplastic lesion growth driven by the death of adjacent normal stem cells. *Proc Natl Acad Sci U S A* **105**, 15034-9 (2008).

31. Brouxhon, S. *et al.* Deletion of prostaglandin E2 EP2 receptor protects against ultraviolet-induced carcinogenesis, but increases tumor aggressiveness. *J Invest Dermatol* **127**, 439-46 (2007).

32. Kang-Rotondo, C.H., Miller, C.C., Morrison, A.R. & Pentland, A.P. Enhanced keratinocyte prostaglandin synthesis after UV injury is due to increased phospholipase activity. *Am J Physiol* **264**, C396-401 (1993).

33. Konger, R.L., Scott, G.A., Landt, Y., Ladenson, J.H. & Pentland, A.P. Loss of the EP2 prostaglandin E2 receptor in immortalized human keratinocytes results in increased invasiveness and decreased paxillin expression. *Am J Pathol* **161**, 2065-78 (2002).

34. Pentland, A.P., Mahoney, M., Jacobs, S.C. & Holtzman, M.J. Enhanced prostaglandin synthesis after ultraviolet injury is mediated by endogenous histamine stimulation. A mechanism for irradiation erythema. *J Clin Invest* **86**, 566-74 (1990).

35. Merk, H.F. Topical diclofenac in the treatment of actinic keratoses. *Int J Dermatol* **46**, 12-8 (2007).

36. Tomasetti, C., Vogelstein, B. & Parmigiani, G. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc Natl Acad Sci U S A* **110**, 1999-2004 (2013).

37. Gatenby, R.A., Cunningham, J.J. & Brown, J.S. Evolutionary triage governs fitness in driver and passenger mutations and suggests targeting never mutations. *Nat Commun* **5**, 5499 (2014).

38. Andor, N. *et al.* Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med* **22**, 105-13 (2016).

39. Di Gregorio, A., Bowling, S. & Rodriguez, T.A. Cell Competition and Its Role in the Regulation of Cell Fitness from Development to Cancer. *Dev Cell* **38**, 621-34 (2016).

40. Liu, N. *et al.* Stem cell competition orchestrates skin homeostasis and ageing. *Nature* **568**, 344-350 (2019).

41. Ellis, S.J. *et al.* Distinct modes of cell competition shape mammalian tissue morphogenesis. *Nature* **569**, 497-502 (2019).

42. Madan, E. *et al.* Flower isoforms promote competitive growth in cancer. *Nature* **572**, 260-264 (2019).

43. Gatenby, R.A., Zhang, J. & Brown, J.S. First Strike-Second Strike Strategies in Metastatic Cancer: Lessons from the Evolutionary Dynamics of Extinction. *Cancer Res* **79**, 3174-3177 (2019).

44. Stankova, K., Brown, J.S., Dalton, W.S. & Gatenby, R.A. Optimizing Cancer Treatment Using Game Theory: A Review. *JAMA Oncol* **5**, 96-103 (2019).

45. Kim, E. *et al.* Senescent fibroblasts in melanoma initiation and progression: an integrated theoretical, experimental, and clinical approach. *Cancer Res* **73**, 6874-85 (2013).

46. Picco, N., Sahai, E., Maini, P.K. & Anderson, A.R.A. Integrating Models to Quantify Environment-Mediated Drug Resistance. *Cancer Res* **77**, 5409-5418 (2017).

47. Cai, D., Cohen, K.B., Luo, T., Lichtman, J.W. & Sanes, J.R. Improved tools for the Brainbow toolbox. *Nat Methods* **10**, 540-7 (2013).

48. Vasioukhin, V., Degenstein, L., Wise, B. & Fuchs, E. The magical touch: genome targeting in epidermal stem cells induced by tamoxifen application to mouse skin. *Proc Natl Acad Sci U S A* **96**, 8551-6 (1999).

49. Vin, H. *et al.* Sorafenib Suppresses JNK-Dependent Apoptosis through Inhibition of ZAK. *Mol Cancer Ther* **13**, 221-9 (2014).

50. Vin, H. *et al.* BRAF inhibitors suppress apoptosis through off-target inhibition of JNK signaling. *Elife* **2**, e00969 (2013).

51. Gonzalez, R.C. & Woods, R.E. *Digital Image Processing (3rd Edition)*, (Prentice-Hall, Inc., 2006).

52. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man and Cybernetics* **9**, 62-66 (1979).

53. Gonzalez, R.C. & Woods, R.E. *Digital image processing*, xvi, 1168 pages (Pearson, New York, NY, 2018).

54. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **9**, 62-66 (1979).

55. Reeves, M.Q., 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* **20**, 699-709 (2018).

56. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods* **11**, 163-6 (2014).

57. Joost, S. *et al.* Single-Cell Transcriptomics of Traced Epidermal and Hair Follicle Stem Cells Reveals Rapid Adaptations during Wound Healing. *Cell Rep* **25**, 585-597 e7 (2018).

58. Joost, S. *et al.* Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity. *Cell Syst* **3**, 221-237 e9 (2016).

59. Bowling, S., Lawlor, K. & Rodriguez, T.A. Cell competition: the winners and losers of fitness selection. *Development* **146**(2019).

60. Andor, N. *et al.* Single-cell RNA-Seq of follicular lymphoma reveals malignant B-cell types and coexpression of T-cell immune checkpoints. *Blood* **133**, 1119-1129 (2019).

61. Andor, N. *et al.* Joint single cell DNA-seq and RNA-seq of gastric cancer cell lines reveals rules of in vitro evolution. *NAR Genom Bioinform* **2**, lqaa016 (2020).

62. Chen, J. *et al.* Single-cell transcriptome analysis identifies distinct cell types and niche signaling in a primary gastric organoid model. *Sci Rep* **9**, 4536 (2019).

63. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309-313 (2016).

64. Li, J., Smalley, I., Schell, M.J., Smalley, K.S.M. & Chen, Y.A. SinCHet: a MATLAB toolbox for single cell heterogeneity analysis in cancer. *Bioinformatics* **33**, 2951-2953 (2017).

65. Efremova, M., Vento-Tormo, M., Teichmann, S.A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* **15**, 1484-1506 (2020).

66. Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).

67. Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res* **42**, D472-7 (2014).

68. Bravo, R.R. *et al.* Hybrid Automata Library: A flexible platform for hybrid modeling with real-time visualization. *PLoS Comput Biol* **16**, e1007635 (2020).

69. Schenck, R.O. *et al.* How Homeostasis Limits Keratinocyte Evolution. *BioRxiv* (2019).

70. Teixeira, V.H. *et al.* Stochastic homeostasis in human airway epithelium is achieved by neutral competition of basal cell progenitors. *Elife* **2**, e00966 (2013).

71. Hall, M.W.J., Jones, P.H. & Hall, B.A. Relating evolutionary selection and mutant clonal dynamics in normal epithelia. *J R Soc Interface* **16**, 20190230 (2019).

72. Adelmann, C.H. *et al.* MEK Is a Therapeutic and Chemopreventative Target in Squamous Cell Carcinoma. *J Invest Dermatol* **136**, 1920-4 (2016).

73. South, A.P. *et al.* NOTCH1 mutations occur early during cutaneous squamous cell carcinogenesis. *J Invest Dermatol* **134**, 2630-8 (2014).

74. Li, Y.Y. *et al.* Genomic analysis of metastatic cutaneous squamous cell carcinoma. *Clin Cancer Res* **21**, 1447-56 (2015).

75. Pickering, C.R. *et al.* Mutational landscape of aggressive cutaneous squamous cell carcinoma. *Clin Cancer Res* **20**, 6582-92 (2014).

76. Tang, Y.J. *et al.* Tracing Tumor Evolution in Sarcoma Reveals Clonal Origin of Advanced Metastasis. *Cell Rep* **28**, 2837-2850 e5 (2019).

77. Kowalczyk, A.P. & Green, K.J. Structure, function, and regulation of desmosomes. *Prog Mol Biol Transl Sci* **116**, 95-118 (2013).

78. Thomas-Ahner, J.M. *et al.* Gender differences in UVB-induced skin carcinogenesis, inflammation, and DNA damage. *Cancer Res* **67**, 3468-74 (2007).