Title
Tumor Heterogeneity and Clonal Evolution of Metastatic Skin Cancer

Author
McCreery, Melissa Q.

Publication Date
2017

Peer reviewed|Thesis/dissertation

Tumor Heterogeneity & Clonal Evolution of Metastatic Skin Cancer

by

Melissa Quino McCreery Reeves

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Acknowledgements

No graduate student completes his or her work without an incredible amount of support from advisors, mentors, family, and friends. I would like to thank in particular a few of those people who have made this journey not only possible for me, but quite enjoyable.

I would like to thank first and foremost my graduate advisor, Dr. Allan Balmain, for his guidance and ideas, for letting me play with the Confetti mouse, and for his support during my graduate career and beyond. I would like to also thank my thesis committee, Drs. Boris Bastian and Eric Collisson, as well as Dr. Rosemary Akhurst for being a co-mentor.

To my family and friends, I don't have the space to thank each of you who has supported me, but thank you all the same. I would like to particularly thank my parents, for encouraging me when going to graduate school seemed like a monumental career change that was, to say the least, a surprise to everyone around me. I would also like to thank my husband, Jon Reeves, who has been there for countless small moments, victories and discouragements alike, and who has never wavered in his unconditional support and belief in me.

On a more formal note, two of the chapters that follow are taken from published works, and I would like to acknowledge my collaborators and co-authors on each of those projects, as well as the journals in which they are published.

Chapter 1 was originally published in Nature Medicine in 2015: Melissa Q McCreery, Kyle D Halliwill, Douglas Chin, Reyno Delrosario, Gillian Hirst, Peter Vuong, Kuang-Yu Jen, James Hewinson, David J Adams & Allan Balmain. Evolution of metastasis revealed by mutational landscapes of chemically induced skin cancers. *Nature Medicine* **21,** 1514–1520 (2015).

Chapter 3 was originally published in the Annual Review of Cancer Biology in 2017: Melissa Q McCreery & Allan Balmain. Chemical Carcinogenesis Models of Cancer: Back to the Future. *Annual Review of Cancer Biology* **1,** 285–312 (2017).

Abstract

Tumors acquire genetic alterations throughout their life, leading them to evolve and to accumulate heterogeneity over time. These characteristics make cancer a particularly complex and challenging disease to treat, as tumor heterogeneity has been linked to drug resistance and overall worse outcomes for patients. Using the 4-color lineage tracing Confetti mouse and the DMBA/TPA chemical carcinogenesis model of skin cancer, clonal and evolutionary dynamics were investigated at every stage of tumor development, from initiation to malignant progression and metastasis. The initiating mutagen, DMBA (dimethylbenzanthracene), leaves a characteristic fingerprint of A>T mutations in each tumor genome, and thus next generation sequencing of these tumors enabled detailed dissection of tumor clones and evolution, and relationships between primary tumors and their respective metastases.

It was found that, while benign tumors can be comprised of multiple cellular populations, only one population contains an initiating A>T mutation in *Hras*. Thus benign tumors are clonal in origin, but surprisingly can subsequently recruit neighboring “normal” skin cells into the tumor. Like initiation, progression to malignancy was found to be a clonal event, driven by the emergence of a single dominant progressing clone. In contrast to both of these processes, however, metastasis was seen not to be clonal—multiple cell populations from the same primary tumor were observed to contribute to metastases in the lymph nodes as well as at distant sites, suggesting that metastasis is a fundamentally different process than initiation or progression. Next generation sequencing analysis further demonstrated that most metastases departed from the primary tumor synchronously rather than traveling first to a lymph node, supporting the parallel rather than the linear model of metastasis. These findings highlight the usefulness of mouse models which faithfully recapitulate all stages of human disease, from benign tumor formation to metastatic spread, in understanding the evolutionary processes which take place in cancer.

Table of Contents

List of Tables

Chapter 1

Chapter 3

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Introduction

The discovery of the genetic origins of cancer—that cancer could be caused by aberrations in DNA sequence or architecture[1–3]—was made roughly concurrently with the discovery that the genomes of tumors are not static entities[4]. Mutations to specific genes, such as *RAS*[1,2] and *TP53*[5], were found to be strongly linked to cancer, as were gains, losses, and rearrangements of whole pieces of chromosomes[6,7]. But cancer cells did not merely acquire these genetic changes in a one-shot instance—rather, they accumulated these changes over time, beginning before the cancer was detectable and continuing throughout its life[4].

It was soon appreciated that this ongoing acquisition of genetic changes could be linked to tumor advancement, or progression[8]—for example, a tumor with only a *RAS* mutation might be benign, but become malignant after acquiring a second genetic alteration, such deletion of *CDKN2A* or *TP53*[9–11]. The accumulations of these mutations, however, was found to be neither linear nor entirely predictable, nor did it simply stop once a tumor reached malignancy. Rather tumors continue to accumulate genomic changes and evolve indefinitely, in a process that has been likened to the Darwinian model of the evolution of species[4,12].

Models for tumor evolution were developed as early as the 1970s using karyotype analysis to visualize the changes in chromosomes of leukemias[4]. In the past decade, the boom of next-generation sequencing has enabled the sequencing of thousands of human tumors, and data from these studies have robustly confirmed many of the early principles of tumor evolution[12]. Sequencing melanomas adjacent to the benign nevi they had arisen from has revealed that the malignant tumors, when compared to their benign counterparts, have acquired both specific progression-related mutations (frequently, deletion of *CDKN2A* or *PTEN*) as well as an overall 3-fold or higher increased burden of genetic alterations in general[13]. Many of these alterations have no known role in cancer to date, but serve to make the tumor a highly complex entity that is quite difficult to model or fully understand the behavior of. More surprisingly, sequencing of multiple regions of individual renal[14] and colorectal[15] tumors revealed that regions

could vary from each other by dozens of mutations. Thus tumors are not only evolving over time, but distinct regions can evolve separately, creating substantial heterogeneity within the tumor[12,14–16].

As more sequencing data has become available from patients, it has become possible to ask whether this intratumor heterogeneity actually matters to patients—and it does. Patients with highly heterogeneous tumors respond worse to therapy and have overall worse clinical outcomes[17]. Heterogeneity has been linked to drug resistance, as a small number of therapy-resistant cells in a tumor are sufficient to grow back into a drug-resistant tumor after therapy[12,18].

Understanding the patterns of tumor evolution and heterogeneity are critical to the study of cancer and to developing better strategies to combat it. We need a better grasp of how the ongoing acquisition of genetic changes in a tumor relates to clinically meaningful and observable changes in the tumor. We also need to better understand the patterns of heterogeneity in tumors, including how and when they arise and the dynamics that govern this heterogeneity. While evolution and heterogeneity are now well-documented phenomena, many aspects of exactly how they occur are still the subject of much debate.

These debates begin at the very origin of tumors. Tumors were historically believed to arise from a single cell that had acquired a mutation, and tumors were thus described as being monoclonal (or simply “clonal”) in origin[4,12]. This theory has been challenged by demonstrations that large patches of normal tissue can contain identical cancer-associated mutations, a phenomenon known as “field cancerization”[19], leading to the possibility that a tumor might arise from many of the cells in the “field” and instead be of polyclonal origin. Studies using chimeric mice have also shown that multiple cellular populations of distinct origins can be found in a tumor[20,21], lending weight to the polyclonal origin hypothesis. As the origin of a tumor defines many of the tumor’s attributes, and further because tumors of polyclonal origin will carry higher intrinsic heterogeneity than monoclonal tumors, this is a question that is important to resolve.

Debate also exists about the nature of progression of a tumor from a benign state to malignancy and to eventual metastatic dissemination to other anatomical sites. While a link has been documented between the accumulation of specific genetic alterations (such as *CDKN2A* or *TP53* loss) and progression to malignancy[9–11,13], it is not known precisely how this progression happens and whether what we observe as "progression" is the result of a single cell acquiring the right genetic changes, or the behavior of a many cells acting together. And while some genetic alterations that can drive a tumor from a benign to malignant state have been identified, our current catalog is far from complete.

Similarly, the nature of metastasis has been the subject of perhaps even greater controversy. Searches for "metastasis genes"—alterations which would drive a tumor to metastasize, the way that loss of *CDKN2A* or *TP53* might drive progression from benign to malignant—have largely been futile. Alternative theories have been proposed, that a tumor metastasizes as the result of changes in environmental factors[22] rather than specific genetic changes, opening the possibility that metastasis may be driven by entire groups of cells[23,24].

The pattern of metastasis is also debated. It was once accepted that when a tumor metastasized, it spread first to the nearby lymph nodes and then subsequently to more distant organs, in what has been described as a "linear" fashion. In contrast to this model, it has been proposed that metastasizing tumor cells might not necessarily travel via a lymph node, but rather disseminate directly to distant organs in parallel to spreading to the lymph nodes[25]. These models have direct clinical implications: if the former, so-called "linear model," is the predominant mode in which tumors metastasize, then it would make sense that a patient going into surgery to remove a tumor would also have any nearby lymph nodes removed—and this is currently frequently done. However if the second model, the "parallel model," of metastasis dominates, then lymph node removal is an unnecessary invasive surgery that will yield no clinical benefit. Gaining insight into the mode of metastasis dissemination is thus of direct and immediate clinical interest.

The uniting theme of the work that follows is a goal of dissecting tumor evolution at all stages of the tumor progression—from initiation to benign, malignant and metastatic—to shed light on these debates and to gain insight into the heterogeneity and evolutionary dynamics present at each of stage of the tumor. Throughout this work, a mouse model of chemical carcinogenesis has been used, in which mice develop benign skin tumors (papillomas) following treatment with the chemical carcinogen dimethylbenzanthracene (DMBA)[26]. Some papillomas progress to fully malignant tumors (carcinomas), and surgery to remove these carcinomas leads to prolonged survival and subsequent development of metastatic disease[27], mimicking the course of human disease and making it a highly suitable model of studying cancer progression. Further, the chemical initiation of these tumors is done with DMBA, a carcinogen which binds to adenosine (A) bases and leads to A>T mutations when these adenosine adducts are misrepaired. DMBA thus leaves signature of its damage in the genome, in the form of hundreds of A>T mutations. This characteristic signature, combined with the fact that each tumor has a unique "fingerprint" of its specific A>T mutations[28], proves to be extraordinarily useful as a tool to dissect the evolutionary processes in these tumors and to track relationships between primary skin carcinomas and metastases to distant organs.

In chapter 1 of this work, a cohort of 15 mice and over 100 tumors are used to investigate genomic changes at all stages of tumor progression. Using the DMBA A>T signature and tumors of different stages, it was possible to identify signature or "early" and "late" genomic alterations, giving insight into the natural course and mode of evolution in these tumors. Sequencing data also enabled interrogation of the relationship between primary tumors and the metastases they give rise to, revealing that parallel dissemination—not linear—is the dominant mode of metastasis in these tumors. The material in Chapter 1 is taken from my first author publication, "Evolution of metastasis revealed by mutational landscapes of chemically induced skin cancers," published in *Nature Medicine* in December 2015[28].

Chapter 2 builds on these findings, combining the DMBA carcinogenesis model with the Confetti mouse, an engineered mouse which contains a 4-color cassette, inserted in its DNA, that permits permanent, stochastic labeling of individual cells with one of the 4 colors and the tracing of their progeny[29]. Using this mouse, the skin is labeled with 4 colors (red, blue, green, and yellow) prior to treatment with DMBA, such that tumors subsequently arise from multi-color skin. While the tumors themselves are found to contain multiple, differently-colored cell populations, careful genetic analysis of these colored cells reveals that only one population in each tumor carries the expected cancer-driving *Ras* mutation. Thus despite being multi-colored, these tumors arise clonally from a single cell, and subsequently recruit nearby, normal skin cells into the tumor. In subsequent experiments, early and late benign tumors were labeled instead of the skin, and data from these experiments revealed that progression from a benign to malignant tumor is driven by a single, dominant cell population. Importantly, many papillomas in this model do not progress to carcinomas during the lifetime of the host. In these terminally benign papillomas, and in contrast to the carcinomas on the same mice, a dominant clone failed to develop, and each papilloma comprised several distinct clones. Evolution of clonal dominance is therefore a distinct requirement for malignant progression. Unlike tumor initiation and progression however, multiple cell populations can contribute to metastasis, suggesting that metastasis is a fundamentally different process.

Chapter 3 takes a step back, and presents the history of chemically induced animal tumor models. The findings from chapter 1 help to illuminate some of advantages of chemical carcinogenesis models, and place the DMBA model in relationship to other chemical models as well as genetically engineered mouse models. The material in this chapter was originally published in the inaugural issue of the *Annual Review of Cancer Biology* in 2017, under the title "Chemical Carcinogenesis Models of Cancer: Back to the Future"[30].

Finally, chapter 4 offers a short vignette in which the relationship between DMBA-initiated tumors and the immune system is explored. Cell lines derived from DMBA-induced

tumors were allowed to form tumors in mice and then treated with anti-PD-1 immunotherapy. It was found that tumors from certain cell lines were responsive to therapy, while tumors from other cell lines were not, and that responsiveness was correlated with the number of mutations in the original cell line. Cell lines with large numbers of mutations responded to therapy while those with few mutations did not, in a pattern paralleling the anti-PD-1 responsiveness of human lung tumors[31]. This suggests that mouse models which carry high mutation burdens will be particularly well-suited to the study of immunotherapy. This material is excerpted and adapted from a publication in progress, focusing on my particular contributions to that project and the elements most relevant to the studies described in the preceding chapters.

Chapter 1

Evolution of metastasis revealed by mutational landscapes of chemically induced skin cancers

ABSTRACT

Human tumors show a high level of genetic heterogeneity, but the processes that influence the timing and route of metastatic dissemination of subclones are unknown. Here, we have used whole-exome sequencing of 103 matched benign, malignant, and metastatic skin tumors from genetically heterogeneous mice to demonstrate that most metastases disseminate synchronously from the primary tumor, supporting parallel rather than linear evolution as the predominant model of metastasis. Shared mutations between primary carcinomas and their matched metastases have the distinct A>T signature of the initiating carcinogen Dimethylbenzanthracene (DMBA), but non-shared mutations are primarily G>T substitutions, a signature associated with oxidative stress. The existence in the same host animal of carcinomas that either did or did not metastasize suggests that there are tumor-intrinsic factors that influence metastatic seeding. We also demonstrate the importance of germline polymorphisms in determining allele-specific mutations, and identify somatic genetic alterations that are specifically related to initiation of carcinogenesis by *Hras* or *Kras* mutations. Mouse tumors that mimic the genetic heterogeneity of human cancers can aid our understanding of clonal evolution of metastasis and provide a realistic model for testing of novel therapies.

INTRODUCTION

Next generation sequencing technologies have revolutionized our understanding of the genomic landscapes of human cancers[32,33], but we have only a poor appreciation of the genetic determinants of metastasis—the leading cause of human cancer death[34]. Metastases may spread linearly, from one organ site to the next, or in parallel, each departing separately from the primary tumor[25], but lack of appropriate human matched primary carcinomas and

metastases has hindered analysis of these questions. We took advantage of a mouse skin tumor model that encompasses genetic and environmental factors, and for which all stages from benign lesions to metastases are available[27]. We show here that chemically induced skin tumors display a diversity of point mutations (single nucleotide variants, SNVs) and gene copy number variants (CNVs) that permit a detailed analysis of clonal evolution from the initiated cell to metastasis.

To mimic the extreme germline and somatic genetic heterogeneity in human samples, we bred *Mus musculus* mice (FVB/N) with a wild-derived strain, *Mus spretus* (SPRET/EiJ), to create a heterogeneous backcross population (henceforth FVBBX) (**Fig. 1a**), and carried out chemical carcinogenesis with dimethylbenzanthracene (DMBA) and 12-O-tetradecanoylphorbol-13-acetate (TPA). Squamous carcinomas (SCCs) in this model almost always carry *Hras* mutations[26], which is also a driver of squamous carcinogenesis in human head and neck[35] and skin[36]. In order to compare routes to carcinoma development driven by *Hras* or *Kras*[27,37], which is mutated in many human cancers[38], we also bred *Hras*$^{-/-}$ FVBBX mice that develop aggressive tumors carrying *Kras* mutations[27]. To further mimic human clinical practice, we surgically removed primary carcinomas, allowing continued survival and subsequent harvesting of distant metastases (**Fig. 1a**)[27]. We selected 103 tumor samples from 15 animals (9 wild-type and 6 *Hras*$^{-/-}$ mice), for which matched normal tissue, benign papillomas, carcinomas, and metastases were available, and carried out whole exome-sequencing and gene expression microarrays.

RESULTS

Mutation spectrum of early and late mutations

The predominant mutation found in tumors from wild type FVBBX mice was the classical A>T transversion in the *Hras* gene[26,39], leading to Q61L mutations (58 tumors). Using a nested PCR approach[40] this mutation can be detected as early as one week after DMBA treatment, but the skin remains morphologically normal for periods up to one year in the absence of further

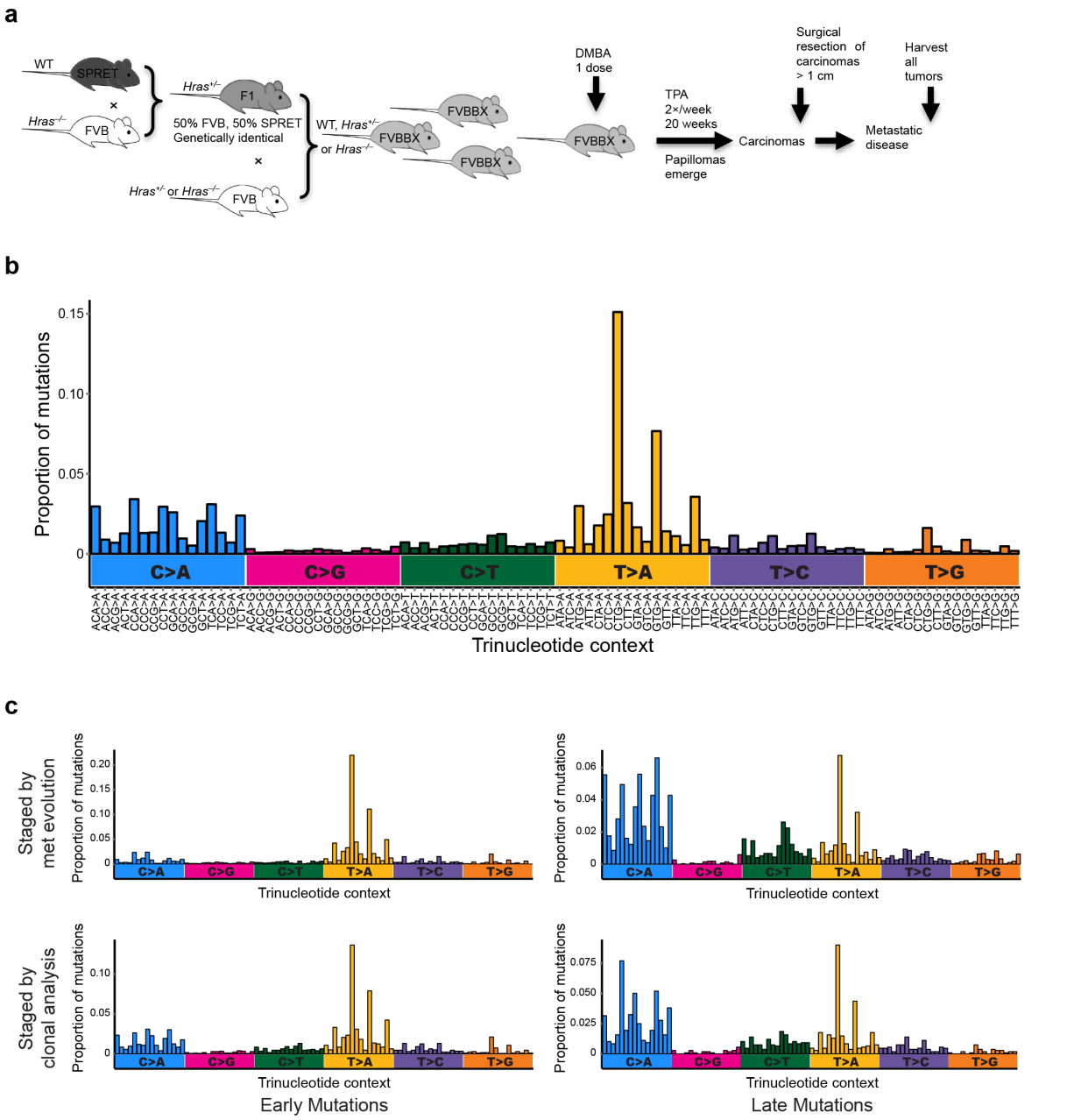


Figure 1. Chemically induced tumors carry a mutation signature of the carcinogen DMBA. (**a**) Genetically unique backcross mice (FVBBX) were generated by crossing FVB/N *Hras*$^{-/-}$mice to SPRET/EiJ mice, and then crossing again to inbred wild-type (WT) or *Hras*$^{-/-}$ FVB/N mice. Tumors were induced in FVBBX mice, carcinomas were resected upon reaching > 1 cm in longest diameter, and mice were allowed to progress to metastatic disease. (**b**) Frequency of mutation observed in each of 96 possible trinucleotide mutation contexts for all mutations across all tumors. Trinucleotide contexts are arranged on horizontal axis grouped by base pair change of the mutation. Peaks are observed at two specific contexts, CTG>CAG and GTG>GAG. The same pattern is observed when only nonsynonymous mutations are considered (not shown). (**c**) Mutations were classified as "early mutations" (left panels) or "late mutations" (right panels) by two strategies. In the upper panels, mutations in metastases are classified by whether they are shared with the ancestral primary tumor (early, upper left panel) or not shared (late, upper right panel). In the lower panels, mutations in tumors of all stages were classified based on whether they were fully clonal (i.e., present in all tumor cells; early, lower left panel) or subclonal (i.e., present only in a fraction of tumor cells; late, lower right panel). Early mutations show an enrichment of T>A mutations. Later mutations show a higher proportion of G>T mutations.

treatment with TPA (12-*O*-tetradecanoylphorbol-13-acetate)[41]. Tumors from *Hras*$^{-/-}$ mice had *Kras* mutations at a range of hotspot sites—G12D, G13R, Q61L, and Q61H (26 tumors). Both *Hras* and *Kras* mutations showed a strong preference for the FVB allele (84% and 100% specificity, respectively) in mice heterozygous at these loci ($P < 0.001$) consistent with previous work indicating allelic preference for *Ras* mutations in mouse models[42,43]. A single mouse carried no *Hras* or *Kras* mutations, but had several tumors with a mutation in PI3-Kinase alpha (*Pik3ca)* at the most common hotspot (H1047L) found in human cancers[44].

Tumors carried an average of 237 single nucleotide variants (SNVs) (5.2 mutations/megabase (Mb)), similar to the mutational burden of human adenocarcinomas and squamous carcinomas of the lung[45]. Papillomas harbored fewer mutations than did primary carcinomas (172 vs. 284 mutations on average, respectively, $P = 0.01$) although all samples from each individual mouse were harvested at euthanasia. The genome-wide mutation spectrum across all 103 tumors showed an enrichment in DMBA signature transversion mutations (A>T or T>A, hereafter A>T; 45% of all mutations), predominantly at 2 of the 96 possible trinucleotide contexts (**Fig. 1b**)[46].

Mutations in metastases that were shared with a matched primary tumor showed an enrichment of A>T DMBA signature mutations (63% of all mutations). Subsequent nonshared mutations were predominantly C>A or G>T mutations (hereafter G>T; 50%, compared to 14% of shared mutations, **Fig. 1c**). Subclonal mutations in tumors of all stages were also much more likely to be G>T than A>T substitutions (**Fig. 1c**). The average mutant allele fraction of A>T mutations was significantly higher than that of G>T mutations (0.265 vs. 0.190, $P = 2.2 \times 10^{-16}$), consistent with A>T mutations being early and clonal, and G>T mutations occurring later and more frequently being subclonal.

Evolutionary trees reveal patterns of metastatic dissemination

We constructed phylogenetic trees[47] demonstrating the evolutionary relationships among papillomas, carcinomas, and metastases from each mouse (**Fig. 2**). In 7 of 8 cases in which multiple metastases originated from a single primary carcinoma, the number and identity of mutations shared with the primary carcinoma were almost identical between metastases, suggesting that dissemination had occurred synchronously and in parallel. However, both primary tumor and metastases continued to evolve, and each tumor accumulated many private mutations following divergence.

Tumor cells can disseminate at an early stage, before the evolution of overt carcinomas[25], but the relationship between these early circulating cells and outgrowth of metastasis is still unclear. In the skin model, metastasis occurred at different points during tumor evolution, in some cases diverging after many mutations had accumulated and in others after relatively few. In mouse 1664 (**Fig. 2a**), 46% of total SNVs, and only 17% of G>T substitutions, were shared between Carcinoma A and its four metastases. By contrast, in mouse 1383 (**Fig. 2b**) nearly 88% of SNVs, and 66% of G>T substitutions, were shared among Carcinoma A and its four metastases. Other mice had more intermediate distributions (**Fig. 2c–f**). We conclude that while dissemination may be able to occur at an early stage, additional factors determine whether these lead to seeding and growth of metastases. It also seems that the timing of dissemination is unrelated to the total mutational burden of the primary tumor.

These data also provide information on possible routes of dissemination. If metastasis occurs first to a regional lymph node and subsequently to distant sites, as proposed in the "linear evolution" model[25], the metastases should be more genetically related to each other than to the primary carcinoma. This is in fact seen in only one animal (Case 1949, **Fig. 2f**). In most cases, each metastasis was genetically distinct, supporting a model of independent, parallel spreading from the primary tumor. Further, metastases to lymph nodes do not always depart first. In mouse 1407, the spleen metastasis departs first (**Fig. 2c**).

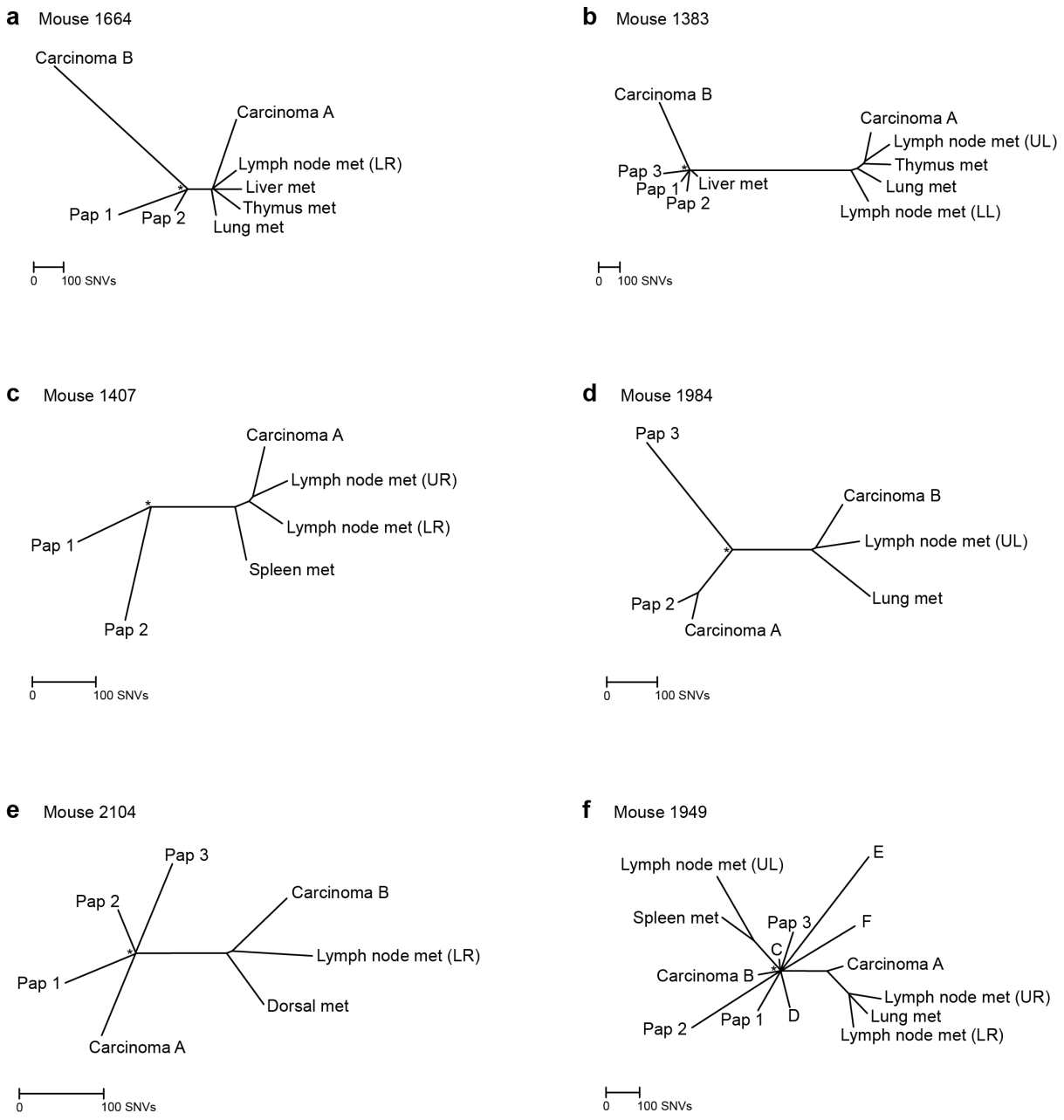


Figure 2. Phylogenetic trees reveal evolutionary relationship between tumors. (**a**) In mouse 1664, Carcinoma A had 4 metastases that diverged early, sharing an average of 46% of mutations (83 mutations), whereas carcinoma B did not form any detectable metastases. Distant metastases to the lung and thymus do not show evidence of disseminating from a lymph node, but rather appear to diverge from the primary carcinoma at approximately the same time as the lymph node metastases. An asterisk (*) denotes normal tissue (root of tree). (**b**) In mouse 1383, Carcinoma A had 4 metastases that diverged relatively late and approximately synchronously, sharing an average of 88% of mutations (822 mutations) with the primary carcinoma**.** (**c**) In mouse 1407, two lymph node metastases depart synchronously from Carcinoma A, and a spleen metastasis diverges slightly earlier. Tumors share an average of 69% of mutations (153 mutations). (**d**) In mouse 1984, a lymph node metastasis and lung metastasis diverge from Carcinoma B synchronously, sharing an average of 58% of mutations (160 mutations). (**e**) In mouse 2104, a lymph node metastasis and dorsal metastasis diverge from Carcinoma B synchronously, sharing an average of 61% of mutations (118 mutations). (**f**) In mouse 1949, two lymph node metastases and a lung metastasis share an average of 52% of mutations (167 mutations) with Carcinoma A, and an additional 90 mutations with one another suggesting one of the metastases may have given rise to the others. This mouse provides the only counterexample we observed to parallel evolution.

In some cases, carcinomas developed that showed no signs of metastasis, even in the presence of other metastatic carcinomas in the same animals (cases 1664, 1383, 2104, 1949, **Fig. 2**, and case 1508 not shown). This suggests that some carcinomas are intrinsically metastatic, while others are restrained from forming metastases at distant sites, possibly by immune surveillance[48], lack of appropriate "niche" factors, or other tumor-specific mechanisms[49]. Comparative analysis of mutations in metastasizing and non-metastasizing primary tumors did not reveal any obvious candidate metastasis driver mutations (see Methods). Gene expression analysis identified *Pdlim7* (also known as *LMP1* in humans, a known downstream target of transforming growth factor (TGF)-β1 signaling[50]) as more highly expressed in metastasizing compared to non-metastasizing primary tumors ($P = 6.4 \times 10^{-6}$). Several other genes were suggestive of being differentially expressed but did not meet the threshold for significance, including *Cd151* ($P = 1.4 \times 10^{-4}$), a cell surface glycoprotein involved in cell adhesion, integrin trafficking, and metastasis[51,52]. Expression levels of *Pdlim7* and *Cd151* are correlated with each other in carcinomas from this data set (Spearman's rank correlation coefficient (ρ) = 0.77, $P = 6.7 \times 10^{-6}$) as well as in an independent set of mouse carcinomas (ρ = 0.52, $P = 2.2 \times 10^{-5}$)[53] and in The Cancer Genome Atlas (TCGA) profiles of human head and neck cancers ($\rho = 0.67$)[54,55].

Mesenchymal-to-epithelial transition (MET) depends on metastatic site

Metastases to all organs except the lung—including lymph nodes, kidney, liver, spleen, and thymus—nearly always matched the squamous or spindle morphology of the primary tumor (21 of 22 cases, 95%). Metastases to the lung were almost always squamous (7 of 9 cases, 78%; one of the seven cases had mixed SCC and spindle morphology), even when they arose from spindle primary tumors. This finding is consistent with previous observations of metastases to the lung with squamous morphology in an inbred FVB/N model[27], however, we were able here to sequence-match these metastases to their respective primary tumors and demonstrate

Mouse	Primary Tumor	Primary Tumor Histology	Metastasis	Metastasis Histology
1383	Carcinoma A	SCC	Lymph Node (LL)	SCC
			Lymph Node (UL)	SCC
			Lung	SCC
			Thymus	SCC
1508	Carcinoma A	SCC	Lymph Node (UR)	SCC
1717	Carcinoma A	SCC	Lymph Node (UL)	SCC
2104	Carcinoma B	SCC	Lymph Node (LR)	SCC
			Dorsal	SCC
1383	Carcinoma B	spindle	Liver	SCC
1407	Carcinoma A	spindle	Lymph Node (UR)	spindle
			Lymph Node (LR)	spindle
			Spleen	spindle
1508	Carcinoma C	spindle	Lung	mixed
1664	Carcinoma A	spindle	Lymph Node (LR)	spindle
			Lung	spindle
			Thymus	spindle
			Liver	spindle
1815	Carcinoma A	spindle	Lymph Node (LR)	spindle
			Lymph Node (UR)	spindle
			Kidney	spindle
1984	Carcinoma B	spindle	Lymph Node (UL)	spindle
			Lung	SCC
1616	Carcinoma A	mixed	Lymph Node (LL)	SCC
1706	Carcinoma A	mixed	Lung	SCC
			Lymph Node (LL)	SCC
1949	Carcinoma A	mixed	Lymph Node (LR)	spindle
			Lymph Node (UR)	spindle
			Lung	SCC

Table 1. Matched primary tumors and metastases. Primary tumors with sequence-matched metastases, including histology information for primary and metastatic tumors. Most metastases show the same histology as their matched primary, with the exception of lung metastases which are nearly always squamous (SCC) regardless of the histology of the primary.

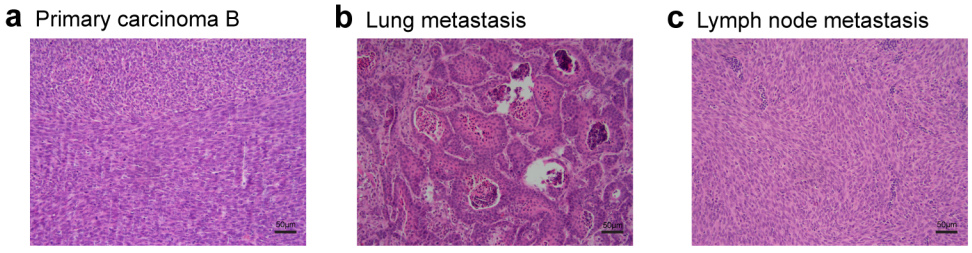


Figure 3. Histology of primary tumors and metastases. Histology of primary and metastastatic tumors from mouse 1984. (**a**) Primary carcinoma B shows spindled tumor cells arranged in fascicles (H&E, 200X). (**b**) Lung metastasis arising from carcinoma B is of squamous histology (H&E, 200X). (**c**) Lymph node metastasis arising from carcinoma B is of spindled histology (H&E, 200X).

that squamous lung metastases could arise from a primary spindle tumor, even when all other metastases from that tumor were spindle (**Fig. 3** and **Table 1**). These data suggest that the requirement for mesenchymal-to-epithelial transition (MET) for the outgrowth of metastases at

distant sites[56], which predicts epithelial-like squamous metastases, may in fact depend on the specific organ site: MET may be favored in the lung, but not in soft tissue sites where spindle metastases were frequently found.

Mutations shared with human SCCs

This analysis identified a large number of recurrently mutated genes, in addition to *Hras* and *Kras*, many of which are reported to be mutated in human SCCs of the head and neck, skin, and lung[35,36,57]. (**Fig. 4**). These included recurrent mutations in *Trp53*, *Fat1*, and the Notch signaling pathway, including *Notch1*, *Notch3*, and *Trp63* (**Fig. 4**). Dysregulation of Notch signaling, which is involved in epithelial differentiation, has been implicated in a variety of human SCCs[35,36,57]. The three mutations we observed in *Notch1* and *Notch3* were all in the N-terminal epidermal growth factor (EGF)-like repeat domains responsible for Ca^{2+} ligand binding, and were probably inactivating mutations, as expected for SCCs, rather than activating mutations as seen in human leukemias[35,36].

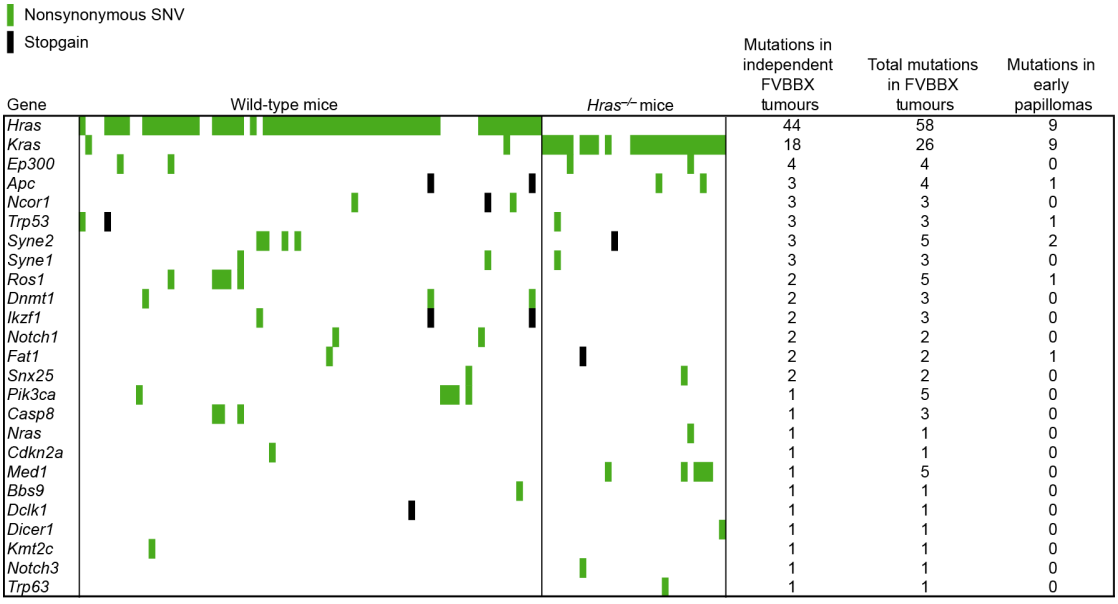


Figure 4. Mutated genes in carcinogen-induced tumors. Nonsynonymous and stopgain mutations in genes observed to be mutated in human cancer. Green squares indicate nonsynonymous mutations, and black squares indicate stopgains. Tumors are arranged on the horizontal axis by mouse genotype (wild type or *Hras*$^{-/-}$). Independent mutations are defined as mutations in tumors that show no phylogenetic relationship (based on shared mutations). Number of mutations observed in early papillomas is also shown.

Recurrent mutations were also seen in *Ep300*, *Apc*, *Ncor1*, *Syne1*, *Syne2*, *Ros1*, and *Dnmt1*, all of which have been observed in human tumors. Single mutations were also seen in *Pik3ca* and *Casp8* (consistent with those seen in head and neck SCC (HNSCC)); in *Bbs9*, *Dclk1*, and *Kmt2c* (consistent with those seen in cutaneous SCCs); and *Keap1* (consistent with those seen in lung SCCs). Among these, stop gains (mutations that create stop codons) were seen in *Apc*, *Ncor1*, and *Dclk1*.

Dependence of CNVs on initiation by *Hras* or *Kras* mutations

We used the whole-exome sequencing data to construct copy number profiles for all tumors. Copy number gains of chromosome 7 (on which *Hras* is located) were seen in tumors from wild-type mice but not $Hras^{-/-}$ mice, pointing to Hras as the driving force behind whole chromosome 7 gains. Undifferentiated spindle tumors did not have chromosome 7 gains, in line with our observation that spindle carcinomas have reduced *Hras* expression[27]. Copy number gains of chromosome 6, previously noted by karyotype analysis[58], were seen in tumors from both wild-type and $Hras^{-/-}$ animals (**Fig. 5a**). *Kras, Braf, Craf* and *Met* all reside on chromosome 6 and are candidate drivers of chromosome 6 gains. Additionally, we observed that SCCs from wild-type mice (i.e., *Hras*-driven carcinomas) have copy number gains on chromosome 1, and these are not seen in any other tumor class. The absence of chromosome 1 gains in papillomas and spindle tumors suggests that these gains are specifically linked to the squamous papilloma-carcinoma conversion.

To understand what might be driving the gain of chromosome 1, we searched for genes on chromosome 1 that were expressed at significantly higher levels in squamous tumors that had chromosome 1 gains compared to those that did not (**Fig. 5a**). *Nuak2*, which encodes an AMP protein kinase (AMPK)-related kinase, showed the greatest significant change in expression (1.7–fold increase, $P = 5.3 \times 10^{-5}$), and changes in *Itpkb* (Inositol 1,4,5-trisphosphate 3-kinase B) and *Cflar* (CASP8 and FADD-like apoptosis regulator) expression were also

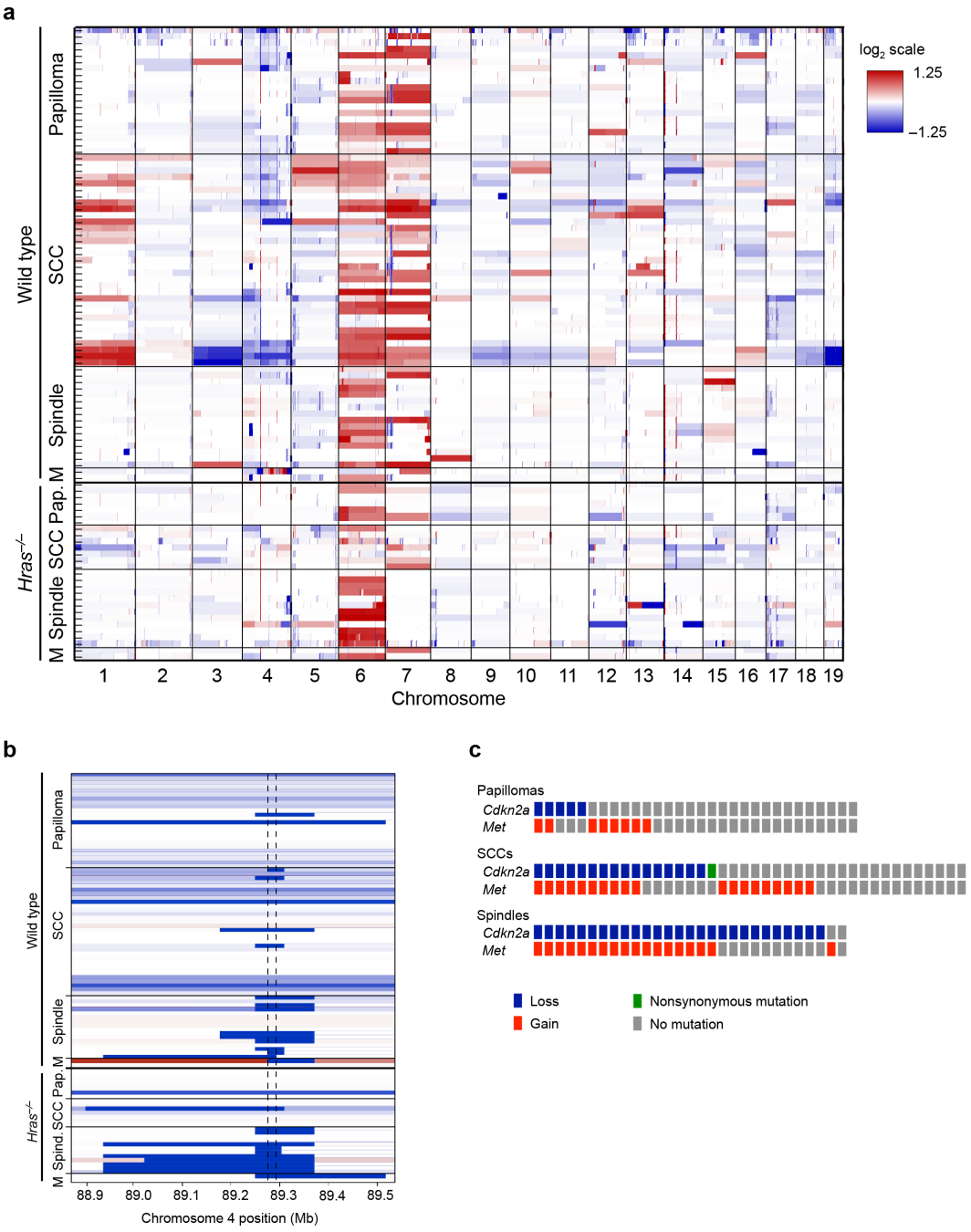


Figure 5. Gene copy number variations in mouse skin tumors. (**a**) CNVs are shown on each chromosome (X-axis) with *Ras* genotype and tumor morphology shown on the Y-axis (M=mixed SCC and spindle morphology). Chromosome numbers are on horizontal axis, and chromosomes are arranged with proximal end on the left on distal ends on the right. (**b**) Focal *Cdkn2a* deletions are visible in many tumors, frequently less than a megabase in size. Dashed lines indicate boundaries of the *Cdkn2a* gene. (**c**) *Cdkn2a* losses and *Met* amplifications in samples displayed by tumor type. Co-occurrence of these events increases substantially with tumor progression, from papillomas (7% co-occurrence) to SCCs (25% co-occurrence) to spindle tumors (59% co-occurence).

significant ($P = 5.1 \times 10^{-6}$ for *Itpkb*; $P = 1.1 \times 10^{-4}$ for *Cflar*, **Table 2**). *Nuak2* expression levels were strongly correlated with *Hras* expression levels in tumors from wild-type mice in this study ($\rho = 0.63$, $P = 1.9 \times 10^{-9}$), as well as in an independently derived data set[53] ($\rho = 0.58$, $P = 1.3 \times 10^{-6}$ respectively). *Nuak2* has been implicated in human cancer development, in particular by copy number gains in human melanoma[59], in breast and liver carcinomas[54,55], and, to a lesser extent, in HNSCC and lung SCC[35,36], and is therefore a possible therapeutic target. However other genes including those shown in **Table 2** may also play some role in promoting development of *Hras*-driven squamous carcinomas.

Gene	d-statistic	fold change
Itpkb	3.746794	1.587584
Nuak2	3.642988	1.737033
Cflar	3.529978	1.670079

Table 2. Differentially expressed genes in the context of chromosome 1 gain. Genes which display an increase in expression in SCC tumors from wild-type mice in which chromosome 1 has been gained, compared to SCC tumors in wild-type mice in which there was no gain. Statistics calculated using the siggenes package in R.

Among the most frequent copy number changes in aggressive human cancers are deletions at the *CDKN2A* locus on chromosome 9p, and amplification or copy number gains of *MET*[57,60], both of which we also observed in our cohort. *Cdkn2a* was deleted in 52 tumors (50%), including nearly all spindle tumors (27 of 29 tumors, 93%) (**Fig. 5b,c**)[27,61]. *Met* showed copy number gains in 47 tumors (46%), as the result of both focal amplifications (6 tumors) and of whole gains of chromosome 6 (41 tumors). Both the occurrence and co-occurrence of these lesions increased dramatically with tumor stage: both events were rare in papillomas, but in SCCs, 63% of tumors had either a *Cdkn2a* loss or *Met* gain, and 25% had both. In spindle tumors, nearly all tumors had a *Cdkn2a* loss or *Met* gain, and 59% of tumors had both (**Fig. 5c**).

Acquisition of SNVs precedes CNVs during progression

To verify the timing of SNVs and CNVs during tumor progression we sequenced a set of 18 early, small (2–4mm) FVB/N papillomas at 12–14 weeks after tumor initiation (9 from wild-type mice, 9 from $Hras^{-/-}$ mice). The overall mutation spectrum was highly similar to that in the

tumors from the FVBBX population (**Fig. 6a**). Interestingly, these early papillomas had already acquired a similar proportion of G>T mutations to late papillomas from FVBBX mice, suggesting that most G>T mutations are acquired relatively early in the life of the tumor, during the phase of proliferation and chronic inflammation that is induced by the tumor-promoting TPA. Analysis of CNVs in early papillomas identified gain of chromosome 7, which contains the mutant *Hras* gene, exclusively in papillomas from wild type mice, as the only recurrent and significant early event (**Fig. 6b**). The striking absence of additional copy number events in these early tumors suggests that acquisition of CNVs seen in the original cohort is a later event associated with tumor progression.

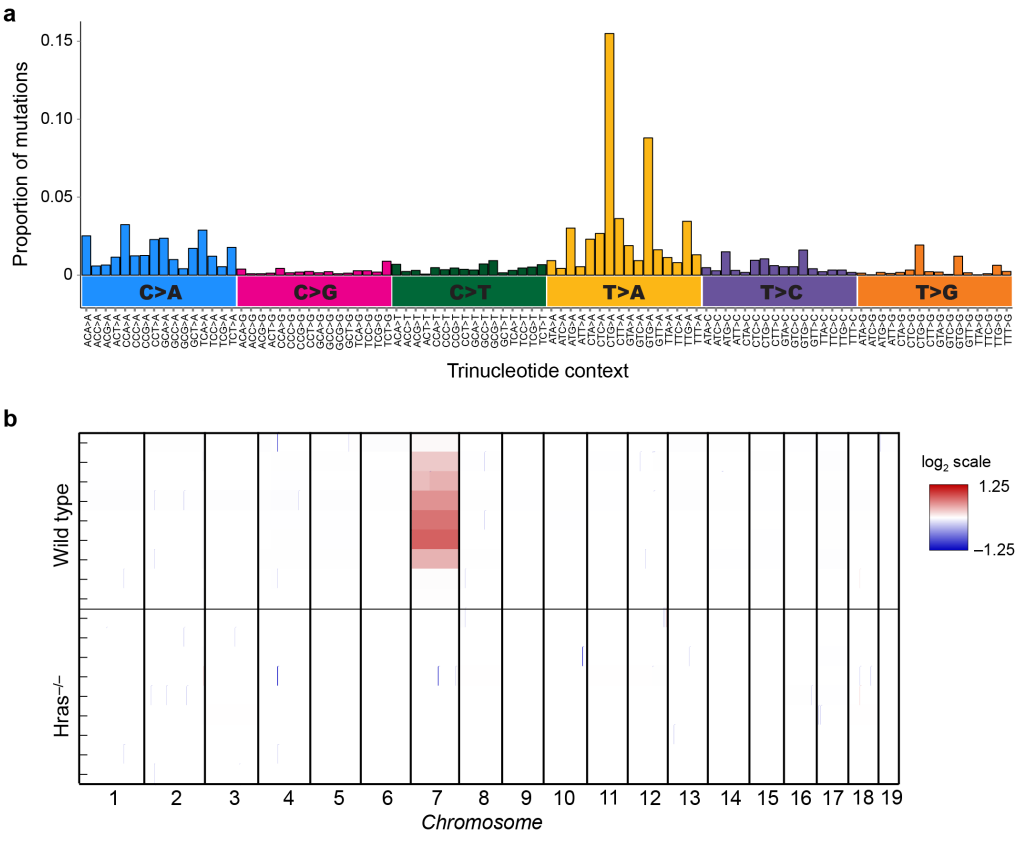


Figure 6. Genetic alterations in 12–14 week papillomas. (**a**) Frequency of mutation observed in each of 96 possible trinucleotide mutation contexts for all mutations in 18 early papillomas. Trinucleotide contexts are arranged on horizontal axis grouped by base pair change of the mutation. DMBA signature of T>A mutations (49% of all mutations) is observed, as are G>T mutations (25% of all mutations). (**b**) Copy number profiles for 18 early papillomas. CNVs are shown on each chromosome (X-axis) with Ras genotype shown on the Y-axis. Two-thirds (6 of 9) of early papillomas from wild-type mice show a chromosome 7 gain. These tumors also have an *Hras* mutation. None of the 9 papillomas from *Hras*$^{-/-}$ mice, which carry *Kras* mutations, show any copy number gains or losses.

DISCUSSION

We have exploited a classical skin tumor model in a genetically heterogeneous mouse population to illuminate the evolutionary history of metastatic tumors by whole-exome sequencing of multiple lesions from the same animals. The ability to distinguish the parental alleles in each mouse shows that both *Hras* and *Kras* mutations are highly allele-specific, demonstrating the important role of germline polymorphisms in determining mutation selection. Furthermore, the nature of the initiating event (*Hras* or *Kras* mutation) influences the subsequent genomic changes that take place, making this model useful for identifying potential drug candidates such as *Nuak2* that may be activated in tumors with *Hras* mutations.

The use of specific initiating mutagens combined with clonal analysis has enabled us to distinguish early from late mutations. While the initiating agent DMBA causes primarily A>T mutations genome-wide, tumors accumulate G>T transversions during the early phase of TPA treatment, presumably due to induction of oxidative stress pathways linked to inflammation or to Ras signaling[62,63]. The high frequency of point mutations in these skin tumors, which contrasts with the paucity of informative mutations in genetically engineered mouse (GEM) models[64,65], provides the opportunity to map, in great detail, the sequential mutations that take place during tumor evolution.

Our data support the “parallel evolution” model as the major route to metastasis, by which each metastasis is seeded directly by the primary tumor, rather than via a lymph node[25,49,66], as only one mouse (Mouse 1949, **Fig. 2f**) showed evidence in favor of the linear model (**Fig. 2**). Interestingly, the number of shared mutations between primary and metastases in these mice was highly variable, suggesting that the seeding could occur either early or late in the primary tumor evolution. Dissemination from the primary tumor can therefore occur very early and probably continuously[25], but the capacity of disseminated tumor cells to grow out into full metastases is controlled by a separate, presently unknown factor, possibly secreted “niche factors”[67,68], which allow progressive growth at distant sites. However, the presence of multiple

primary carcinomas in some animals, only one of which showed clear signs of metastasis, suggests that seeding is also controlled by tumor-intrinsic factors, possibly the repression or absence of neo-antigens that would stimulate recognition by the immune system[69].

Our study design also allowed us to compare primary carcinomas that metastasized and other carcinomas from the same animals that had not. The availability of these unique matched samples allowed us to carry out gene expression analysis to identify a network of genes more highly expressed in metastasizing carcinomas than in their non-metastasizing counterparts, including *Pdlim7* and *Cd151* which have been previously linked to integrin signaling and tumor invasion[51,52]. Further functional studies of these candidates in this model will be the subject of future investigations.

Another recent study of evolution of metastasis by next generation sequencing of a mouse model of small cell lung cancer reached the conclusion that lymph nodes may be the gateway to distant metastases[65], in agreement with the linear model. However, evidence for linear seeding was based mostly on a primary lung cancer and two metastases, one in the lymph node and one in the liver, from a single mouse. The reasons for the different conclusions from this study are unclear, but may be due to the relative numbers of samples analyzed, or to the use of a GEM model with very few nonsynonymous point mutations. During the revision stage of this manuscript, another study appeared that utilized sequencing to identify relationships between primary skin tumors and their metastases[70]. In this study, the conclusions regarding tumor spread were, again, made on the basis of one informative mouse with two metastases. Thus, no clear conclusions can be drawn from these studies regarding the predominance of the linear vs parallel evolution models of metastasis.

The parallel evolution of metastases seen in this study is supported by a substantial body of clinical data showing that lymph node removal has no impact on patient survival[71–73]. The observation that primary and metastatic tumors in some mice share a very low number of mutations supports prior observations that early stage tumors can disseminate even before

reaching overt malignancy[25,66]. Our data indicate that not only must these tumors have begun dissemination early, but they also acquired the ability to seed metastases probably before the primary tumor advanced to a clinically recognizable carcinoma stage. Clearly, efforts to prevent dissemination or metastatic seeding in such cases would be futile, and other therapeutic approaches would be required. The availability of mouse models of cancer, that mimic both the genomic heterogeneity and natural evolution of metastasis, such as that described here, will provide the opportunity to address these questions in more detail in a preclinical setting.

METHODS

Mouse backcross and carcinogenesis. Wild type male SPRET/EiJ *Mus spretus mice* were crossed to female *Hras*$^{-/-}$ FVB/N *Mus musculus* mice to generate *Hras* heterozygous F1s. Female F1s were crossed with male *Hras*$^{-/-}$ or *Hras* heterozygous FVB/N *Mus musculus* mice to generate FVBBX mice. Fifteen FVBBX mice were used in this experiment, 13 males and two females, chosen based on the availability of papilloma, primary carcinoma, and metastasis samples for sequencing. Early papillomas were harvested from four female and three male FVB/N *Mus musculus* mice, either wild type or *Hras*$^{-/-}$, at 12–14 weeks after initiation with DMBA, at a size of 2–4mm diameter.

Mice were shaved and treated with 25mg DMBA (Sigma-Aldrich) dissolved in 200uL acetone at 8 weeks, and subsequently received TPA (Sigma-Aldrich) (200uL of a 10^{-4} M solution in acetone) twice a week for 20 weeks, following established chemical carcinogenesis protocol[5]. Carcinomas were surgically resected when they reached a size of >1 cm in longest diameter. Mice were sacrificed when disease progressed, per animal care requirements. At sacrifice, papillomas and carcinomas were removed from skin, and all internal tumors were resected. Matched normal skin was also taken. A section of each primary carcinoma and metastasis was embedded in paraffin at time of removal for histological evaluation. Remaining tumor was flash

frozen and stored at –80 C, as was normal skin. No randomization or blinding was used in these experiments, as all mice underwent an identical treatment protocol. All animal experiments were approved by the University of California San Francisco Laboratory Animal Resource Center.

Histological classification. Formalin-fixed paraffin-embedded tumor was sectioned to 6 μm and H&E stained for histologic assessment.

Nucleic acid extraction. DNA was extracted from flash frozen tumor and skin samples. Frozen tissue was ground and digested in 3mg proteinase K (Bioline) overnight. DNA was separated by addition of phenol/chloroform and use of 5 PRIME Phase Lock Gel Heavy Tubes (Fisher Scientific), precipitated with isopropanol, washed with 70% ethanol, and re-suspended in nuclease-free water. Concentration and quality were determined by Nanodrop spectrophotometry and by PicoGreen (Invitrogen).

RNA was extracted from flash frozen tumor and skin samples using TriZol reagent (Invitrogen), following manufacturers' instructions, and cleaned with Qiagen RNeasy kit following instructions. Concentration was determined by Nanodrop spectrophotometry and RNA integrity number (RIN) determined by Agilent 2100 bioanalyser. All samples used for microarray analysis had a minimum RIN score of 6.5.

Exome sequencing, alignment, and quality control. Exome enrichment and sequencing genomic libraries were prepared using the Illumina Paired End Sample Prep Kit following manufacturer instructions. Enrichment was performed as described previously[74] using the Agilent SureSelect Mouse All Exon kit following the manufacturer's recommended protocol. Each exome was sequenced using a 76bp paired-end protocol on the Illumina platform (GAII or HiSeq2000).

Tumor .bam files were aligned to the GRCm38/mm10 version of the *Mus musculus* genome using BWA (version 0.5.10)[75]. After alignment, duplicates were marked and mate information was fixed using Picard (version 1.72; **http://picard.sourceforge.net/).** We then recalibrated base quality score and realigned reads around indels using the Genome Analysis Toolkit (GATK) (version 1.5-9)[76]. Finally, alignment and coverage metrics were collected using Picard. We sequenced an average of 66 million unique on-target reads per tumor. Targeted bases were sequenced to a mean depth of 50, and 78% of targeted bases were sequenced to 20X coverage or greater. There were no significant differences in depth of coverage or proportion of regions covered to 20X between tumor types.

Variant calling. SNVs were called using somatic variant detection program MuTect (version 1.1.4)[77]. Each tumor was called against its matched normal tissue, and calls were filtered against a database of known *Mus musculus* FVB and *Mus spretus* SPRET germline SNPs available at ftp-mouse.sanger.ac.uk (release 1303, mgp.v3), as well as against a panel of normal skins from this experiment. MuTect was run with the required_maximum_alt_allele_mapping_quality_score parameter set to 60, as the mixed genetic background of our FVBBX cross resulted in a number of poorly mapping reads we wished to exclude. Results were further filtered to calls with a minimum read depth of 10, and with a minimum mutant allele fraction of 10%. Variants were annotated using Annovar (downloaded on 2/13/2014)[78], and these annotations were used as the basis for assessing exonic variants as synonymous, nonsynonymous, stopgain, or stoploss. Mutations were compared to mutations observed in human cancers using the COSMIC database (version 68, **http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/download**)[79]. Genes were selected for display (**Fig. 4**) based whether they were commonly mutated in HNSCC or cutaneous skin SCC, and genes on a previous published list of cancer driver genes[80].

Copy number calling was done with CNVkit[81], and tumors copy number status was called against a panel of normal skins from this experiment. Raw copy number gain was adjusted to remove dilution from contaminating normal cells, with contamination determined by PyClone estimate (see below). Whole chromosome gains considered to be those in which over 90% of the chromosome was gained and the weighted mean amplification $\log_2$ was greater than 0.4, implying trisomy in at least 30% of cells. Focal copy number events were determined as those for which the average copy number change for all exons of the gene had a $\log_2$ increase of 0.6 or $\log_2$ decrease of –1 (corresponding to a 50% gain or loss of alleles).

Mutation Validation. To validate our mutation calls, we sequenced 205 SNVs to a median depth of 7,080 reads using MiSeq. We obtained an 86% validation rate. **PCR and Sequencing:** Primers for each candidate SNP were designed as nested pairs using Primer3[82,83]. External amplicons were fixed at 400–800 bp, internal amplicons were fixed at 250-300 bp. Internal primer pairs had partial Illumina adaptor sequences added to allow the construction of barcoded Illumina libraries. Tail sequences are: left adaptor 5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT 3', right adaptor 5' TCGGCATTCCTGCTGAACCGCTCTTCCGATCT 3'. PCR was performed using hot start Taq DNA polymerase following the manufacturers instructions (Thermo Scientific). PCR1 (external primers) used a touchdown PCR approach for increased specificity. The products of PCR1 were diluted 100x and used as a template for PCR2 which added the partial Illumina adaptor tails. Candidate SNP PCRs for the same tumor were then pooled and each tumor was barcoded with a different full length Illumina adaptor barcode in a short PCR3 (15 cycles) so each tumor could be decoded after sequencing. Resulting tumor specific PCRs were pooled, cleaned up with 0.8x AmpureXP beads (Beckman Coulter), quantified and run on an Illumina Miseq 150bp PE run. **Data Analysis:** Illumina adaptor sequence and low-quality bases (<20) were trimmed from reads using cutadapt[84] (**https://code.google.com/p/cutadapt/).** Reads were mapped to the

GRCm38 build of the mouse genome using BWA mem[75] (**http://bio-bwa.sourceforge.net/).** Duplicate reads were not marked due to it being amplicon sequencing. Samtools mpileup and bcftools call[85] (**http://www.htslib.org/)** were used to output reference or alternate base calls and allele depths for the candidate positions. Samples were analysed as groups of matched normal skin and tumor samples. SNVs were considered validated if they contained at least 70 reads supporting the expected mutation.

Mutation context. For mutation spectrum analysis, SNVs in all tumors were annotated with 1 of 96 possible trinucleotide context substitutions (6 types of substitutions x 4 possible flanking 5'-bases x 4 possible flanking 3'-bases), and counts of each mutation context were summed.

Clonality. Clonality analysis was performed with PyClone (version 0.12.7)[86]. For inputs to PyClone, reference and variant read depths were taken from MuTect output, and copy number at each locus determined from CNVkit output. Clonal and subclonal mutation clusters were determined from PyClone results table, and compared with PyClone cellular frequency plots for confirmation. Clonal mutations were those present in all tumor cells; subclonal mutations were those present in only a fraction of tumor cells.

Phylogenies. To build phylogenetic trees, absolute distance matrices were calculated based on the presence of mutations in the sample, based on filtered MuTect calls. Rooted trees were built with use of the Analyses of Phylogenetics and Evolution (APE) package and manhattan calculation method implemented in R version 2.15. Metastases were matched to primary tumors on the basis of shared mutations.

Identification of Early and Late Mutations. Early and late mutations were identified with two methods. In the first method, mutations from 26 metastases with matched primaries were

divided into those that were shared with the primary (early) and those that were not shared (later). Tumors used in this method included a representation of SCCs and spindle tumors, and of *Hras*-driven and *Kras*-driven tumors. In the second method, mutations in 12 tumors where multiple clones were identified with PyClone were divided into fully clonal mutations (i.e., those present in all tumor cells) (early) and subclonal mutations (i.e., those present in only a fraction of tumor cells) (late). Tumors used in this method included a representation of papillomas, SCCs, and spindle tumors; *Hras*-driven and *Kras*-driven tumors; and primary tumors and metastases.

Evaluation of Candidate "Metastasis Genes." To assess whether we had any candidate "metastasis genes" that might explain why certain primary tumors metastasized while others did not, we first examined genes that were recurrently mutated in metastasizing carcinomas. We found 22 genes that were mutated more than once in a metastasizing carcinoma, however, 19 of these were also mutated in non-metastasizing carcinomas at a similar frequency. The remaining three, which were *Foxn4*, *Scn9a*, and *Sspo*, were found mutated in one or more papillomas, and so also did not present good candidates. We found *Pdlim7*, *Tecpr2*, and *Cd151* were expressed at higher levels in metastasizing carcinomas, although of these *Pdlim7* was significant after multiple test correction ($p = 6.4 \times 10^{-6}$). None of these genes carried any mutations in any primary or metastatic carcinomas. No genes were expressed at significantly lower levels in metastasizing carcinomas.

Allele specificity analysis. To determine the allele specificity of *Hras* and *Kras* mutations, 75bp reads that contained both the mutation and nearby SNPs were extracted from .bam file using mpileup[56], and genotype of mutated allele was determined. For *Hras* codon 61 mutations, SNPs used to determine genotype were chr7:141192537, A (FVB) or G (SPRET) and chr7:141192567, G (FVB) or A (SPRET). For *Kras* codon 61 mutations, SNPs used were chr6:145234318, T (FVB) or C (SPRET) and chr6:145234388 G (FVB) or A (SPRET). For *Kras*

codon 12 and 13 mutations SNPs used were chr6:145246755, G (FVB) or A (SPRET) and chr6:145246782 C (FVB) or T (SPRET). Significance was evaluated using a chi-squared test.

Expression microarrays, normalization, and differential expression.

Gene expression was quantified using Affymetrix MoGene ST 1.1 arrays hybridized on an Affymetrix GeneTitan instrument. Affymetrix MoGene arrays were normalized using the oligo package[87] and a probe database prepared for FVBBX mice to avoid probes which intersect known FVB/N or SPRET/EiJ SNPs[88]. Differential expression for genes on Chromosome 1 and for comparison of metastasizing vs non-metastasizing carcinomas was done using the siggenes package in R, with a false discovery rate threshold of 0.01. (Microarray data available: GEO accession GSE63967).

Chapter 2

Cellular and genetic heterogeneity of multistage skin tumor evolution

INTRODUCTION

The hypothesis that tumors evolve according to a process resembling Darwinian selection dates back at least as far as 1976, when Peter Nowell drew an analogy between tumor development and the evolution of species[4]. Tumor cells frequently acquire new mutations as the tumor grows, leading to heterogeneity within the tumor and the co-existence of multiple, genetically distinct subclones[4,12]. These subclones may cooperate to drive tumor growth or progression[89,90]; under selective pressure however a subclone may acquire a growth advantage and expand disproportionately, causing a "clonal sweep" in which the advantaged subclone comes to dominate the tumor.

Clonal dynamics at each stage of tumor advancement—from initiation to malignant progression to metastasis—are unique, and it is possible that single or multiple individual cells may contribute to the advancement from each stage to the next. At the initiation stage, tumors are commonly thought to arise from the clonal expansion of a single, initiated cell[4,12], but this model has been challenged based on analysis of both genetic and cellular heterogeneity in mouse models and human tissues (for review see Parsons[19]). The discovery of a wide variety of oncogenic point mutations in groups of cells in histologically normal tissue[91,92] raises the possibility that multiple cells within a field can participate in the earliest stages of cancer development. It has also been proposed that microenvironmental changes that lead to inflammation can cause "field cancerization" with the consequence that multiple individual cells within a field can contribute to the appearance of premalignant lesions (for review: Goruppi and Dotto[93]) . In support of this "polyclonal origin" model, several independent studies using a mouse model of multi-stage carcinogenesis, in which skin tumors are initiated by a chemical mutagen such as dimethylbenzanthracene (DMBA), and promoted by chronic 12-O-

tetradecanoylphorbol-13-acetate (TPA) exposure, showed that benign papillomas can harbor multiple distinct cellular populations[20,21,94]. The nature of these distinct populations has remained obscure, as has the extent of their contributions to tumor progression.

Studies using the DMBA/TPA model have also revealed that benign and malignant tumors exhibit distinct growth dynamics, with evidence for hierarchical, stem-cell-like cells driving growth of benign tumors, while malignant tumors are driven by cells expanding geometrically, with limited capacity to differentiate[95]. This implies that clonal dynamics and the forces that shape tumor heterogeneity are distinct between benign and malignant stages. The transition between these states and the cell population or populations that drive it, however, are not well understood.

Further debate has focused on whether metastasis is a clonal event. While it was once widely thought that metastases were the clonal outgrowths of single disseminated tumor cells, this idea was challenged by experiments revealing that invasive tumor cells could move in groups[23,96], that circulating tumor cell clusters could be found in the blood, and that these clusters had higher metastatic potential than individual circulating tumor cells[24,97]. Deep sequencing of human prostate cancer and melanoma has suggested that multiple matched subclones could be identified in a primary tumor and matched metastasis, arguing for polyclonal seeding[98,99]. Tumors in a genetically engineered mouse model of pancreatic cancer that generates polyclonal tumors were also able to give rise to metastases that contained multiple cell populations from multiple tumor clones[97]. One caveat that applies to all of these studies is that the subclones identified could either have arisen from a pre-existing dominant clone within the tumor, or could be an independent cell population arising from a separate target cell. Analysis of both cellular lineages and genetic heterogeneity within tumors will be essential to resolve the relative importance of these alternative possibilities.

The evolution of cancer from initiation to metastasis is fostered by high levels of intratumoral genetic heterogeneity, which facilitates selection of specific cell populations at

different stages[100]. In contrast to genetically engineered mouse tumors, carcinomas inititated chemically are highly heterogeneous, harboring similar numbers of point mutations and genomic aberrations to human carcinogen-associated cancers[28,64,65]. Here we combine the multi-stage DMBA/TPA cancer model, the most widely used mouse model for studying multi-stage carcinogenesis, with multi-colored Confetti lineage tracing[29] to interrogate clonal dynamics at multiple stages of tumor advancement, from initiation to metastasis. Our data confirm earlier studies demonstrating that chemically initiated papillomas contain multiple cell populations that arise from distinct cells[20,21] compatible with the polyclonal origin model. However a combination of genetic analysis and cell labeling shows that only one dominant clone in these lesions carries the initiating mutation in the *Hras* gene[26]. The other populations are distinct in that they have no initiating mutation, have a very low overall mutation burden and lack gene copy number changes seen in the dominant clones. We propose that these cells are recruited into papillomas during growth rather than being present at the earliest stages, in a manner that mimics the recruitment of normal stem cells into a healing wound[101]. Multi-color labeling further allowed us to observe the distinct patterns of clonal behavior associated with progression to carcinomas, documenting emergence of a dominant clone in tumors that progress to malignancy. In contrast to the clonal events associated with initiation and progression, metastases in this model can involve the participation of different cells present in the matched primary tumor. Metastasis is therefore not a clonal event at the cellular level, suggesting it may be driven by fundamentally different processes compared to mechanisms that select initiated cells or foster benign-malignant progression.

RESULTS

Confetti labels the skin efficiently and durably

The Confetti mouse, which has been used to study clonal dynamics of stem cells in normal tissue, contains a four-color cassette, which upon activation with Cre labels each cell

stochastically and permanently with one of four fluorophores: GFP (nuclear), YFP (cytoplasmic), RFP (cytoplasmic), or CFP (membrane-bound)[29]. We labeled the back skin of *K5CreER-Confetti* mice by administering 4 doses of topical tamoxifen to 8 week old mice, which labeled K5+ skin cells in both the intrafollicular epidermis and the hair follicle (**Fig. 1a**) and achieved 75% labeling. We chose a topical application of tamoxifen to limit the labeling to the back skin, and to avoid labeling internal organs that were expected to be future sites of metastasis.

Cre-mediated recombination in the Confetti cassette gave rise to cells of all four colors, but as has been previously reported, there was a bias in recombination frequency[29] that yielded considerably less GFP labeling than expected and a slight preference for RFP (**Fig. 1b**). Confetti labeling was durable, with skin from 18 month old mice continuing to show labeling of both hair follicles and interfollicular epidermis.

Mice were treated with the carcinogen DMBA 10 days after their final dose of tamoxifen, allowing time for tamoxifen to clear to avoid additional Confetti recombination[102]. At 10 days post-tamoxifen, we observed that cells in the hair follicle did not appear to have yet undergone division, however we observed in the interfollicular epidermis some small patches of matching-colored cells, with an average patch cross-section of 6-7 cells. We subsequently treated with TPA twice weekly for 20 weeks, following the standard DMBA/TPA carcinogenesis protocol[27] (**Fig. 1c**) and observed benign papillomas beginning at 6-8 weeks.

Multiple cell populations contribute to benign papillomas

Papillomas were initially collected from a cohort of tamoxifen-treated *K5CreER-Confetti* mice sacrificed at 20 weeks post-DMBA. The majority of these papillomas (75%) showed a multi-color Confetti labeling pattern, with the tumor comprised mainly of a single color (the “bulk color”) but with several, smaller populations of distinct colors also visible. Observed under a whole-mount fluorescent microscope, these additional populations appeared to originate from the base of the tumor and formed “streaks” up the side of the papilloma (**Fig. 2a**). These

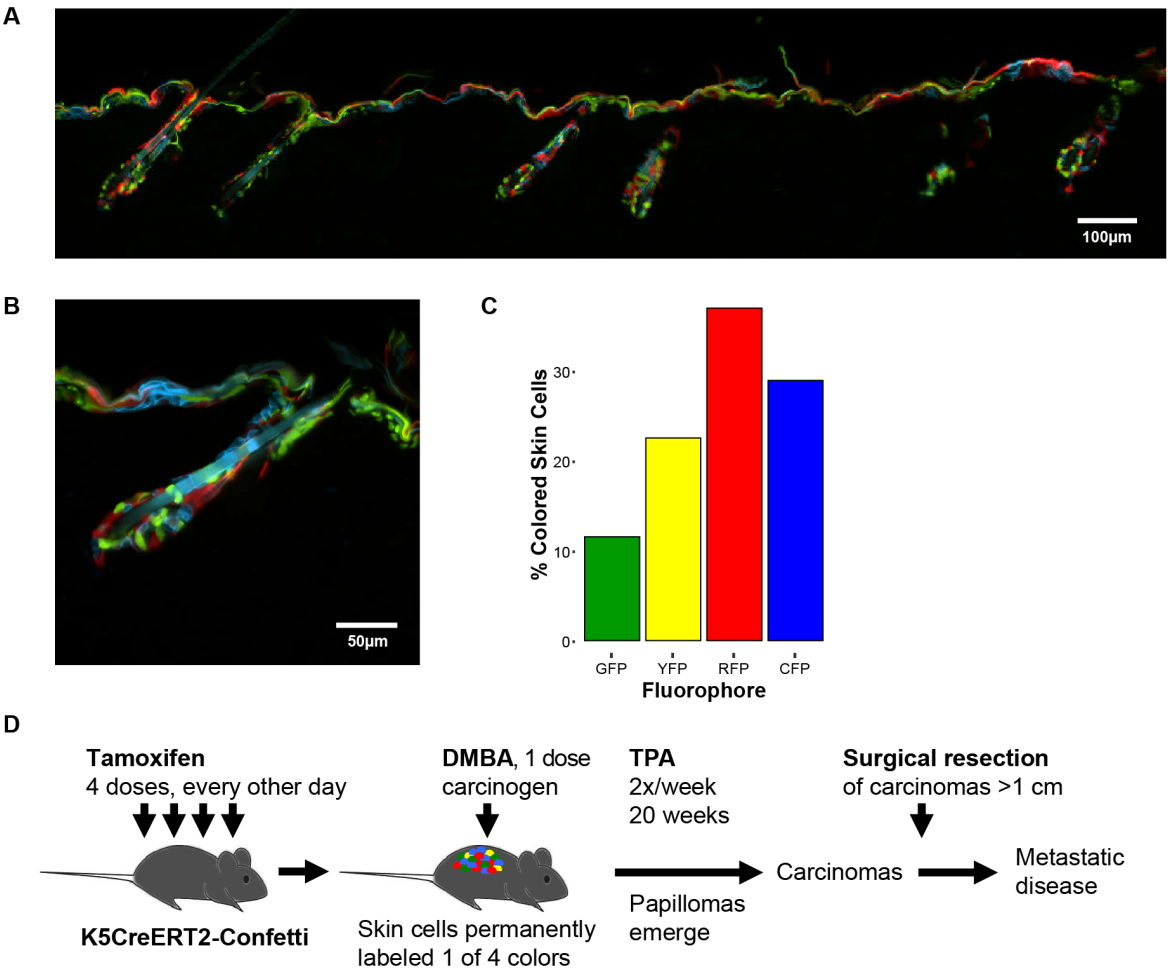


Figure 1. Tamoxifen-induced Confetti labeling of the skin. (**a, b**) Skin sections 10 days after final dose of tamoxifen, 10x (**a**) and 40x (**b**) magnification. (**c**) Proportion of colored skin cells expressing each Confetti fluorophore. (**d**) Schematic of tumorigenesis strategy. Mice are treated 4x with tamoxifen, and then given DMBA 10 days after final tamoxifen dose, followed by biweekly treatments with TPA.

patterns were similar to those observed many years ago using a completely independent approach involving chimaeric mice[21].

In order to determine whether these patterns were established early in papilloma development or evolved during growth, we analysed early stage papillomas from a separate cohort of mice harvested at 12 weeks after DMBA treatment. In contrast to the 20 week papillomas, the majority of 12-week tumors were single-colored (**Fig. 2b,c**), and showed little evidence of cellular heterogeneity, suggesting that the streak patterns developed over time during papilloma growth, rather than being present at the time of initiated cell selection. Analysis of papillomas from a third cohort of mice with terminal disease harvested between 27 and 49

weeks after TPA treatment showed patterns identical to those observed at 20 weeks. The increase in “streaked” papillomas between 12 weeks and both later timepoints was statistically significant (p=0.006 and p=0.00001 between 12 vs. 20 weeks and 12 weeks vs. terminal papillomas respectively) (**Fig. 2c**). Cross-sectioning of “streaked” tumors confirmed the visual pattern (**Fig. 2d**), and H&E staining of serial sections revealed no histological differences between cells that belonged to a streak and adjacent cells that did not (**Fig. 2e,f**), indicating the streak populations are a true part of the tumor that cannot be distinguished pathologically.

We monitored tumors in two cohorts of control mice to determine whether stochastic recombination of the Confetti cassette could affect our observations. In one control cohort *K5CreER-Confetti* mice were given no tamoxifen (217 tumors, 10 mice); in the other, Confetti mice lacking the *K5CreER* cassette were treated with tamoxifen (177 tumors, 7 mice) prior to treatment with DMBA. In both cohorts, we observed an extremely low level of Confetti leakiness. In a small number of papillomas—4 papillomas from the “no tamoxifen” cohort (2.3%) and 10 papillomas from the “no *K5CreER*” cohort (4.6%)—one or two individual spots of color were observed. These were significantly smaller in size than the “streaks” described above (<15 cells) and lacked any trends in localization. Further, no leakiness was observed in carcinomas or metastases from either cohort (0/11 and 0/9 carcinomas and 0/8 and 0/9 metastases in “no tamoxifen” and “no *K5CreER*” cohorts respectively). From this we conclude that leakiness did not contribute to the streak patterns or other Confetti color phenotypes we observed.

Minor cell populations are normal recruited cells, and papillomas are monoclonal in origin

To address the question of whether the “streaks” that develop during papilloma growth were genetically initiated cells, four multi-colored tumors were separated by FACS into red (RFP), blue (CFP), and green (YFP and GFP) components. Two bulk-RFP, one bulk-CFP, and one bulk-YFP tumor were sorted (**Supplementary Fig. 1,** see end of chapter), to avoid any

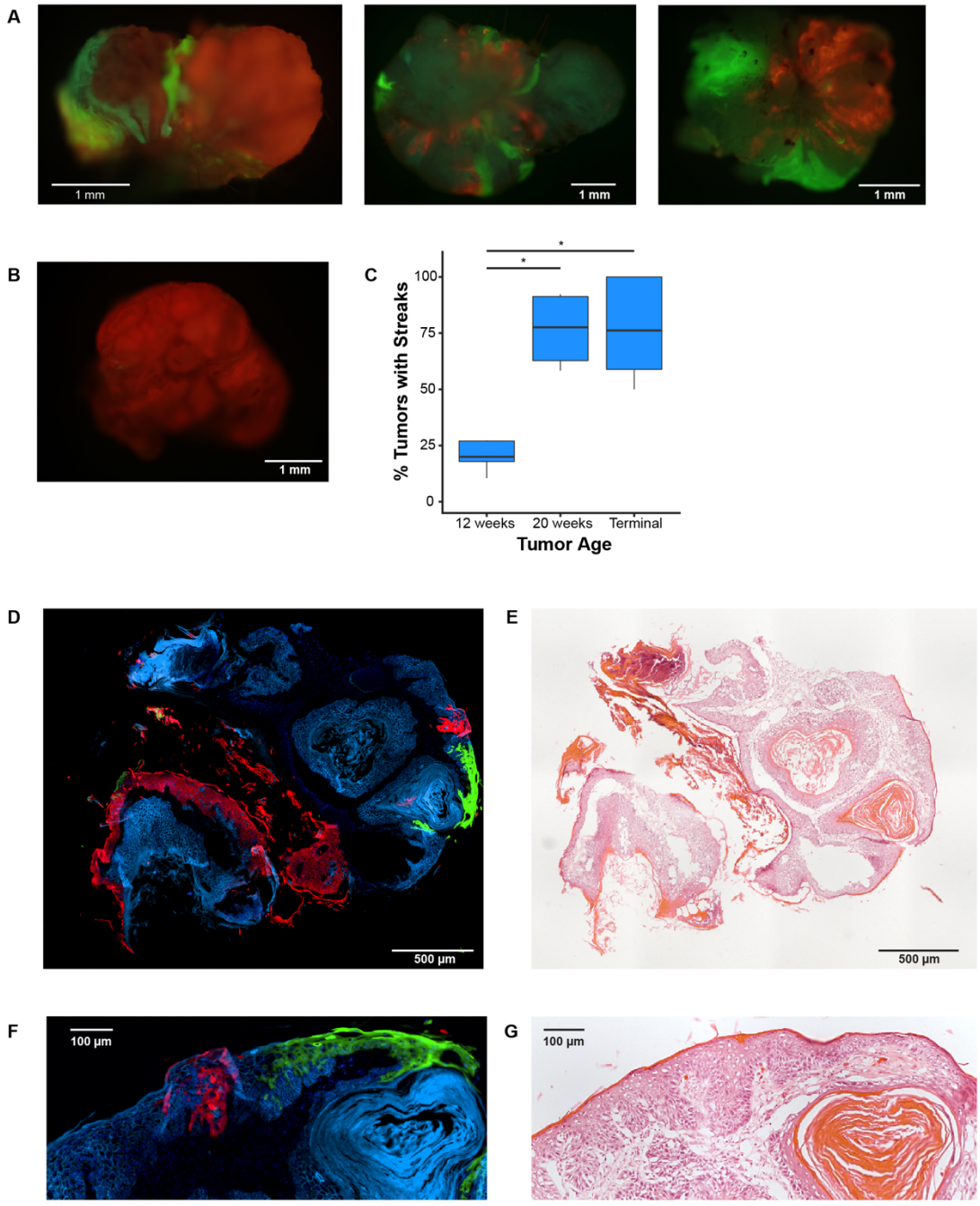


Figure 2. Streaked appearance of papillomas arising from Confetti-labeled skin. (**a**) Whole papillomas with multi-color streaking pattern, characteristic of 20-week papillomas, viewed from fluorescent dissecting microscope. Left panel views the papilloma from the side, center and right panels view papillomas from the bottom. (**b**) Whole papilloma lacking streaking pattern, characteristic of 12-week papillomas, viewed from fluorescent dissecting microscope. (**c**) Percentage of papillomas in each mouse that exhibited multi-color pattern. Data based on 222 12-week tumors, 77 20-week tumors, and 81 terminally benign tumors. (**d**) Cross-section of a streaked papilloma. (**e**) H&E of adjacent section of papilloma shown in panel (**d**). (**f**) RFP+ and YFP+ streaks in a CFP+ papilloma at 40x magnification. (**g**) H&E of adjacent section of papilloma region shown in panel (**f**).

fluorophore-based bias in quality of sorting; we found that while RFP, YFP, and GFP were all bright and easily isolated, CFP expression was weaker and led to lower total yield of CFP+ cells. Following isolation of "bulk" and "streak" cells from each tumor, we PCR-amplified and Sanger sequenced *Hras* in each population. Previous sequencing has shown that an *Hras* activating mutation, *Hras* Q61L*,* is the initiating mutation present in 95% of papillomas[26,28]. We therefore asked whether the bulk tumor and the streaks each had *Hras* mutations. We found that the bulk tumor population, in all four cases, had a heterozygous *Hras* Q61L mutation, while all the streak populations were wild-type for *Hras* by Sanger sequencing (**Fig. 3a**).

We then sequenced whole-exomes of the sorted "bulk" and "streak" cell populations. We found that the streaks carried fewer total mutations (4.8 vs 13.3 mutations per megabase in the bulk population), although this was not statistically significant due to low number of samples (**Fig. 3b**). The streaks however carried a distinct mutation signature compared to that in the bulk color population of cells. DMBA treatment is associated with a specific A>T mutation signature resulting from mis-repair of adducts formed with adenosine residues in DNA[103]. While mutations in the bulk populations carried the expected, dominant A>T DMBA mutation signature as well as a smaller number of G>T mutations, consistent with previous sequencing of tumors from this model[28,70], the streak populations carried almost no A>T mutations and instead had a mutation signature primarily comprised of G>T mutations (**Fig. 3c**). The near absence of A>T mutations in in streak populations indicates that the streaks, in addition to having no *Hras* mutations, almost completely lacked the signature of DMBA insult.

Our previous work demonstrated that, while A>T mutations were early mutations associated with DMBA, tumors in this model accumulate G>T mutations later and over time, possibly due to oxidative stress induced by the tumor promoter TPA[28]. In support of this, G>T mutations in these populations showed no strand bias, consistent with being reactive oxygen species (ROS)-induced and in contrast to the carcinogen-induced A>T mutations (**Supplementary Fig. 2**). Mutations in the streaks were thus likely accumulated in the tumor

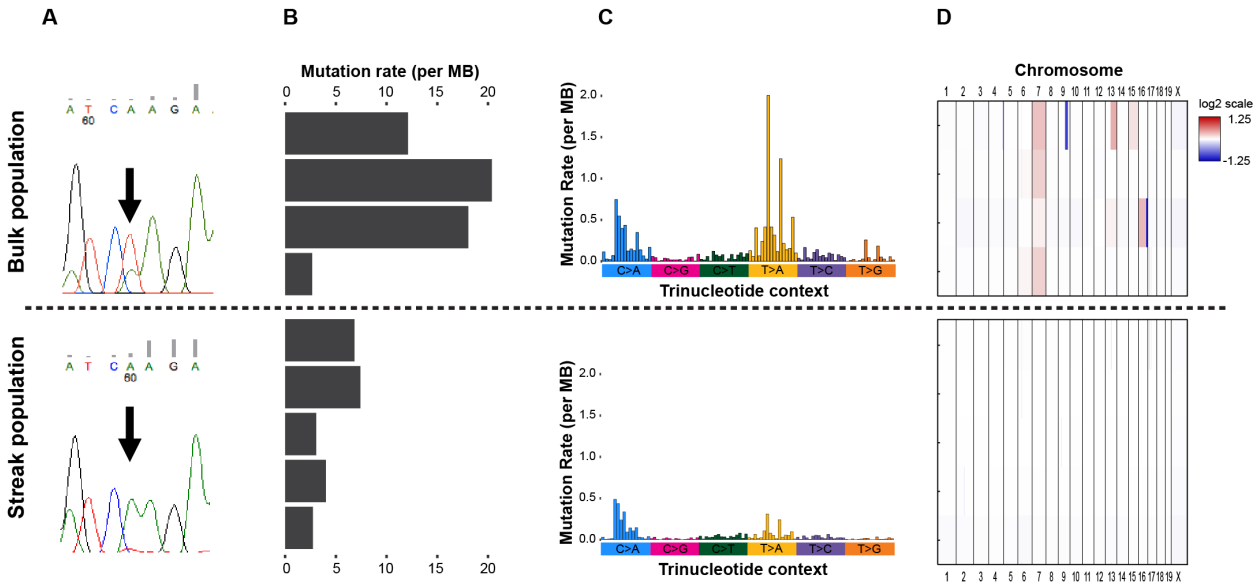


Figure 3. Genetic analysis of bulk and streak papilloma populations. (**a**) Sanger sequencing of Hras locus. Arrow points to chr7:141192550; when mutated to T, as seen in the bulk population, this results in an *Hras* Q61L mutation. (**b**) Mutation rate per megabase of 4 bulk population samples and 5 streak population samples. Bulk and streak populations taken from the same 4 papillomas. (**c**) Trinucleotide context of all mutations in the same 4 bulk and 5 streak population samples. X-axis denotes 96 possible trinucleotide contexts, grouped by (labeled) base pair change. T>A mutations, in yellow, are frequent in bulk population but near-absent in streak populations. (**d**) Copy number alterations in the same 4 bulk and 5 streak population samples. Chromosomes are arranged along the X-axis, samples on the Y-axis in the same order as panel (**b**).

environment after incorporation into the growing papillomas. Further, while the bulk population carried gains of chromosome 7, effectively duplicating the mutant copy of the *Hras* gene located on this chromosome[28,58], copy number profiles of the streak populations were completely silent (**Fig. 3d**). The mutation and copy number data together suggest that these streaks are derived from normal, neighboring K5+ keratinocytes (evidenced by the Confetti labeling) which were co-opted to grow and proliferate abnormally in the tumor.

Because these streaks are non-initiated cells lacking *Hras* mutations, because they dramatically increase in number between 12 and 20 weeks, and because even at 20 weeks they remain a small fraction of the total tumor (on average, 6% of total colored cells by FACS) we conclude that they are not co-initiators of papillomas, but are incorporated later during tumor growth, and that papillomas are in fact of monoclonal origin.

Non-progressing papillomas harbor multiple equipotent clones, but a single clone drives progression to malignancy

Most papillomas that arise as a consequence of sequential DMBA/TPA treatment do not progress to malignant carcinomas during the lifetime of the host animal. This observation has led to debate about the mechanisms of tumor progression, and whether this process is purely stochastic or linked to acquisition of additional genetic changes. In support of the latter possibility, *Trp53* mutations have been linked to progression rather than initiation in this model[9–11]. There is also evidence for the existence of "high risk" papillomas, with a high probability of becoming malignant[104], but whether this property is associated with new genetic events is unclear. An alternative, or additional, scenario is that high risk papillomas may arise from a different target cell within the skin epithelium that has more stem cell-like properties[105].

In order to investigate the cellular dynamics of malignant progression, we activated Confetti labeling in established papillomas and followed clonal evolution patterns over time. *K5CreER-Confetti* mice were treated at 8 weeks with DMBA and TPA according to the standard protocol, and were given 2 doses of tamoxifen 8 weeks later when papillomas were visible on the skin and typically 0.5-2mm in size (**Fig. 4a**). This resulted in widespread labeling of papillomas with all four Confetti colors (**Fig. 4b**), activating fluorescence in approximately 45% of tumor cells, as well as labeling of adjacent skin.

We monitored these mice over a period of 6-7 months after labeling, to identify clonal dynamics in papillomas that did not progress to malignancy, compared with those that underwent malignant conversion. In papillomas that did not progress to carcinomas over a 6 month time period, we observed that typically several large mono-color clones emerged (**Fig. 4c,d**). This is consistent with papilloma growth being driven by a relatively small number of stem-cell-like cells, consistent with previous reports[95]. Notably, the pattern of multiple side-by-side clones persisted for the majority of papillomas that did not progress and could still be observed 6 months later at the time the mouse was sacrificed. We conclude that in these non-

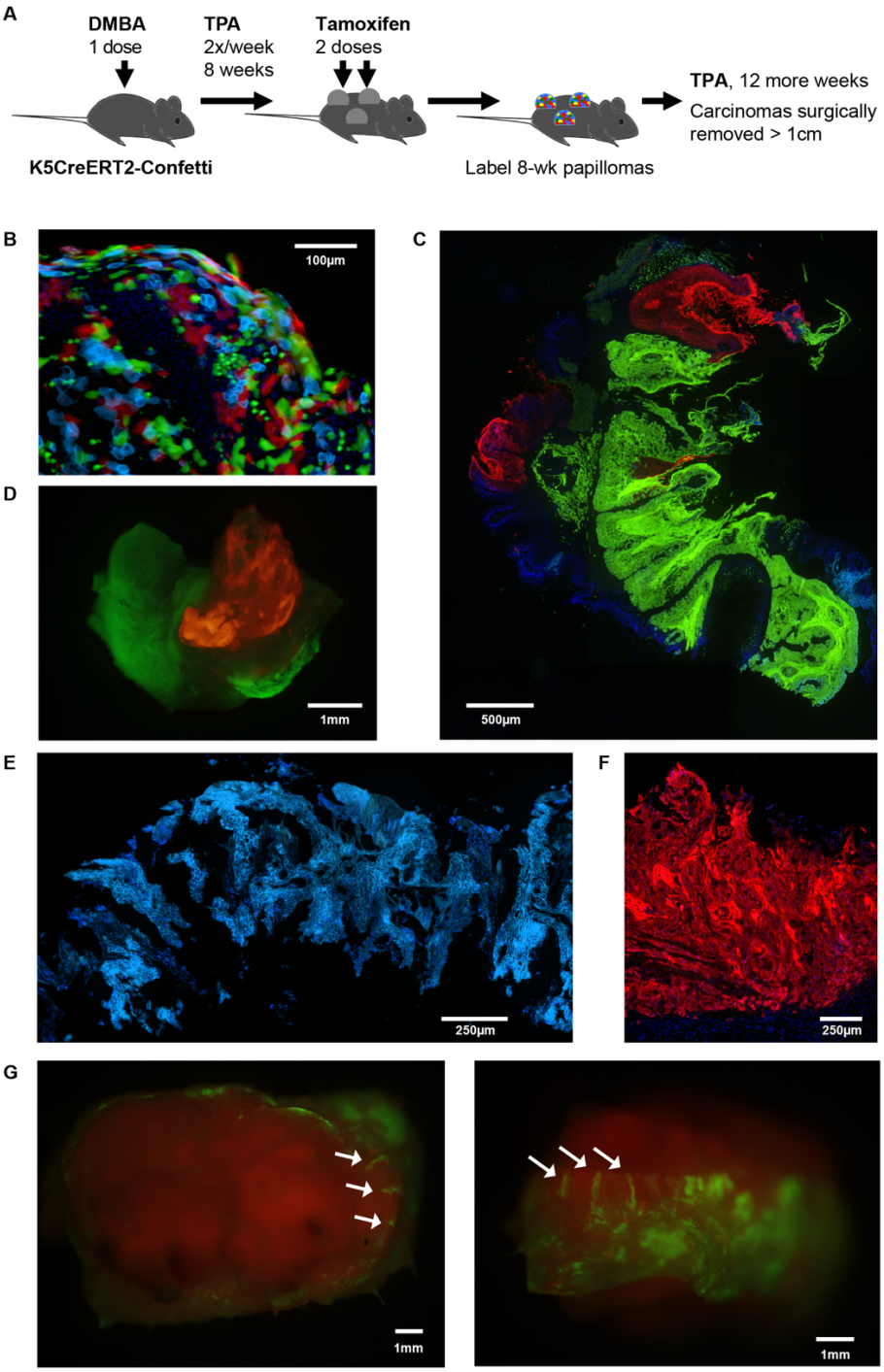


Figure 4. Tumor evolution following Confetti labeling at 8 weeks. (**a**) Schematic of experimental design. DMBA/TPA-induced papillomas were allowed to grow for 8 weeks. Confetti labeling was activated 8 weeks after DMBA treatment with 2 doses of tamoxifen, and TPA treatment was continued. (**b**) Cross-section of a papilloma labeled at 8 weeks post-DMBA, 3 days after final dose of tamoxifen, exhibiting Confetti labeling. (**c**) Cross-section of a papilloma labeled at 8 weeks post-DMBA, 8 weeks after tamoxifen. (**d**) Whole papilloma labeled at 8 weeks post-DMBA, 27 weeks after tamoxifen, viewed from with side with a fluorescent dissecting microscope. (**e, f**) Cross-sections of carcinomas from mice in which Confetti labeling was activated 8 weeks post-DMBA. (**g**) Whole carcinoma from a mouse treated with tamoxifen at 8 weeks, viewed from the top (left) and side (right). Streaks, indicated by arrows, are visible but confined to the periphery.

progressing papillomas, no single dominant clone emerged that had taken over the bulk of the cells within the tumor.

In contrast, a different pattern was seen in carcinomas that developed from papillomas in this same cohort of mice. Carcinomas that were surgically resected when they reached 1cm in diameter (13 carcinomas from 4 mice) were all comprised of a single color clone (**Fig. 4e,f**). Approximately half (6/13) were entirely colored (3 RFP, 2 YFP, 1 CFP), while the other half were entirely uncolored. Single-color patterns were confirmed by FACS for 11 of 13 tumors (the remaining 2 carcinomas were not sorted) and by sectioning. To assess the significance of this, we compared this single-color pattern in carcinomas with the number of visible colored lobes observed in papillomas that were age-matched to a carcinoma or obtained at sacrifice. While carcinomas were comprised of only one color, papillomas contained on average 2.5 distinctly colored, externally visible lobes ($p = 0.0003$). This number is likely an underestimate, as only externally visible papilloma lobes could be scored.

It should be noted that in assessing carcinoma color, we focused on the tumor core. It is common for a carcinoma to grow partially beneath the skin, and so at the periphery it is not unusual to see hair follicles and interfollicular epidermis that differ in histology as well as color pattern from the tumor (**Supplementary Fig. 3**), which we excluded from analysis. Further, although the streaks observed in papillomas could be detected in these carcinomas as well as carcinomas that developed from mice in the skin-labeling experiment described above, they were restricted to the periphery of the tumor (**Fig. 4g**). This suggests that whatever role these genetically “normal” streaks might play in papilloma development, they are not essential to the tumor after progression to malignancy.

We conclude that the clonal evolution within non-progressing premalignant tumors is fundamentally different from that in benign lesions that progress to malignancy within the same time frame. Progression to malignancy is accompanied by the sweep of a single clone, which

displaces or outgrows the subclones it previously co-existed alongside, and this sweep is specific to progression.

Emergence of dominant cell clones during carcinoma progression

Labeling of the early stage (8 week) papillomas showed that the non-progressing lesions maintained multiple distinct cell clones over a long period of time, while all of the progressing malignant lesions had selected only a single dominant clone. In order to investigate the dynamics of selection of the dominant clone(s), we labeled papillomas at a later stage, treating with tamoxifen to activate Confetti labeling at 24 weeks (**Fig. 5a, b**). As in the 8-week labeling experiment, papillomas which did not progress to carcinomas over the following 6 months exhibited multiple colored regions. Carcinomas that emerged in this experiment, however, showed a range of labeling patterns, which fell into three categories: multi-colored, speckled, and single-colored. These patterns appeared to correlate with the latency between tamoxifen labeling and carcinoma appearance and harvest (**Fig. 5c**), suggesting that we were able to observe snapshots of tumor dynamics at distinct points in time surrounding progression, depending on how far ahead of progression the papilloma was labeled.

In 5 carcinomas harvested between 2 and 6 weeks after labeling, we observed numerous distinctly-colored clones growing side by side (**Fig. 5d,e**). Evidently, at the time we labeled these 24 week papillomas, the dominant progressing clone had already emerged, even though the tumor still appeared benign by visual inspection, and the various colored compartments in each carcinoma were derivatives of this progressing clone that were capable of contributing to continued carcinoma growth. We observed that at the intersection of colored subclones, mixing of cell populations could be seen (**Fig. 5e**), and that subclones showed no distinction from each other by H&E (**Fig. 5d**, lower panel). Distinct subclones, interestingly, displayed differential proliferative capacities, as shown by differential levels of Ki67 staining

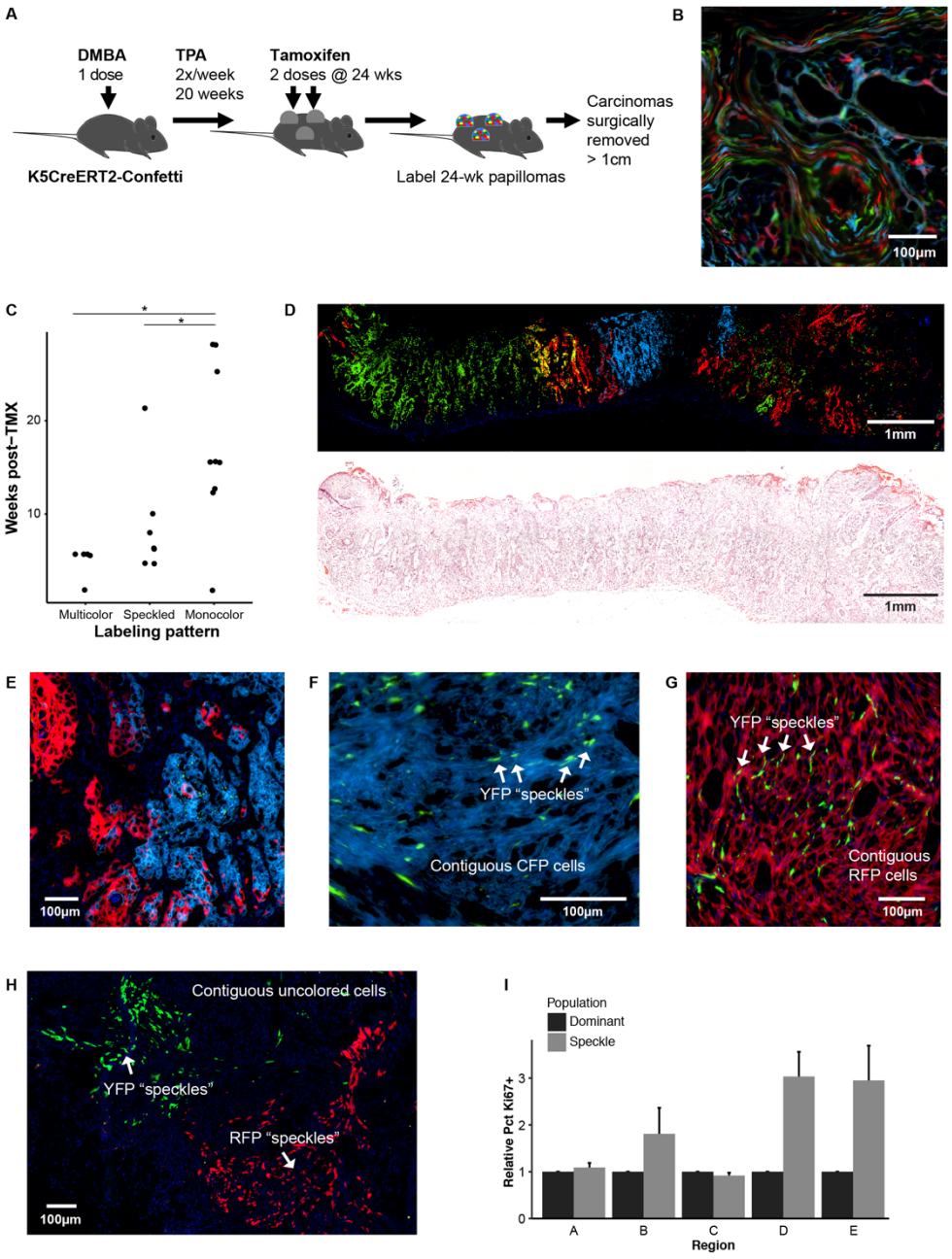


Figure 5. Carcinoma evolution following Confetti labeling at 24 weeks. (**a**) Schematic of experimental design. DMBA/TPA-induced papillomas were allowed to grow for 24 weeks. Confetti labeling was activated 24 weeks after DMBA treatment with 2 doses of tamoxifen. (**b**) Cross-section of a papilloma labeled at 24 weeks post-DMBA, 3 days after final dose of tamoxifen, exhibiting Confetti labeling. (**c**) Time between tamoxifen labeling and carcinoma harvest, compared with labeling pattern observed in carcinoma at harvest. (**d**) Cross-section of a multi-color carcinoma exhibiting patches of all Confetti colors, showing Confetti labeling (top) and H&E of adjacent section (bottom). (**e**) Magnified region of multi-color carcinoma shown in panel (**d**), showing intermixing of colored cells at a boundary between RFP, CFP, and GFP clones. (**f, g**) Examples of carcinomas containing "speckles." CFP-dominant carcinoma with YFP speckles (**f**) and RFP-dominant carcinoma with YFP speckles (**g**). Example "speckle" cells indicated with arrows. (**g**) Carcinoma containing multiple patches of differently-colored speckles. Carcinoma is dominated by uncolored cells and contains YFP and RFP speckles, each localizing to distinct regions.

(**Supplementary Fig. 4**). Whether the more proliferative subclone(s) would eventually come to dominate the carcinoma is unclear.

In contrast to this highly multi-color pattern in carcinomas that emerged close to the time of labeling, carcinomas harvested between 5 and 10 weeks after tamoxifen labeling typically exhibited a dominant color clone, however, many of these carcinomas also exhibited "speckled" patches in which cells of a distinct color were locally intermixed with the contiguous, dominant color clone (**Fig. 5f,g**). Carcinomas could have multiple such local speckle patches of distinct colors (**Fig. 5h**). These speckles displayed a pattern reminiscent of the border between two clones in the multi-color carcinoma (**Fig. 5d**), and were indistinguishable by H&E. Squamous tumor speckles were positive for K14 (**Supplementary Fig. 5**), consistent with this pattern being the result of intermixing of two tumor clones.

To address the possibility that the speckled subclones were the remnants of clones being outcompeted by the dominant clone, we quantified localized Ki67 levels in both populations. These data revealed that the speckled subclones were nearly always growing at the same rate or faster than the dominant clone with which they were locally intermixed (**Fig. 5i**). These observations, along with proliferation data from the multi-color tumors, rather indicate that carcinoma growth is driven by several, intermixed subclones, and further that these subclones emerge after the clonal sweep associated with progression. This pattern is distinct from the equipotent clones observed in papillomas, which tended to remain localized to specific lobes rather than displaying the broad intermixing seen in carcinomas. This property of intratumoral clonal mixing may reflect the increased capacity for invasion and migration that is associated with progression to malignancy.

Finally, we also observed carcinomas which were single-colored, the majority of which were harvested 12+ weeks after tamoxifen labeling. In these cases induction of Confetti labeling occurred far ahead of progression, and so we expect that the clonal sweep associated with progression occurred after labeling, as in the 8-week papilloma labeling experiment described

above. Based on these snapshots of clonal dynamics in carcinomas at distinct points in time, we conclude that, while an initial clonal sweep is associated with progression, this is followed by the emergence of multiple subclones which drive carcinoma growth, and which readily intermix with one another. These subclones can exhibit differential proliferation rates and it is possible for a new dominant clone to emerge, resulting in ongoing tumor evolution.

Metastasis can be polyclonal

The evidence for the clonal development of metastasis in controversial. While the existence of driver genes and mutations for initiation and tumor progression is well documented (e.g., *RAS*, *APC*, *TP53* or *PTEN*), there is presently no consensus regarding the existence of specific common driver mutations for metastasis. An alternative viewpoint is that metastasis is driven by a stem cell gene expression program, and that multiple cells within the primary tumor may be endowed with the capacity to disseminate and seed distant metastatic growth. The process of metastasis is extremely complex, itself involving multiple stages of invasion through the basement membrane at the primary site, intravasation and dissemination followed by extravasation at distant sites[106]. For this reason, many studies of metastasis focus on mouse models in which tumor cells are injected via the tail vein or aorta, thus missing steps that may involve clonal selection. Subcutaneous or orthotopic injections of tumor cell lines circumvent this problem, but also do not replicate the clonal evolution and genetic heterogeneity that occurs during primary tumor growth. Metastases from the skin carcinogenesis model, which develop after surgical resection of the primary tumor thus mimicking the course of human clinical practice, uniquely allow interrogation of the patterns of clonal evolution that take place during metastasis.

We investigated patterns of metastasis in all cohorts of mice in which labeling was induced either early (pre-initiation or 8 weeks) or late (24 weeks). Mice were monitored until primary carcinomas reached a size of 1cm diameter, at which point they were surgically

removed, thus enabling prolonged survial of the animals and subsequent harvesting of any metastases that developed. Metastases from mice in which labeling occurred at either early timepoint (pre-initiation, or in mice with 8 week papillomas) were uniformly single-colored (**Fig. 6a, b**), consistent with the carcinomas in these mice. Further, in the majority of mice in which multiple metastases were collected, metastases to distinct locations were the same color as one another, with as many as 8 matching-colored metastases found in one mouse (**Fig. 6a, b**). This is in agreement with previously published phylogenetic trees from this model demonstrating that most commonly all metastases arise from the same primary tumor[28].

In contrast to the uniform clone colors in metastases from these early labeling experiments, we observed metastases comprised of multiple distinctly-colored cellular populations in 3 of the 5 mice (60%) in the 24 week labeling experiment (**Fig. 6c**), the only experiment in which multi-color carcinomas were present. In one case, metastases to two lymph nodes and to the lung were all comprised of a mix of RFP and YFP cells (**Fig. 6c**); notably, no GFP or CFP cells were observed in any of these.

We submitted samples from two of the three cases of multi-color metastasis for exome sequencing, to determine whether the different cell populations in each metastasis were from the same primary tumor. Our previous work has demonstrated that DMBA-induced tumors carry a unique signature of dozens to hundreds of A>T mutations[28], thus enabling confident assignment of metastases to their primary tumor of origin. The first case we sequenced was a lymph node metastasis in which the majority of cells were uncolored, but 25% of K14+ cells (i.e., tumor cells) in the lymph node were GFP+ (**Fig. 6d**) and intermixed with the uncolored cells in a speckling pattern. This particular mouse had three primary carcinomas, and exome sequencing revealed that only one (Carcinoma A) had given rise to the lymph node metastasis (**Fig. 6e,f**), sharing 350 mutations (**Fig. 6d**). We could not initially be certain whether these 350 mutations were present in both the uncolored and GFP+ cells in the metastasis, so we sought to determine whether it was possible the two populations had originated from two distinct primary

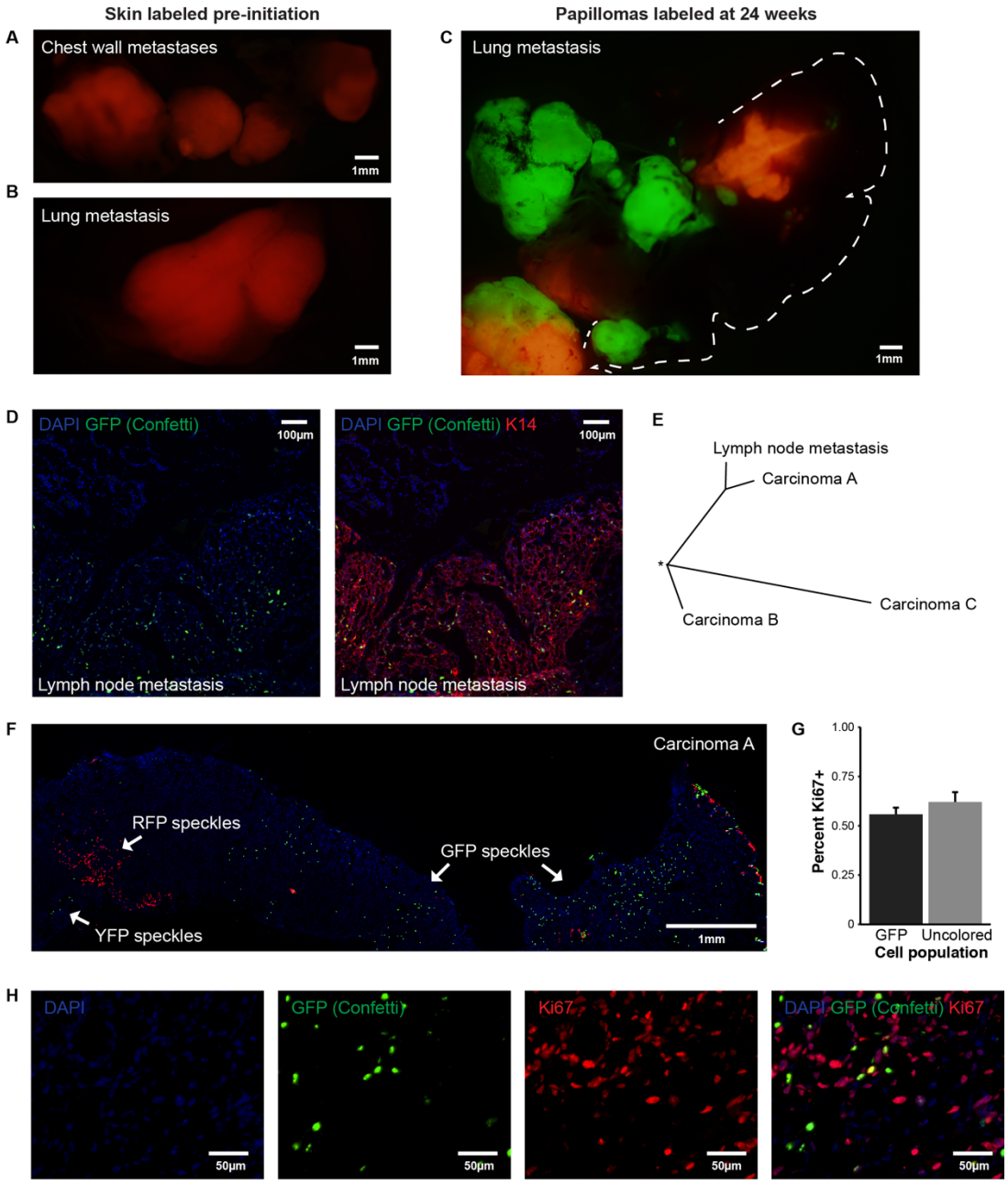


Figure 6. Evidence for polyclonal seeding of metastasis. (**a**,**b**) Metastases to the chest wall (**a**) and lung (**b**) from a mouse in which skin was labeled pre-initiation; all metastases are from the same mouse. (**c**) Lung metastasis from a mouse in which papillomas were labeled at 24 weeks. (**d**) Lymph node metastasis exhibiting dominant uncolored cell population and GFP+ speckles (left) and K14 staining of the same panel (right). (**e**) Phylogenetic tree showing relationship between all tumors in mouse bearing lymph node in panel (**d**). Carcinoma A (shown in panel (**f**)) and the lymph node metastasis share 350 mutations. (**f**) Cross-section of sequence-matched Carcinoma A that gave rise to lymph node metastasis in panel (**d**). Carcinoma contains YFP, RFP, and GFP speckle patches indicated with arrows. (**g**) Quantification of Ki67 staining of GFP+ and uncolored cells in the GFP speckle region in Carcinoma A, adjacent to right-most arrow in panel (**f**). (**h**) Illustrative Ki67 staining of GFP speckle region, quantified in panel (**g**).

tumors. We asked whether any of the mutations detected in the other carcinomas in this mouse (Carcinoma B and Carcinoma C, **Fig. 6e**) were present at low levels in the metastasis, particularly in loci well-covered by sequencing reads (>50x), and found they were not. We also examined mutations in the metastasis that were not shared with the matched Carcinoma A for contributions from an unidentified primary tumor; if these nonshared mutations were contributions from a separate primary tumor, we would expect to see a fingerprint of DMBA-induced T>A signature mutations[28]—however, we did not, and found only 4 T>A mutations in the metastasis that were not present in Carcinoma A. We conclude that no evidence exists in the sequencing data that would support the possibility that the uncolored cells and GFP+ cells in this metastasis originated from distinct primary tumors but rather, both cell populations arose from the same, sequence-matched primary tumor, Carcinoma A.

Interestingly, Carcinoma A was a predominantly uncolored tumor with a significant GFP+ speckle region, visible in the middle and right of the tumor cross section (**Fig. 6f**). Both the uncolored cells and GFP+ cells in this region were highly proliferative (**Fig. 6g,h**, 62% and 56% Ki67+ respectively). This carcinoma also contained YFP+ and RFP+ speckle regions, visible at the left side of the cross section (**Fig. 6f**). However despite the presence of these YFP+ and RFP+ cells in the primary tumor, only the GFP+ and uncolored populations contributed to metastasis.

We performed exome sequencing on a second case of multi-color metastasis as well, a lymph node metastasis from another mouse which contained both RFP+ and uncolored K14+ tumor cells (**Supplementary Fig. 6a**). In this case, we sorted the RFP+ and uncolored populations by FACS, sequenced, and again confirmed that they originated from the same primary tumor on the basis of shared mutations and shared copy number alterations (**Supplementary Fig. 6b**).

We conclude that metastases, in contrast to the earlier stages of tumor progression, do not arise from a single cell within the primary tumor that has acquired metastatic properties, but

rather that multiple cells within progressed lesions have the capacity to disseminate and seed at distant sites. These data therefore agree with a polyclonal model of metastatic dissemination.

DISCUSSION

Papillomas have multiple cell populations but clonal origins

The hypothesis that cancers arise from a single cell was originally based on estimates of mutation frequency in cells treated with a chemical mutagen[4,107]. The argument was that because the mutation frequency is very low, the probability that two adjacent cells would both suffer the initiating mutation and give rise to a polyclonal tumor was so low as to be negligible. Analysis of X-linked markers in human tumors derived from heterozygous females[108,109] or chimaeric mice[19,21] yielded evidence both for and against this hypothesis.

The debate regarding cellular origins of cancer was further complicated by the observation of patches of histologically normal cells carrying mutations in cancer associated genes[91,92] and by the concept of “field cancerization” in which many cells within a tissue region are predisposed, possibly by an inflammatory process or activated stromal cells, to develop into progressively growing lesions (for review see Dotto[110]).

We have exploited multi-color lineage tracing using the Confetti mouse to investigate the clonal dynamics that govern each stage of tumor progression from initiation to metastasis. Our observations of multiple cell populations in skin papillomas induced by treatment with DMBA and TPA were strikingly consistent a previous report by Winton et al, who used H2-chimaeric mice to demonstrate the existence of multiple cells of origin within early papillomas[21]. These authors concluded that papillomas always had a dominant clone, but that secondary clones could also be detected, leading to the speculation that papillomas were sometimes polyclonal in origin. Sequencing specifically of the streak populations however allowed us to conclude that these cells were not co-initiators of the papilloma, as they lacked genetic changes including the initiating *Hras* mutation. The streaks we observed were reminiscent of a similar radial streaking

pattern seen in the skin when Krt15+ hair follicle bulge stem cells were labeled prior to wound healing[101]. These authors observed the growth of streak populations emanating from the stem cell region and growing into the healing wound over time, calling to mind the analogy of a tumor to a wound that never heals[111].

Whether the streak populations in papillomas, which increase substantially in size and number between 12 and 20 weeks post-initiation, play a positive role in papilloma growth, for example by supplying growth or niche factors, is presently unclear. However, their presence appears to not be necessary for progression to carcinoma, as they become increasingly marginalized during malignant conversion and residual streaks are only found on the periphery of squamous carcinomas. Interestingly, Krt15-positive stem cells also make only a transient contribution to normal healing wounds and do not persist over long periods of time[101,112].

The paucity of point mutations and the typical DMBA mutation signature in these streak cell populations is surprising in view of the fact that they must have been adjacent to cells that suffered many mutations and gave rise to clonal papillomas. The whole of the back skin was treated with the same concentration of DMBA, suggesting that different cell populations in the exposed skin are not in fact mutated equally. This could be due to differential expression of the metabolic enzymes that lead to activation of DMBA to its ultimate carcinogen DMBA-*trans*-3,4-dihydrodiol-1,2-epoxide[113], or differential repair of adducts formed with DNA. Differential immune responses to mutant cells in particular locations in the epithelium may also contribute to this discrepancy. For example, cells of the bulge region have been proposed to be immune privileged, and therefore possibly protected from immune attack[114]. Notably, the non-mutated cells are capable of giving rise to actively growing progeny and must have some stem cell capacity at the time of mutagen treatment. The properties of these cells, and where they are located in the skin, are questions of significance for future investigation.

Rapid emergence of clonal dominance is seen in carcinomas but not in papillomas

Labeling of established papillomas at 8 weeks or 24 weeks after initiation allowed us to detect distinct patterns of clonal evolution during progression to malignancy. Activation of the Confetti allele at 8 weeks led to efficient labeling of epithelial cells, but after 6 months of growth papillomas consisted mainly of 2-4 distinct colored clones. Whether these clones arise from a subpopulation of rare papilloma stem cells[95] or from a process akin to neutral drift that has been shown to govern patterns of clonal growth in normal stratified epithelia[115] is presently unclear. However, carcinomas and their metastases from this same cohort of mice were exclusively comprised of a single clone, indicating that a clonal sweep had taken place resulting in the emergence of one dominant clone. This comparison was particularly striking even in age-matched papillomas and carcinomas from the same mice, indicating that rapid clonal emergence is a feature of malignant progression.

Labeling papillomas at a later time point (24 weeks) enabled us to visualize additional features of clonal competition leading to selection of a dominant cell population. Five papillomas that progressed to carcinomas within 6 weeks of labeling showed the presence of multiple adjacent clones similar to what was seen in long lived papillomas, with clones intermixing at the boundaries of each compartment. Papillomas labeled further ahead of progression frequently showed a dominant clone, intermixed in a speckling pattern with cells from other clones. This intermixing of cells of different clonal origins was not seen in papillomas, and reflects the intrinsic propensity of carcinoma cells to invade and migrate both into adjacent tumor as well as normal tissue. The presence of these intermixed, proliferating populations reveals that carcinoma growth is driven by multiple cellular populations following the initial clonal sweep. However, new dominant clones may subsequently emerge and be responsible for ongoing evolution after progression, and future studies of these distinct populations will be useful in shedding light on this continuing evolution.

Multiple cell populations can contribute to metastases

Metastasis arises as a consequence of dissemination from the primary tumor site, followed by seeding and progressive growth in local lymph nodes or at distant sites. Sequencing studies have demonstrated that metastases share many trunk mutations with corresponding primary tumors, and therefore arise from the original dominant clone. This is seen in our analysis of metastases derived from papillomas that were originally labelled at 8 weeks, as these gave rise to carcinomas and matched metastases from the same colored clone. There is no consensus that metastasis is driven by new driver mutations in single primary tumor cells, resulting in a further clonal sweep at this stage. Indeed, at the cellular level, there is evidence that metastasis can involve multiple independent cells derived from the primary carcinoma, either through concerted invasion of groups of cells, or by sequential seeding of different single cells from the same primary tumor[23,24,96]. Deep sequencing of human tumors has offered support for this theory, with genetic events from distinct primary tumor subclones being simultaneously detected in metastases[98,99].

In metastases from the tumors we labeled at 24 weeks, we observed that multiple cellular populations in fact frequently participated in metastasis. These findings align with data from a genetically engineered mouse model of pancreatic cancer, in which polyclonal tumors give rise to circulating tumor cell clusters and metastases containing cells from multiple tumor clones[97]. However primary tumors in this model arise from a field induction of activating *Kras* mutations and *Trp53* deletions, yielding polyclonal tumors in which every clone has been identically genetically initiated and thus is independently malignant. Our findings represent the first case of successful lineage tracing of multiple cellular populations from a single, clonally selected primary tumor to the same metastasis, in a model that accurately recapitulates the natural progression of human disease.

Lineage tracing offers a distinct advantage beyond sequencing, in that it enables the identification of distinct cellular populations in which differentiating genetic events may not be

detectable, and further offers information about the spatial organization of distinct clones, which may provide clues as to the nature of the polyclonal contributions. In two cases we observed, some regions of the metastasis were dominated by one contributing clone or the other, while a limited number of regions showed intermixing. It is possible that this pattern arose from sequential seeding of the metastasis by one clone followed by another, or alternatively that each clone was better suited to grow in particular local microenvironments over the other. In the third case, however, the metastasis was thoroughly "speckled" with intermixed GFP and uncolored cells in all regions of the lymph node. In this case, where no localization of either populations was observed, it is tempting to speculate that the metastasizing cells must have arrived together and co-seeded the metastasis. Further study of these patterns and their frequencies, as well as genetic and molecular studies of the contributing populations themselves, will be informative in understanding how metastasis come to arise.

Finally, we observed that while multiple cellular populations from a primary tumor could contribute to metastasis, not all primary tumor populations did. In the GFP-speckled metastasis, the matched primary tumor contained RFP and YFP cellular populations which did not metastasize. In another case, we observed that multiple metastases to distinct anatomical locations carried the exact same two cellular populations. Whether these different populations have specific functions that interact to promote tumor growth or invasion, as suggested by other mouse models[89,90,116] or are simply derivatives of the same trunk clone with similar properties, is presently unclear.

The combination of Confetti lineage tracing with the DMBA/TPA skin carcinogenesis model has enabled interrogation of clonal dynamics of all stages of tumor progression, revealing that while initiation and progression are clonally driven events, metastasis is a fundamentally different process involving the participation of multiple cells. In contrast to genetically engineered mouse models which contain few genetic alterations and low levels of heterogeneity and cellular diversity, models which recapitulate the clonal and evolutionary dynamics and the

heterogeneity of human tumors are critical for advancing our understanding of how genetically complex tumors evolve, and how they can ultimately be targeted and treated.

METHODS

Confetti labeling. Tamoxifen (Sigma) was dissolved overnight in sunflower seed oil at a concentration of 10mg/mL. Mice were treated with 400μL (4mg) per dose, applied to the back skin topically, and were shaved the day prior to first treatment. For skin activation experiment, mice were given 4 doses every other day for a total of 16mg. For tumor activation experiments, mice were given 2 doses for a total of 8mg, spaced 2 days apart, at either 8 or 24 weeks.

Carcinogenesis. Male and female *K5CreER-Confetti* FVB/N mice were shaved and treated with 25mg DMBA dissolved in 200μL acetone either 10 days after final dose of tamoxifen (skin activation experiment) or at 8 weeks of age (tumor activation experiments). Mice subsequently received TPA (200μL of a 10^{-4} M solution in acetone) twice a week for 20 weeks, following established chemical carcinogenesis protocol[27]. Carcinomas were surgically resected when they reached a size of >1 cm in longest diameter, and mice given 0.24mL meloxicam (Boehringer Ingelheim, 5mg/mL solution) for recovery. Mice were sacrificed when disease progressed, per animal care requirements. At sacrifice, papillomas and carcinomas were removed from skin, and all internal tumors were resected.

Tissue harvesting. Skin, tumor, and metastatic tissues harvested for sectioning were kept at 4°C in 10% formalin overnight, in a gradient of 15% / 20% / 30% sucrose on the second day, in 30% sucrose overnight the second night, and then embedded in OCT (Tissue-Tek, Sakura) and stored at –80°C until sectioning.

Digestion for FACS. For tumors that were FACS sorted, a piece of the tumor was first removed for embedding in OCT and imaging. For carcinomas, skin tissue along the edge of the carcinoma was also removed. Tumors were then finely chopped, washed with PBS, digested in 4mg/mL Collagenase A (Sigma-Aldrich) for 1 hour at 37°C, and then resuspended in 0.25% Trypsin-EDTA (Gibco Life Technologies) and incubated on a shaker at 37°C for 1 hour. Trypsin was neutralized with an equal volume of FBS, and digested tumor filtered through a 40μm filter, pelleted, and resuspended in FACS buffer (2% FBS) for sorting.

Quantification of Confetti labeling. To quantify Confetti activation in the skin, a strip of skin oriented along the spine was collected for 5 mice 10 days after the final dose of tamoxifen, e.g., the day that DMBA treatment would have begun. To quantify Confetti activation in tumors, papillomas were surgically removed 3 days after the final dose of tamoxifen. Tissues were harvested and embedded in OCT as described above, and 5μm sections were taken for quantification. For each sample, a minimum of 10 images were taken at 40× with a 6D Nikon microscope, and nuclei and colored cells of each color were manually counted in ImageJ.

Quantification of multi-color papillomas. At sacrifice, back skin was removed intact from mouse and imaged whole on an MVX10 fluorescent stereoscope, which could distinguish RFP, CFP, and YFP/GFP. YFP and GFP were not distinguishable from each other under this microscope. Tumors on each back skin were counted, and color(s) recorded. For statistical calculations (e.g., Figure 2C), mice which had fewer than 3 papillomas were excluded as they did not provide sufficient data (this led to the exclusion of 2 mice).

Nucleic acid extraction. DNA was extracted from tumors using Qiagen DNeasy Blood & Tissue Kit, following manufacturer's instructions. In cases where DNA was extracted after FACS sorting, protocol was modified as follows to improve yield: (1) After sorting, FACS collection tube

was spun down at 3,000rpm for 10 minutes and cells resuspended in 200μL PBS before initial lysis step; (2) Wash step with buffer AW2 was done twice. DNA concentration and quality were determined by Nanodrop spectrophotometry.

Sanger sequencing of *Hras* locus. Hras locus containing codon 61 was PCR amplified using primer pair AAGCCTGTTGTTTTGCAGGA (forward) and GGTGGCTCACCTGTACTGATG (reverse). PCR product was purified using Exonuclease I (USB) and Shrimp Alkaline Phosphatase (Affymetrix), and Sanger sequencing was performed using the forward primer listed above by MCLAB. Images were taken using FinchTV.

Exome sequencing, alignment, and quality control.

DNA samples were submitted to Otogenetics Corporation (Atlanta, GA) for mouse exome capture and sequencing. Illumina libraries were made from qualified fragmented gDNA using SPRIworks HT Reagent Kit (Beckman Coulter) and the resulting libraries were subjected to exome enrichment using SureSelectXT Mouse All Exon (Agilent) following manufacturer's instructions. Enriched libraries were tested for enrichment by qPCR and for size distribution and concentration by an Agilent Bioanalyzer 2100. The samples were then sequenced on an Illumina HiSeq2500 using Rapid v2 SBS chemistry which generated paired-end reads of 106 nucleotides.

Sequence alignment, processing and quality control

Reads were mapped to the GRCm38/mm10 version of the *Mus musculus* genome using BWA (version 0.7.12)[75] with default parameters. The Picard MarkDuplicates module was used to remove duplicates from the data (version 1.131; **http://broadinstitute.github.io/picard**). The Genome Analysis Tool Kit (GATK-Lite) toolkit (version 2.3-9)[117] module IndelRealigner and BaseRecalibrator were used to preprocess the alignments.

During base quality recalibration, dbSNP variants were used as known sites, according to GATK Best Practices recommendations[76]. Finally, alignment and coverage metrics were collected using Picard. We sequenced an average of 42 million unique reads per sample. Targeted bases were sequenced to a mean depth of 50, and more than 75% of targeted bases were sequenced to 20× coverage or greater.

Variant calling. SNVs were called using somatic variant detection program MuTect (version 1.1.7)[77]. Each tumor was called against its matched normal tissue (tail), and calls were filtered against a database of known *Mus musculus* germline SNPs available at ftp.ncbi.nih.gov/snp/organisms/mouse_10090/VCF/genotype, as well as against a panel of normal tails from this experiment. Results were further filtered to calls with a minimum read depth of 10 at the locus for both tumor and matched normal, and to calls where at least one alternate read had a mapping quality score of 60 or higher. Variants were annotated using Annovar (downloaded on 2/4/2016)[78], and these annotations were used as the basis for assessing exonic variants as synonymous, nonsynonymous, stopgain, or stoploss.

Copy number calling was done with CNVkit[81], and tumor copy number status was called against a panel of normal tails from the same sequencing batch.

Mutation context. For mutation spectrum analysis, SNVs in all tumors were annotated with 1 of 96 possible trinucleotide context substitutions (6 types of substitutions x 4 possible flanking 5'-bases x 4 possible flanking 3'-bases), using MuTect output, and counts of each mutation context were summed.

Phylogenies. To build phylogenetic trees, absolute distance matrices were calculated based on the presence of mutations in the sample, based on filtered MuTect calls. Rooted trees were built

with use of the Analyses of Phylogenetics and Evolution (APE) package and manhattan calculation method implemented in R version 2.15. Relationships between metastases and primary tumors were determined on the basis of shared mutations.

Quantification of colored lobes in papillomas labeled at 8 weeks. Papillomas were imaged whole on an MVX10 fluorescent stereoscope. Images were taken to record observations, and these whole-tumor images were used to count the number of distinctly-colored lobes on each tumor. It was possible to identify RFP, CFP, YFP/GFP, and uncolored lobes, however YFP and GFP could not be distinguished from each other using this microscope. Selected tumors from this cohort were also embedded in OCT as described above, sectioned, and 5μm sections were imaged with a Nikon 6D scope to confirm lobe quantifications done with the whole-mount tumor. Results from imaging these sections correlated well with whole-mount observations. For statistical analysis, in comparing number of colored lobes in papillomas and carcinomas, we excluded fully-uncolored papillomas and carcinomas because it is impossible to distinguish a monoclonal tumor from a poorly-labeled polyclonal tumor with multiple uncolored subclones; thus, we included only tumors where at least one Confetti color was visible in the analysis.

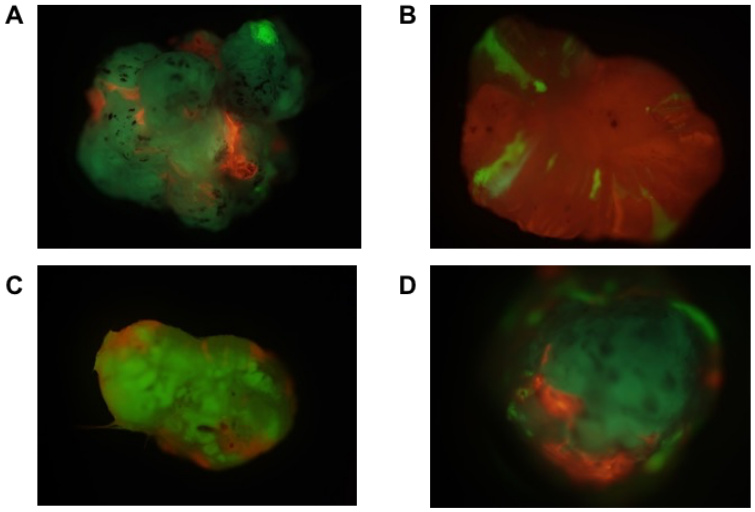
Classification of labeling pattern in tumors labeled at 24 weeks. For carcinomas harvested in 24 week labeling experiment, all carcinomas were embedded in OCT as described above, sectioned, and imaged with a Nikon 6D microscope in order to classify their labeling pattern. At least two distinct pieces of each carcinoma were used, and at least 3 serial sections taken 100μm or more apart. Tumors classified as multi-color contained large, contiguous patches of at least 2 of the 4 Confetti colors. Tumors classified as speckled showed one or more “speckle” populations of cells that were a distinct color from the surrounding tumor cells, and which formed localized patches of non-contiguous cells (in contrast to contiguous patches observed in “multi-color” tumors), and which could be identified in at least 3 serial sections from the tumor.

Tumors classified as single-color contained only one color population of cells in the tumor, excluding hair follicles and intrafollicular epidermis that was sometimes present at the edge of the tumor.

Immunofluorescent staining. Slides with 5μm tumor sections were brought to room temperature, post-fixed in 4% paraformaldehyde for 8 minutes, and washed in PBS for 5 minutes. For Ki67 staining, slides were blocked in 5% goat serum/0.3% Triton-X100 for 1 hour. Ki67 antibody (Cell Signaling #9129) was used at a concentration of 1:300 in 2% goat serum/0.3% Triton-X100 and left on slides overnight. Slides were washed 3x in PBS, and incubated with goat anti-rabbit Alexa Fluor 647 (Thermofisher) secondary antibody at a concentration of 1:200 for 90 minutes, and then washed 2x with PBD (PBS with 0.1% Tween-20) followed by PBS. For K14 staining, slides were blocked in 10% donkey serum/0.1% Triton-X100 for 15 minutes. K14 antibody (Biolegend #905301) was used at a concentration of 1:2000, and allowed to incubate on slides for 2 hours. Secondary antibody of either donkey anti-rabbit Alexa Fluor 594 (Thermofisher) or Alexa Fluor 488 (Thermofisher) was selected to avoid interference with Confetti colors in relevant samples, and slides were incubated with secondary antibody at a concentration of 1:500 for 1 hour and washed 3x with PBS.

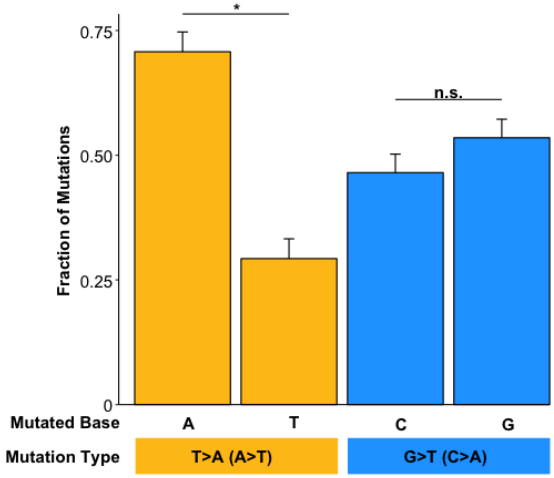
Quantification of Ki67. Tumor sections stained for Ki67 were imaged with a Nikon 6D microscope. Quantification was based on manual counts of at least 3 images taken at 40x magnification, where individual cells could be identified and classified based on both presence or absence of Ki67 and Confetti labeling color.

Supplementary Fig. 1



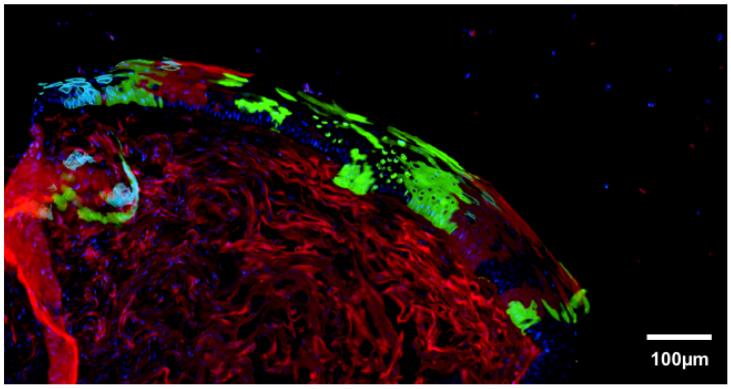
Whole-tumor strereoscope images of the 4 multi-color papillomas which were separated by FACS for sequencing of ”bulk” and “streak” populations. Populations for which sufficient DNA existed to sequence were as follows: (A) CFP bulk, RFP streak; (B) RFP bulk, YFP streak; (C) YFP bulk, RFP streak; (D) CFP bulk, RFP and YFP streak.

Supplementary Fig. 2



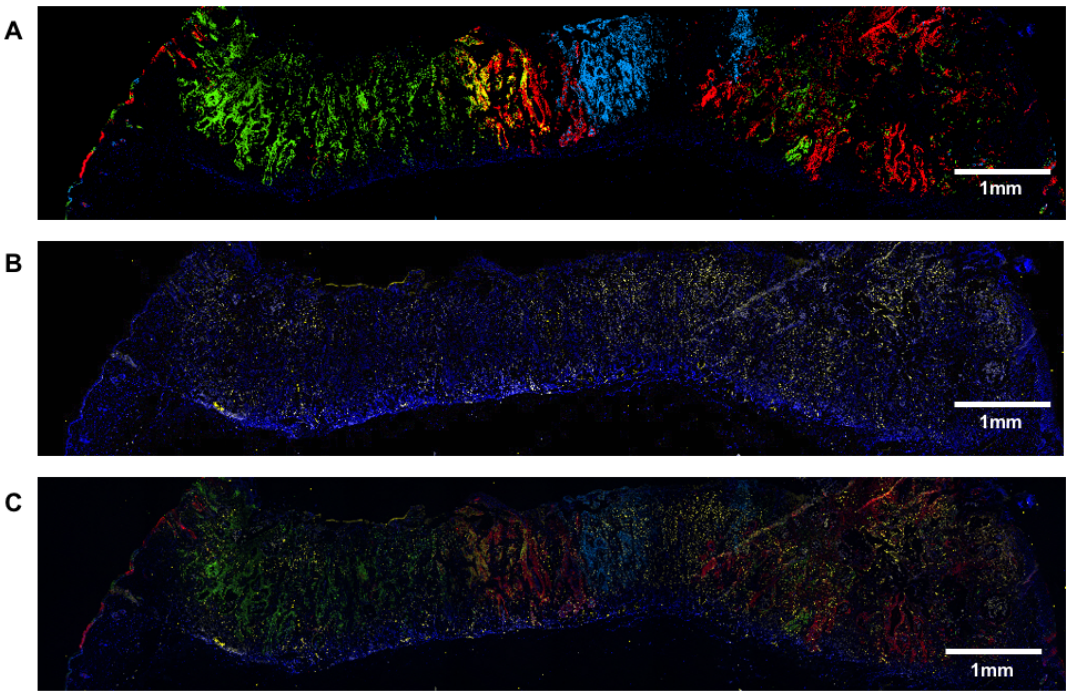
For each mutation type, frequency of each base pair on the coding strand being mutated. In T>A(A>T) mutations, it was significantly more common for the coding strand to contain an “A” than a “T” at the mutated site. For G>T(C>A) mutations, there was no bias as to whether a “G” or a “C” base was present on the coding strand at the mutated site.

Supplementary Fig. 3



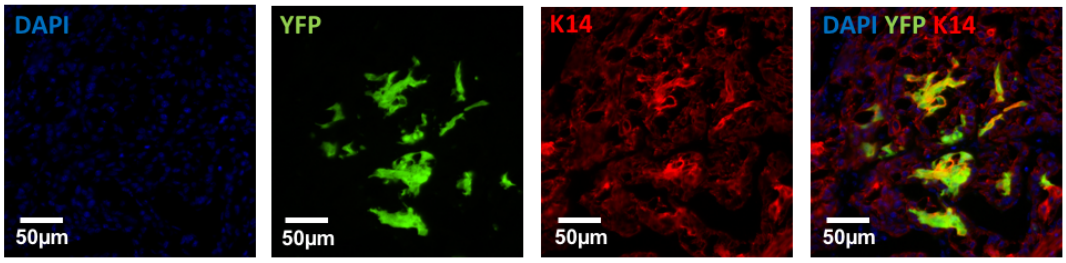
The edge of a single-colored RFP+ carcinoma, showing the tumor growing under normal intrafollicular epidermis. The epidermis here, as elsewhere across the entire back skin, contains patches of all Confetti colors; however these are morphologically distinguishable from the carcinoma itself.

Supplementary Fig. 4



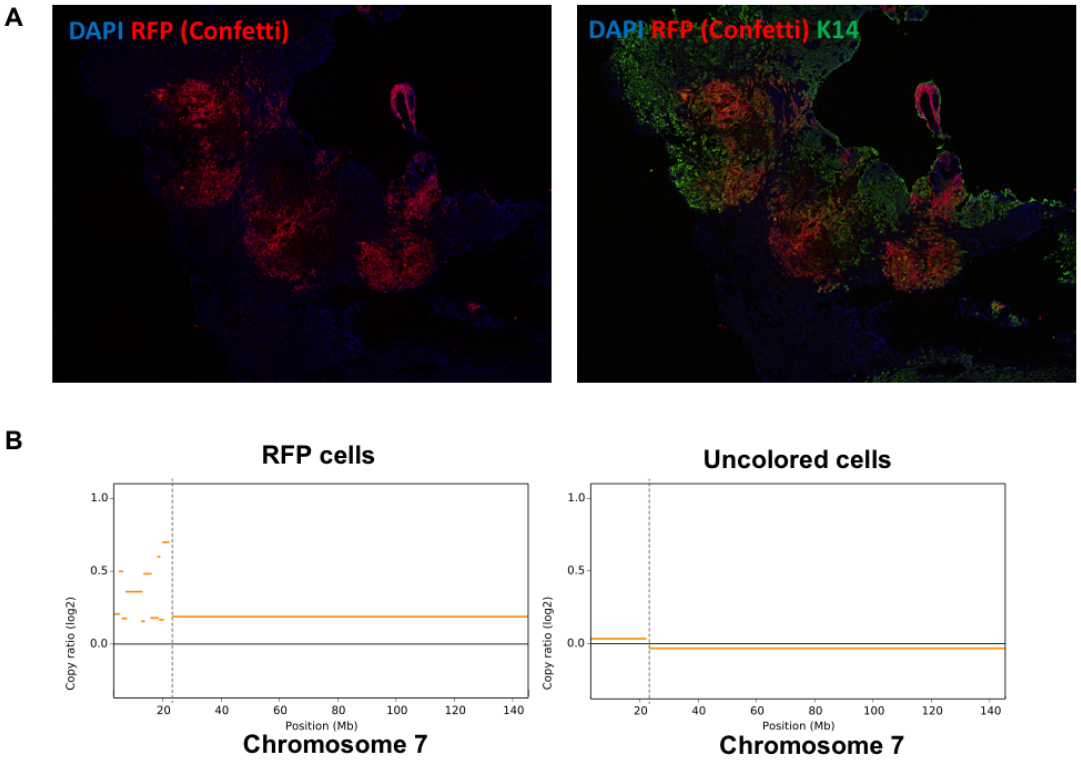
Ki67 labeling of a multi-color carcinoma from the 24-week labeling experiment, showing differential levels of Ki67 in different color clones. (A) Confetti labeling (B) Ki67 staining (C) merge, with Confetti colors muted to improve visualization of Ki67 localization.

Supplementary Fig. 5



The edge of a single-colored RFP+ carcinoma, showing the tumor growing under normal intrafollicular epidermis. The epidermis here, as elsewhere across the entire back skin, contains patches of all Confetti colors; however these are morphologically distinguishable from the carcinoma itself.

Supplementary Fig. 6



Analysis of a multi-colored lymph node metastasis. (A) Cross-section of lymph node (left) and K14 staining (right) (B) Copy number plot of chromosome 7, with proximal end of the chromosome at left and distal end at right. A focal copy number gain at the proximal end is seen in both the RFP+ (left) and uncolored (right) fractions. Note contamination of the uncolored cell fraction with lymphocytes results in dilution of signal and lower resolution; however the identical breakpoint (dashed line) is observable in both samples.

Chapter 3

Chemical Carcinogenesis Models of Cancer: Back to the Future

Introduction

We have known for over 250 years that cancer can be caused by exposure to an environment rich in toxic chemicals, but it is only in the last 50 years or so that the causal mechanistic links between chemical exposures, DNA damage, and cancer initiation have become clear. Following the early observations of scrotal cancers in chimney sweeps exposed to soot and coal tar[118], Ichikawa et al developed the first animal model of cancer by painting rabbit skin with coal tar to induce tumors[119], thus providing a definitive causal link between treatment and subsequent cancer development. In a series of classic experiments, mouse models were refined to show that cancer can develop through specific stages that could be associated with different types of chemicals[120]. A single exposure to a mutagenic chemical could lead to initiation, where DNA damage permanently disposed the treated tissue to cancer. However, tumors only grew out after subsequent and repeated exposure to other agents, known as tumor promoters, which appeared to act not through mutation induction but by stimulation of proliferation and inflammation[121,122]. These pioneering studies led to the concept of multistage carcinogenesis that is now widely accepted today as being applicable to both mouse models and human cancers, particularly in epithelial tissues such as the colon, skin, pancreas, and mammary gland.

We now have quite a sophisticated view of the mechanisms by which different classes of chemicals interact with DNA, forming adducts or causing base damage that, if improperly repaired, result in cancer-initiating mutations[123]. We have also learned, thanks to the human genome project and the huge databases of normal and tumor genomes that it has spawned, that hundreds if not thousands of genes, when mutated, can contribute to cancer initiation or progression. The view of cancer that has emerged is one of increasing complexity and

heterogeneity at the cellular and genetic levels, a microcosm of Darwinian evolution leading to selection of those cells that are best suited to their particular environment.

The purpose of this review is to chronicle the development of our present understanding of how tumors are initiated by exogenous chemical agents, and how these tumors acquire the capacity for malignant progression, evade the immune surveillance system, and ultimately metastasise throughout the organism. It may be asked why, in this era of precision genomics that has led to development of so many sophisticated genetically engineered mouse cancer models (GEMMs), would we focus on models involving “non-specific” mutagens that rarely if ever can be targeted to a specific tissue type or cell population? The reason is quite simple: while GEMMs have provided unprecedented opportunities to alter, at will, the germline of mice to induce specific events that increase cancer susceptibility, it is increasingly clear that they portray a vastly oversimplified view of the numbers and types of mutations that are found in human cancers[64,65]. GEM models frequently contain only a handful of point mutations, while human cancer genomes present a rich tapestry of point mutations, gene copy number changes and complex genomic events that are in large part due to exposure to exogenous agents that mold genome architecture. Analysis of thousands of human tumor genomes by whole exome or genome sequencing has shown that the frequency of point mutations can vary over several orders of magnitude, from less than 0.1 to greater than 50 mutations per megabase[80]. Tumors from patients exposed to a high concentration of carcinogens, for example lung cancers from heavy smokers, or skin cancers due to prolonged exposure to mutagenic UV radiation, have extremely high numbers of point mutations as well as considerable genetic heterogeneity. Other highly exposed tissues such as the esophagus, head and neck, and gastrointestinal tract also suffer many insults that increase point mutation burden, while the incidence of these lesions is lower in tumors from other tissues such as the prostate, mammary gland and brain, which are in a relatively less exposed environment. Point mutations and large scale genomic alterations in cancers are important for many reasons: they can confer properties that are critical

determinants of individual patient prognosis, responses to therapy or development of drug resistance[124]. Chemical carcinogenesis models in the mouse, which like environmentally induced human tumors carry a high mutation burden, are therefore uniquely able to replicate some of these cardinal genetic features of human cancers, and provide routes to addressing the many stubborn questions that remain unanswered in the struggle to understand and ultimately prevent or successfully treat human cancer.

The smoking gun of chemical exposure

Chemical carcinogenesis models also offer us new ways to definitively identify environmental agents that play causative roles in human cancers. The presence of carcinogens in the environment leaves an imprint, a kind of "smoking gun" signature, in the human genome that can act as a chronological record of mutagen exposure[46,125]. UV-induced melanomas carry a well-known pattern of C>T mutations predominantly at dipyrimidine sites, which are the target for UV-induced crosslinking[126,127], while lung carcinomas from smokers carry a mutation signature dominated by a G>T mutation pattern, which precisely replicates the trinucleotide context of G>T mutations induced by Benzopyrene in vitro in mouse cells[46,128]. From a total of 31 "mutation signatures" that have been identified by deep sequencing of human tumors, however, only 7 are attributable to known causative agents such as cigarette smoke or UV exposure[125]. The vast majority of these signatures have unknown causes, suggesting the existence of environmental agents that influence human cancer development in important ways, but which have not yet been identified. These important data have implications for cancer prevention as well as early detection of exposures in the human workplace. Just as was seen with the G>T signature of Benzopyrene, tumors initiated by carcinogenic agents leave the same genetic imprint in the genomes of mice as they do in humans[28,64,70,126]. Mouse lung tumors induced by a single treatment with N-methyl-N-nitroso-urea (MNU) show a genome-wide pattern of G>A transition mutations, while an alternative initiating mutagen, urethane, induces tumors

with either A>T or A>G mutations genome-wide[64,129,130]. Thus once the mutagen signatures are known, the identity of causative agents can be deduced from inspection of tumor sequences. These data offer exciting new possibilities to combine human tumor genome analysis and rodent chemical carcinogenesis models to associate mutation signatures with their causative agents, which could lead to identification of some of the many unknown environmental factors that are contributing to human cancer development.

Mutagenic activation of driver mutations by chemical carcinogens.

The discovery of human *RAS* oncogenes that were activated by single point mutations[1,131–134] raised the possibility that chemical mutagens could initiate cancer by targeting and introducing these same mutations. This possibility was first tested using the original mouse skin model of 2 stage carcinogenesis involving initiation with a potent chemical mutagen, 7,12-dimethyl-benzanthracene (DMBA), and promotion by multiple exposures to 12-O-tetradecanoyl-phorbol-acetate (TPA)[135]. Although three members of this gene family (*HRAS*, *KRAS* and *NRAS*) had been shown to be mutated in human tumors and cell lines, analysis of activation patterns in both early stage premalignant papillomas and late stage carcinomas from mouse skin showed remarkably consistent activation of the same gene, *Hras*, in almost all tumors[39]. The specific activating mutation was shown to be an A:T>T:A transversion at codon 61 of *Hras*[26,136], and was present in over 90% of benign and malignant tumors. Codon 61 encodes glutamine (CAA) containing 2 target adenosines, both of which when mutated to T can result in *Hras* activation. DMBA was much less likely to cause Ras activation by mutation at codons 12 or 13 (GGA and GGC respectively) as adducts with G are much less common. Tumors initiated by an alternative carcinogen (N-methyl-N-nitrosourea (MNU)), however, exhibited G:C>A:T transitions leading to activation of *Hras* by mutations at codons 12 or 13[137]. A similar activating mutation had also been seen in mammary carcinomas in rats treated with MNU[138] .These mutation patterns were in complete agreement with prior data showing that DMBA forms

adducts predominantly with adenosine residues in DNA[103], while MNU methylates guanines, predominantly at GG dinucleotide positions[130,139], thus establishing a mechanistic link between carcinogen exposure and specific point mutations in cancer genes.

A series of additional experiments demonstrated the induction of carcinogen-specific mutations in target genes by other chemical mutagens and in other tissues. In chemically induced lung tumors, *Kras* was found to be mutated much more commonly than *Hras*, but the chemical signatures were the same. MNU typically induced mutations in codon 12 of *Kras*[140], while urethane, which preferentially causes A:T>T:A and A:T>G:C mutations, predominantly induced mutations in *Kras* codon 61[129]. Polycyclic hydrocarbons such as Methylcholanthrene (MCA) or Benzopyrene primarily form adducts with guanine residues, and mis-repair of these bulky lesions results in G:C>T:A transversions, which was the main mutation type seen in Ras genes and in the p53 tumor suppressor gene in tumors induced by these agents (**Table 1**, see end of chapter)[128,137]. Additional examples exist of common agents in the human environment that cause cancer through induction of mutations consistent with their known genome-wide mutation signatures. Aflatoxin, a fungal metabolite found in peanuts, binds to guanines, resulting in G>T mutations at a recurrent hotspot in TP53[141,142] as well as throughout the genome[143]. Aristolochic acid (AA) is a constituent of plant extracts that have been used for medicinal purposes throughout antiquity. This agent acts as a strong mutagen both in humans and in rodent cancer models, primarily by forming adducts with adenosines, resulting in A:T>T:A mutations both in *Ras* oncogenes[144] and genome-wide[145]. These studies provided the rationale for present efforts to identify human environmental carcinogens based on mutation signatures in DNA from a wide variety of tumor types[46].

Factors leading to induction and selection of oncogenic mutations

A unique feature of mouse models of chemical carcinogenesis is the insights that they provide into the complex mechanisms leading to selection of specific mutations in oncogenes.

As seen above, carcinogens have a propensity to cause mutations at specific locations in the genome. Following the initial chemical insult, a variety of selective processes operating at the level of metabolism, gene target sequence, DNA repair, tissue type, cell of origin, and genetic background all factor into the eventual outgrowth of a tumor carrying a specific set of mutations (**Fig. 1**). As we will endeavor to show below, chemical carcinogenesis models are therefore not simply a "pepper-spray" approach to cancer modeling, but can be used to address specific mechanisms by which cancers can be initiated or promoted in subsets of tissues or cells by activation of select cancer driver pathways.

Beyond favoring mutations of particular DNA bases in particular trinucleotide contexts, patterns of carcinogen mutations are influenced by the genomic architecture of their target genes and target cells. Carcinogen-induced mutations in human cancers are clustered in regions of closed chromatin, resulting in mutation patterns that depend on the tissue in which tumors arise[146]. The particular DNA strand—coding or noncoding—on which adducts form can also play a major role in determining mutation specificity, with accumulating data suggesting that coding strand mutations are favored. Mutations in oncogenes and tumor suppressor genes are predominantly caused by formation of adducts with residues on the non-transcribed coding strand[147], presumably because similar adducts on the transcribed strand are subject to efficient transcription-coupled repair. A recent elegant study demonstrated how this mechanism can also account for the observation of *Hras* or *Braf* mutations in mammary tumors induced by different carcinogens[148]. In mammary tumors that were induced by expression of a *Wnt* transgene followed by carcinogen exposure, treatment of transgenic mice with DMBA resulted in almost 100% of driving mutations being CAA>CTA mutations at codon 61 in the *Hras* gene. In stark contrast, ethylnitrosourea (ENU) treatment resulted in 100% mutations in the *Braf* gene, which can be uniquely activated by altering Val637 (equivalent to human Val600) from GTG>GAG. ENU is a potent mutagen that can cause A:T>T:A mutations in mutagenesis assays[149]. Although this mutation is the same as that induced by DMBA, ENU acts through binding to T residues

rather than the A residues that bind DMBA. This specificity of specific *Hras* vs *Braf* mutations can be explained by the fact that ENU forms thymine adducts on the coding strand of Val637, a mutation that cannot be efficiently induced by DMBA as the target adenosine is on the "wrong" strand. Such strand bias is seen in mutation signatures genome-wide[46], but would appear to be particularly important for strongly selected cancer driver genes such as *Ras*, *Braf*, and *Trp53*. Although in the liver model system DMBA has also been shown to cause the same V637E mutation in *Braf*, this is much less frequent than the standard activating CAA>CTA at codon 61 of *Hras*[150]. These data may well be due to tissue specific variation in transcription coupled repair, but this is clearly an area worthy of further investigation.

Initiated cell selection by tumor promoting agents

The term "tumor promotion" was first operationally described in the mid 20th century[151,152], and various aspects of this process have been reviewed several times over the past 60 years[135,153,154]. The consensus view, through the study of agents that promote skin tumors, is that tumor promoters function by non-mutational mechanisms to stimulate the signaling pathways downstream from Protein Kinase C (*Pkc*), which acts as the major receptor for the best known tumor promoting agent, TPA[155,156]. A major consequence of pathway engagement is induction of proliferation, inflammation, and outgrowth of initiated cells, but the precise mechanisms by which these initiated cells are selected remains obscure. Agents exist that efficiently induce both proliferation and inflammation, but nevertheless are very inefficient tumor promoters[157]. Further, distinct promoters exhibit specific selection for cells carrying particular oncogenic mutations. While the combination of DMBA, MNU or MNNG as initiator followed by promotion with TPA in skin tumor induction strongly selects for tumors carrying *Hras* mutations[137], treatment with DMBA followed by an alternative tumor promoter (mezerein)[158] or by over-expression of ornithine decarboxylase (*Odc*)[159] instead leads to a higher frequency of *Kras* mutant skin tumors. Some of these differences may be due to selective activation or

inhibition of various Pkc isoforms, which could result in either positive or negative effects on selection of *Ras* mutant tumors[160]. Some molecules structurally related to TPA have been described that can act as Pkc agonists, but not only lack tumor promoting activity[161,162], but can specifically inhibit the growth of tumors carrying activated *Kras* (but not *Hras*) oncogenes by interfering with non-canonical *Wnt* signaling[163].

In other mouse carcinogenesis model systems such as the liver, promoting agents can select one pathway while simultaneously inhibiting another. Phenobarbital (PB) is an antiepileptic drug that can promote liver tumor development when administered subsequent to treatment with a mutagenic initiator, most commonly diethylnitrosamine (DEN). In contrast to tumors induced only by DEN exposure, which have a high frequency of *Hras* codon 61 mutations, the PB treatment selects for tumors that lack these alterations, but instead harbor mutations in exon 2 of beta-catenin[164]. Since the DEN treatment precedes promotion with PB, cells carrying *Hras* mutations must be present in the liver tissue, but are not promoted by treatment with PB. Clearly, much still needs to be learned about the mechanisms by which promoters act in stimulating or inhibiting the selection of oncogenic pathways.

Tissue-specific selection of oncogenic Ras mutations.

An early conclusion reached on the basis of chemical carcinogenesis models of cancer was that mutations in the various members of the Ras family of oncogenes were exquisitely tissue-specific as well as carcinogen-specific. While *Hras* codon 61 mutations were consistently found in DMBA-initiated skin tumors[39], lung tumors caused by systemic injection of carcinogens often carried mutations in *Kras* rather than *Hras*[140]. This was the case even in skin and lung tumors from the same animals[165]. Similar conclusions were reached by analysis of a range of tumors induced in rats by transplacental exposure to MNU[166]. All tumor types had activating G>A transitions, but in brain tumors these were in the *Neu* oncogene, while mammary tumors and kidney mesenchymal tumors had *Hras* and *Kras* mutations respectively. Notably, Ras

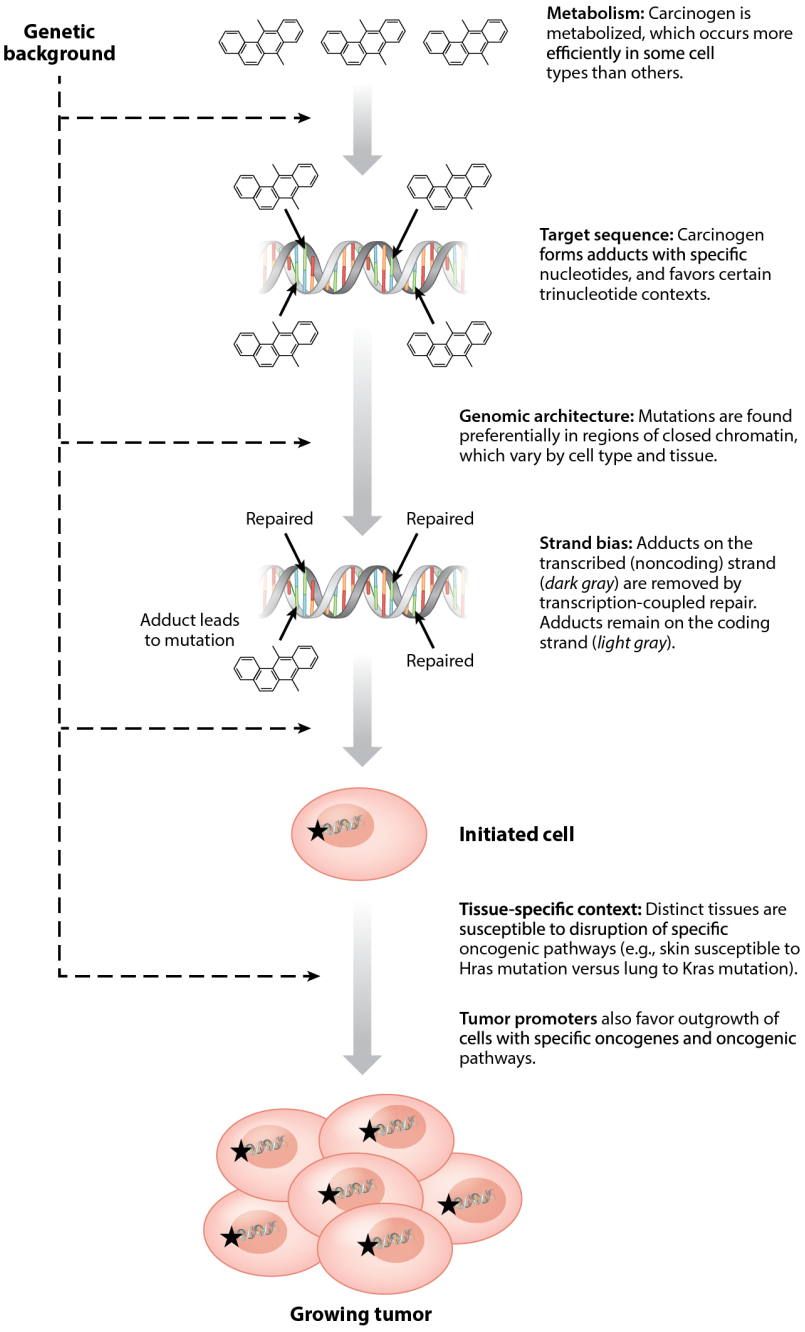


Figure 1. Factors influencing the selection of chemically initiated cells

mutations in human tumors follow a similar pattern. Lung adenocarcinomas have a high incidence (36%) of *KRAS* mutations, while squamous carcinomas of the skin, head and neck and lung, although they have a low overall RAS mutation frequency, tend to have a relatively higher proportion of *HRAS* mutations[35,36,57]. Interestingly, skin carcinomas are increased in

frequency in human melanoma patients who are treated with the *BRAF* inhibitor vemurafinib, and around 60% of these harbor the *Hras* Codon 61 CAA>CTA mutation that is seen in mouse tumors initiated with DMBA[167]. The short latency for development of these tumors suggests that they pre-exist in normal skin in these patients, and are rapidly promoted by vemurafinib treatment. This particular mutation would therefore appear to be under strong positive selection in squamous epithelia, for reasons that are still not clear. The reader is referred to other more comprehensive reviews of the possible mechanisms that underlie this selection induced by a well known cancer drug[168,169].

A trivial explanation of tissue specific *Hras* and *Kras* mutations would be that *Hras* is expressed in squamous epithelial tissues and *Kras* in the simple epithelia of the lung or gastrointestinal tract. This explanation is however untenable, as certain promoting agents such as mezerein or overexpression of *Odc* can promote growth of cancers with *Kras* mutations from normal mouse skin. *Kras* mutations are also seen in a high percentage of skin tumors that arise in *Hras* knockout mice[27,37]. Similarly, *Hras* when inserted into the *Kras* locus can lead to rapidly growing *Hras* mutant lung tumors after treatment with a lung carcinogen such as urethane[170] showing that *Kras*, although mutated in almost all chemically induced lung tumors, is not in fact essential for lung tumor development.

While both *Hras* and *Kras* mutations can lead to transformation of epithelial cells in skin and lung, analysis of skin tumors carrying these mutations highlights some important differences. Skin papillomas carrying *Kras* mutations are less frequent but arise earlier and are more likely to progress to carcinomas than corresponding lesions with *Hras* mutations[158]. *Kras* mutant tumors in *Hras* knockout mice are also reduced in frequency, but more likely to give rise to distant metastases[27]. The reasons for these differences remain obscure, but one model proposed is that *Kras* and *Hras* are expressed in different stem and progenitor cell populations within the skin[171]. In this model, mutation of the endogenous genes gives rise to phenotypes that reflect the cell of origin in addition to any biochemical differences in signaling activated by these

different oncogenes. This is compatible with results of transgenic experiments in which mutant *Hras* was artificially directed to stem-like and differentiated cell populations in mouse epidermis[105], resulting in development of benign lesions with distinctly different propensities for progression to carcinomas depending on the cell targeted. It is likely that both cell of origin and biochemical properties underlie these *in vivo* results, but this is an important area that will require more detailed investigation.

Influence of genetic background on mutation selection.

The selection processes present due to genomic architecture, target tissue, and choice of promoter are further refined by naturally occurring polymorphisms in different mouse strains. Schwarz and colleagues have shown that while DEN-induced liver tumors in susceptible C3H mice can have either *Hras* codon 61 or *Braf* codon 637 mutations, both of which lead to activation of the Mapk pathway, the relative frequency of *Braf* mutations is significantly higher in C57BL/6 mice that are more resistant to this tumor induction protocol[150]. These results are compatible with other data showing the pleiotropic effects of strain background on multiple factors that influence tumor development in mouse models, ranging from effects on stem cell selection[172,173], carcinogen metabolism[174], inflammation[175], as well as allele-specific somatic mutations or copy number changes[28,42,43,176]. Genetic background can also influence the specific mutation acquired in a given oncogene. Over 90% of urethane-induced lung tumors from wild-type mice carry a Q61R *Kras* mutation, but this switches to Q61L in *Kras* heterozygous ($Kras^{+/-}$) mice[64]. The reason for this switch is unknown, but implies that the balance between wild type and mutant *Kras* introduces novel facets of *Ras* biology that influence selection of specific mutants.

What have we learned from sequencing mouse tumor models?

The development of cost-effective next generation sequencing approaches has ushered in a new era of research on chemical carcinogenesis models. For the first time we are now able to study not only the types of mutations induced in single driver genes by specific carcinogenic agents, but can identify the genome-wide mutation spectra, as well as the patterns of gene mutations at different stages of tumorigenesis. Whole exome sequence analysis of mutations in chemically induced lung tumors[64] demonstrated that the genome-wide mutation signature exactly replicated the expected carcinogen signatures. MNU produces G>A transition mutations, and this was the predominant mutation seen across the genome in MNU-induced tumors. Tumors induced by urethane, on the other hand, exhibited genome-wide A>T or A>G mutations in a wide range of target genes, including many known cancer driver genes. In contrast with chemical models, tumors produced by a GEM model—carrying the same primary driver oncogene mutation and on the same genetic background—showed a very different genetic architecture. Two GEM models of small cell or non small cell lung cancer both gave rise to tumors with very few point mutations[64,65]. In these models, the sequential acquisition of gross chromosomal events leading to gene copy number alterations would appear to be the major cause of initiated cell selection and tumor progression[64,65,177]. This comparison highlights the importance of making the right choice of mouse cancer model to address the research question being addressed, whether it be the effect of targeted drugs, mechanisms of drug resistance, or responses to immunotherapy.

The skin model offers some specific advantages for analysis of the sequential events that influence multiple stages of tumor progression, because of the availability of lesions representing benign, locally invasive, and metastatic stages. In agreement with the mutation pattern seen previously in the initiating *Hras* gene, the genome-wide mutation spectrum in all tumors initiated by a single DMBA treatment consisted primarily of A>T mutations in a range of potential cancer driver genes[28,70]. Exome sequencing of papillomas, carcinomas, and metastases showed an average of 172 mutations in papillomas compared to 284 mutations in

carcinomas and 250 in metastases[28]. Whether the reduced number of mutations in papillomas compared to more progressed lesions is significant biologically remains to be determined. All of the A>T mutations were induced at the same time for each tumor, and all papillomas were harvested together with carcinomas and metastases for each mouse. A larger mutation burden may therefore facilitate tumor progression to carcinomas, and papillomas with high mutation load may therefore be under-represented in this cohort. Alternatively, early stage papillomas with high mutation load may be recognized by immune surveillance and removed, leaving only progressed lesions or papillomas that have escaped immune recognition. Interestingly, recent analysis of mutation burden in matched melanocytic nevi and melanomas from patients also found a lower mutation burden in the nevi, suggesting that both in mouse and human systems, early lesions, at least in terms of histology, exhibit fewer point mutations[13].

Mutations in melanomas and most other human tumors are acquired through chronic exposure to carcinogens (e.g. UV light or cigarette smoking), making it difficult to determine the timing of mutational events just based on signatures. In chemical carcinogenesis models it is however possible to clearly distinguish early from late mutations. In the case of a single initiation with DMBA or MNU, the resulting carcinogen-specific mutations are fixed within a few days of carcinogen exposure, and their timing can be precisely determined. Application of this approach to skin tumor mutation profiles clearly identified patterns of late mutations that were distinct from those initiated by the carcinogen DMBA. In contrast to the signature A>T mutations induced by DMBA, subclonal mutations that were acquired during growth and metastasis had a G>T mutation signature which we have presently attributed to oxidative stress[28]. Further analysis of these data may help to determine whether likelihood of progression is encoded in the DNA mutation profile immediately after initiation, or is more likely to be acquired through random acquisition and selection of G>T mutations during growth.

The large number of mutations present in chemically-induced tumors, and the acquisition of subsequent subclonal mutations over time, also make chemical models of cancer

well-suited to studying primary tumor heterogeneity and subclonal behavior, as well as evolution of metastases. Whether metastases arise in a linear fashion by seeding lymph nodes, followed by spread to distant sites, or disseminate in parallel from the primary site, has been debated for many years[25,178]. Based on the DMBA/TPA model, matching of metastases to their primary tumors using hundreds of specific mutations has shown that metastases generally disseminate from the matched primary tumor in parallel, spreading to all organs at approximately the same time rather than traveling "linearly" via a regional lymph node[28].

Gene copy number alterations are also observed in skin tumors from the DMBA/TPA model, and these increase quite dramatically in more advanced tumors. Chromosome 7 can be amplified during early papilloma development, and at that time is the only copy number alteration observed, suggesting it is the first copy number event to occur. Older tumors will also often have multiple whole chromosome gains or losses, which commonly include gain of chromosome 6. It is likely that these copy number changes reflect increased signaling through the Ras/MapK pathway, as several key genes are located on these chromosomes (*Hras* on chromosome 7, and *Kras*, *Braf* and *Raf1* on chromosome 6). About 50% of *Hras* mutant SCCs amplify chromosome 1, and such amplification is never seen in papillomas or in spindle carcinomas, or in SCCs with *Kras* mutations. This may therefore represent a genetic event that is both stage specific and stimulated by signaling through *Hras* but not *Kras*. Other copy number changes seen during progression include deletions of *Cdkn2a*, and copy number gains of *c-Met*, both of which are seen predominantly in the most advanced tumors[27,28]. These changes also replicate some of the most common genomic events seen in human squamous cancers of the lung and head and neck [35,57].

Other chemical models of cancer

Numerous chemical models exist in addition the common ones described above, with varying amounts of experimental work done to characterize them (**Table 1**

). In some cases, the mouse has proven to be refractory to development of certain tumor types after chemical exposure, while other species such as the rat or hamster, have provided more tractable models. For example, pancreatic cancer is one of the major causes of cancer-related morbidity in humans, but mice treated systemically with mutagens rarely develop pancreatic ductal adenocarcinomas. The Syrian golden hamster is however susceptible to development of pancreatic cancers after treatment with N-nitrosobis(2-oxopropyl)amine (BOP), and these share several genetic features with the equivalent human lesions, including mutations in *Kras* and *Cdkn2a*[179]. The National Toxicology Program has carried out extensive studies of both rats and mice exposed to a range of known and suspected human carcinogens, revealing both tissue-specific and species-specific effects of many of these agents (see Hoenerhoff et al. 2009 for review[180]). These models, as well as others that have led to the identification of important food-borne carcinogens[181,182] will provide an invaluable resource of tumors for molecular studies aimed at defining the range of chemicals that make a significant impact on mutational signatures in human cancers.

The future of chemical carcinogenesis

Mouse models of cancer based on tumor induction by chemical agents were, for the most part, supplanted in the 1980s with more precise genetically based models that allowed manipulation of the mouse germline to induce, in a spatial and temporally defined manner, specific types of genetic alterations found in human cancers. These GEMMs have proven themselves repeatedly to be a font of information that can reveal previously unknown facets of tumor biology[183–186]. Furthermore, GEMMs have been incorporated into “co-clinical” trials of new targeted drugs, as they allow assessment of the on-target effects of novel drugs designed to inhibit specific signaling pathways[187]. Such models will continue in the future to be important tools for discovery and validation of cancer drugs, and their combinations, for treatment of human cancer.

However, chemical carcinogenesis models have specific advantages for addressing certain important but unanswered questions in cancer biology. The first and most obvious is the question of how different tissues and cell types interact with environmental agents linked to human cancer. The environment has a huge impact on cancer development in humans, and the impact of some of these agents can be seen in the mutation signatures that have been identified by whole genome sequencing of thousands of human cancers[46]. The vast majority of mutagen signatures have no obvious causal agent, and sequence analysis of tumors in mice or rats that were induced by these chemicals may help to provide this causal link.

Another important question relates to the cell of origin of cancer within different tissues. Different target cells within the same tissue could give rise to tumors of different histological subtypes, or lesions of the same subtype that have different propensities for malignant progression. Both of these issues have been addressed largely using GEM models in which oncogenic mutations are targeted to different cell types within a tissue using specific gene promoters. Skin tumors with high risk of malignant progression were found to be located within the hair follicle bulge region rather than the more differentiated interfollicular cell population[105], while in the intestine, adenomas were shown to arise from Lgr5-positive stem cells rather than the committed transit amplifying cell population[188]. One caveat with these and other similar studies[189,190] is that the target cell population is predetermined by the investigator through the choice of gene promoter used to activate the appropriate oncogenic stimulus *in vivo*. While these approaches demonstrate that some cells within a population *can* give rise to tumors, whether they do so under conditions of environmentally induced cancer formation remains an open question. An alternative approach using a combination of carcinogen models with GEMMs in which target cell compartments are neutrally labeled with a reporter gene offer the potential to reveal which cells are in fact the true cell of origin when an entire tissue is exposed to a carcinogen[94,191].

Finally, tumor heterogeneity has become a major focus for studies of cancer therapy using both targeted approaches and more general chemo- or immuno-therapy. Heterogeneity can arise at both the cellular level, through participation of multiple distinct cells in tumor development, or at the genetic level through emergence of sublones carrying particular mutations. The latter can contribute to development of drug resistance after therapy, and the level of genetic heterogeneity itself is an indicator of poor prognosis in patient samples[17]. Chemical carcinogenesis models replicate these features of human tumors more accurately than GEM models as they have a high frequency of point mutations and substantial intratumoral heterogeneity. As such, they should be valuable tools for analysis of responses to both targeted drugs and chemotherapy, in which resistance is often driven by pre-existing subclones within tumors. The enormous impact of immune checkpoint blockade inhibitors in human clinical trials highlights the requirement for immunocompetent mouse models that can be exploited for preclinical testing of the many possible combinations of new drugs that might enhance the still relatively poor response rates in cancer patients. Checkpoint blockade therapies have been most effective in patients with tumors carrying high mutation loads[31] which create neoantigens recognizable to the immune system. These findings suggest that the high mutation load present in chemically induced mouse tumors will make these models particularly suitable for studies of immunotherapy. Notably, the concept of immunoediting[48,69] and development of immune checkpoint blockade inhibitors[192,193] have both already exploited the availability of chemically induced mouse tumor cell lines carrying abundant point mutations that stimulate antigenicity. Further development of *in vivo* chemical carcinogenesis models of disseminated disease may provide the stringent tests required to identify successful combinations of targeted, chemo-, and immuno-therapy drugs in preclinical trials.

Cancer Type	Carcinogenesis protocol	Primary Mutation Signature	Associated oncogenes	Tumors formed	Select References
Skin	DMBA (alternatively, MNU or MCA, single dose) + repeated tumor promoter (e.g., TPA).	DMBA: T>A MNU: G>A MCA: G>T	*Hras*; *Kras* in *Hras*$^{-/-}$ mice	Premalignant papillomas & malignant carcinomas (of either squamous or spindle morphology). Surgical resection of the primary tumor leads to metastasis to lymph nodes, lung, and other sites.	[26–28,70,135,158]
Liver	DEN (repeat doses, or in combination with PB promotion)	G>A, T>C	*Hras* (DEN only); *Ctnnb1* (DEN/PB)	Premalignant hyperplastic foci & malignant HCCs. Micrometastases can be observed in lung, particularly with lower doses of DEN in rats.	[164,194–197]
Lung	Urethane or MNU (single i.p. injection)	Urethane: A>T, A>G, and G>A MNU: G>A	*Kras*	Premalignant adenomas & malignant adenocarcinomas.	[64,140,198]
Sarcoma	MCA (injection into mouse leg)	G>T	*Kras* or *Nras*	Fibrosarcoma.	[193,199–202]
Urothelial	BBN (oral gavage or in drinking water)	G>A	*p53*; less commonly, *Hras*	Urinary bladder carcinoma.	[203–205]
Colon	PhIP or MeIQx; plus DSS promoter	PhIP: G>T and G>A; -1G deletion		Adenocarcinoma.	[206]

Cancer Type	Carcinogenesis protocol	Primary Mutation Signature	Associated oncogenes	Tumors formed	Select References
Prostate & Colon	PhIP (rat)	G>T and G>A; -1G deletion	Colon: *Ctnnb1* & *Apc*	Colon adenocarcinoma and prostate carcinoma.	[182]
Breast	DMBA or MNU (rat)	DMBA: A>T MNU: G>A	*Hras*	Multiple breast carcinoma subtypes.	[138,207–209]
Pancreas	BOP (Syrian golden hamster)	G>A	*Kras*	Ductal adenocarcinoma.	[210,211]

Table 1. Selection of common mouse and other rodent models of cancer. Chemical name abbreviations: Dimethylbenzanthracene (DMBA); 12-O-tetradecanoylphorbol-13-acetate (TPA); methyl-nitrosourea (MNU); methylcholanthrene (MCA); N-nitrosodiethylamine (DEN); phenobarbital (PB); N-butyl-N-(4-hydroxybutyl)nitrosamine (BBN); phenylimidazopyridine (PhIP), 2-aminodimethylimidazoquinoxalin (MeIQx), dextran sodium sulfate (DSS), N-Nitrosobis(2-oxopropyl)amine (BOP)

Chapter 4

Responsiveness of chemically-induced tumors to immunotherapy

A great deal of excitement has been generated in recent years about the potential to use the immune system to fight tumors. At the forefront of immunotherapy approaches in the clinic have been checkpoint blockades, α-CTLA-4 and α-PD-1, which function by blocking immune inhibitory receptors and "taking the brakes off" exhausted T cells to improve the anti-tumor immune response. But despite curing metastatic disease in a small percentage otherwise-untreatable patients, checkpoint blockades still yield no clinical benefit to most patients. It is therefore critical to identify models in which we can better study tumor-immune interactions and develop strategies to both predict and improve tumor responses to checkpoint blockades as well as other emerging immunotherapies.

Not surprisingly, considerable investigation has taken place into the factors that correlate with patient responses to immunotherapy. In a study of α-PD-1 therapy (pembrolizumab) in lung cancer patients, it was observed that tumors with a high mutation burden generally responded better to treatment. The explanation for this is believed to be that nonsynonymous mutations, which result in mutant proteins, can lead to the expression of tumor-specific neoantigens that T cells can recognize as "non-self" and use to identify and kill tumor cells. Thus having a high mutation burden increases the number of candidate neoantigens in the tumor, and therefore the likelihood of the tumor being recognized by T cells. In some cases, it has even been possible to pinpoint specific mutation-induced neoantigens that successfully elicit an immune response against the tumor[212].

We hypothesized that chemically-induced skin tumors, which we have shown carry a relatively high mutation burden on par with human lung tumors[28], might therefore be suitable targets for checkpoint blockade therapy. To test this, we made use of several tumor cell lines derived from DMBA-induced skin carcinomas. For comparison, we also used tumor cell lines

derived from a genetically engineered mouse (GEM) tumor model (*Kras*$^{LSL\text{-}G12D}$), which we expected would carry a low mutation burden similar to other GEM models [64,213]. Exome sequencing of these tumor cells lines revealed that the chemically induced cell lines carried a range of intermediate to high mutation burdens (**Fig. 1a**). The highest mutation burden line, CCK168, carried 945 mutations, 297 of which were nonsynonymous and predicted to result in a change in protein sequence. Cell lines from GEM tumors by contrast carried fewer than 40 nonsynonymous mutations (**Fig. 1a**).

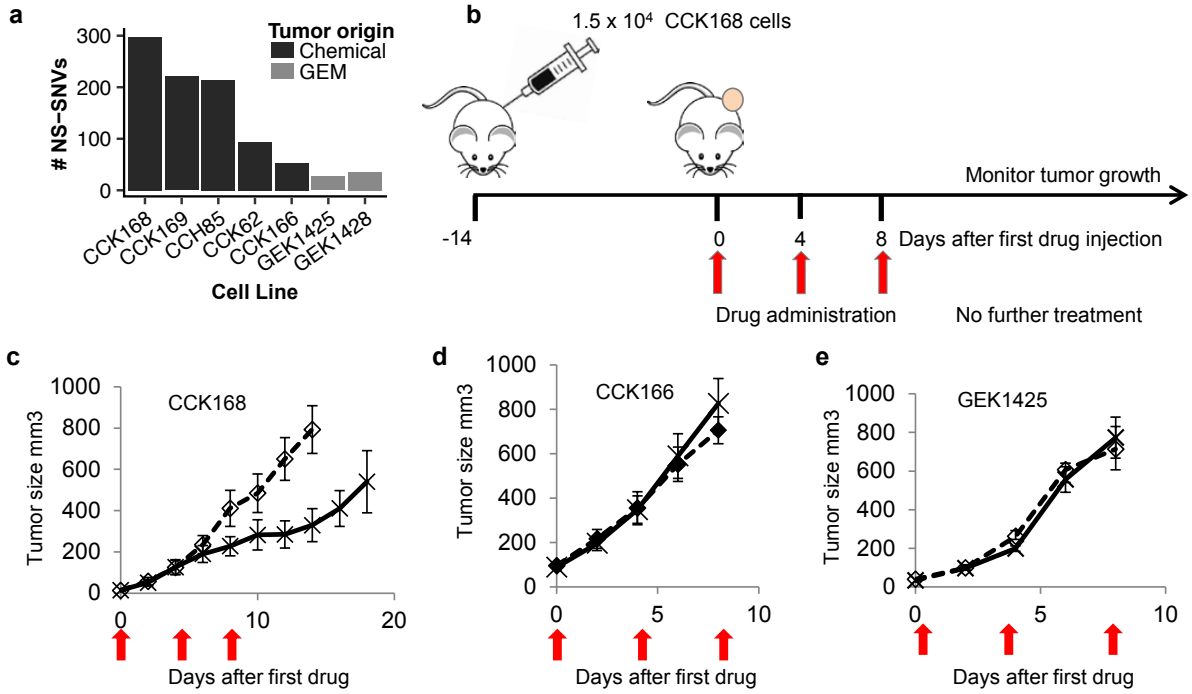


Figure 1. Effects of α-PD-1 on chemically induced and GEM derived tumors. (**a**) Total nonsynonymous mutation burdens of five chemically-induced and two GEM-derived tumors, as determined by whole exome sequencing. (**b**) Scheme for tumor implantation and drug treatment protocol. After implantation of 1.5×10^4 tumor cells by dorsolateral subcutaneous injection, tumors grew for 14 days until they reached ≥5mm diameter. Mice were then treated by intraperitoneal injection of drug into the contralateral side, with three drug administrations, each four days apart (considered days 0, 4 and 8). (**c-e**) Tumor growth with and without α-PD-1 therapy was measured at least every other day from the time of first drug administration for (**c**) CCK168, (**d**) CCK166, and (**e**) GEK1425 in vivo. Red arrows indicate timing of drug administration.

Mice were injected subcutaneously with each of these tumor cell lines, and after 2 weeks when tumors had reached approximately 5mm in diameter, mice were subsequently treated with three doses of α-PD-1, each four days apart (**Fig. 1b**). We found the mice bearing tumors

derived from the two cell lines carrying the highest mutation burdens, CCK168 and CCK169, showed partial responses to α-PD-1 treatment (CCK168 **Fig. 1c**; CCK169 not shown), while mice with tumors derived from the other cell lines did not. This was true for both the other chemically induced lines (**Fig. 1d**), which carried intermediate mutation burdens, and the GEM lines (**Fig. 1e**), which carried very low mutation burdens. Interestingly, even though the mutation burden in CCH85 was nearly as high as that in CCK169 (**Fig. 1a**, 214 vs. 221 nonsynonymous mutations, respectively), this line was entirely unresponsive to α-PD-1 treatment (not shown).

The unresponsiveness of CCH85 in contrast to CCK169 suggests that while mutation burden is a contributing factor toward sensitivity, other factors are relevant in this model as well. This mirrors observations in patients—a high mutation burden correlates to a good response, but is not sufficient to be predictive[31]. The level of immune infiltrate in the tumor is known to play a role—although it is not presently known whether immune infiltrate levels are influenced by the antigenicity or mutation burden of the tumor—and other, yet undiscovered factors are believed to be present. Further, evidence has accumulated from both patients and mouse models suggests that a single or small number of neoantigens may drive tumor rejection following immunotherapy[212,214], thus painting a picture in which a high number of total mutations increases the likelihood of a strong neoantigen being present, but by no means guarantees it.

A number of mice bearing CCK168 tumors exhibited partial responses to α-PD-1 therapy, showing initial tumor regression followed by resistance and outgrowth. In four such mice, the outgrowing tumors were harvested after the onset of resistance. These outgrowth lines were found to be capable of giving rise to tumors in naïve mice, and after passage through two generations of mice, were subsequently collected for sequencing. We speculated that resistance to α-PD-1 might be associated with the outgrowth of a resistant subclone of CCK168, perhaps carrying a lower mutation burden or lacking a particular mutation responsible for a strong neoantigen. To our surprise, however, all four outgrowth lines had substantially more mutations than the parent CCK168 line (**Fig. 2a**). Analysis of these mutations revealed that all

four outgrowth lines had retained nearly all the parental CCK168 mutations, with no consensus between the outgrowth lines as to which mutations were excluded. Further, these lines had accumulated between 650 and 857 new mutations, which were predominantly G>T (C>A) mutations (**Fig. 2b**). We do not presently know whether the majority of these G>T mutations were accumulated during the development of resistance, or subsequently during passage through naïve mice. We have previously seen that G>T mutations are accumulated in primary DMBA/TPA-induced tumors during the course of tumor development, possibly due to the presence of reactive oxygen species[28] (also see Chapter 2), making it likely that at least some of these mutations were acquired during passage. To resolve this in future experiments, resistant tumors will need to be sequenced immediately upon outgrowth. Nonetheless, the presence of such an exceptionally high mutation burden in tumors that had acquired α-PD-1 resistance points to a mechanism of resistance that does not rely on neoantigen silencing and is itself of interest for future investigation.

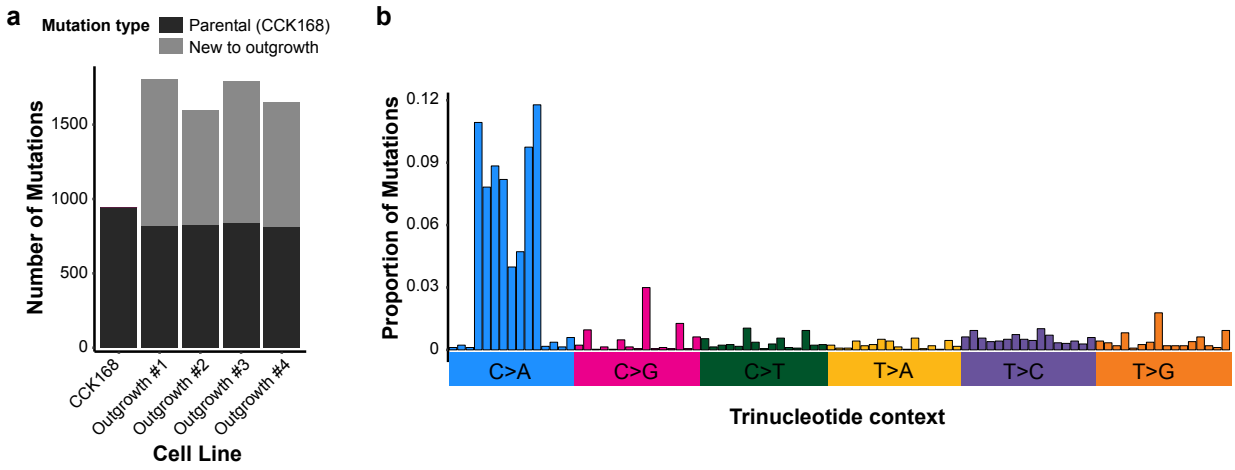


Figure 2. Mutations in CCK168-derived cell lines with acquired resistance to α-PD-1. (**a**) Number of mutations in the parent cell line, CCK168, as compared to four outgrowth lines as identified through whole exome sequencing. Mutations shared with the parent line are shown in dark grey, while mutations acquired in the resistant lines are in light grey. (**b**) Trinucleotide context of non-parental mutations in the outgrowth lines (i.e., light grey mutations in panel (**a**)), revealing that the new mutations were predominantly G>T (C>A). X-axis denotes 96 possible trinucleotide contexts, grouped by (labeled) base pair change.

We have seen that cell lines derived from chemically induced tumors recapitulate key features of human tumors in their responses to immunotherapy, in particular, demonstrating a

correlation between responsiveness to therapy and initial tumor mutation burden. Additional factors, however, are clearly present, as we observed that cell lines with nearly-identical mutation loads could vary in their ability to respond to checkpoint blockades, and that even in the most responsive cell line, CCK168, individual responses to α-PD-1 varied animal to animal. The DMBA/TPA skin model, and cell lines derived from these carcinomas, therefore provide a useful platform for dissecting the anti-tumor immune response and for identifying novel factors that can influence or predict responsiveness to checkpoint blockades, as well as for studying mechanisms of resistance.

ACKNOWLEDGMENTS

This vignette is the result of a collaboration with Drs. Rosemary Akhurst, Eswari Dodagatta, and Dominique Meyer. The mouse experiments in this chapter, including treatment with α-PD-1, were completed by E.D. and D.M., while my (M.Q.M.'s) contributions were in analysis of the sequencing data.

METHODS

Mouse experiments. FVB/NJ mice (6-week old females) purchased from the JAX labs and acclimatized to the UCSF Laboratory Animal Research Center facility for one week, were used for experiments. All animal procedures adhered to NIH Guidelines for the Care and Use of Laboratory Animals and undertaken under authorization of the UCSF Institutional Animal Care and Use Committee in an AAALAC approved facility.

Generation of tumor cell lines. Chemically-induced tumor cell lines were generated on wild type or *Hras*$^{-/-}$ mice[215] of the FVB/NJ strain using a standard DMBA/TPA protocol[27]. Carcinomas appear between 25-45 weeks following a single topical treatment with DMBA (25μg) in acetone applied to previously shaved dorsal skin) followed by twenty weeks of biweekly TPA application

(200μl per treatment of a 10^{-4} M solution in acetone). GEM-derived tumor cell lines were generated by intercrossing FVB mice carrying a *Kras*$^{LSL\text{-}G12D}$ allele[216] and *Lgr5-CreER*[217]. Activation of Cre within Lgr5-expressing keratinocytes was induced by a single topical application of 4-hydroxytamoxifen (200μl of 25 mg/ml in 100% ethanol), causing replacement of *Kras* with an activated *KrasG12D* allele by Cre-induced recombination of *Kras*$^{LSL\text{-}G12D}$. Papillomas arose 6-8 weeks after induction of a full thickness wound. Papillomas >5mm were excised and allowed to grow back, after which they converted to squamous cell carcinomas. GEM-derived carcinomas developed 12-15 weeks following papilloma resection.

Primary carcinomas were removed from freshly euthanized mice and, under sterile conditions, chopped vigorously into small pieces before plating into cell culture with DMEM supplemented with 10% fetal calf serum, non-essential amino acids and penicillin/streptomycin. Attached cells were expanded for 3-4 weeks before stocks were frozen down, and all experiments were undertaken on cells at ≤ four passages.

***In vivo* experiments.** After expansion in DMEM plus 10% FCS, 1.5 x 10^4 tumor cells were injected subcutaneously and unilaterally into the dorsal flank of 7- to 8-week-old mice. 14 days later, when tumors measured approximately 5mm diameter, mice were injected i.p. with three doses of therapeutic antibodies, administered at four day intervals. Antibodies used were rat α-PD-1 IgG2a antibody (RMP1-14, BioXCell) or its rat isotype control (2A3) each at 250μg per dose.

DNA extraction & whole exome sequencing. Cell line samples were submitted to Otogenetics Corporation (Atlanta, GA USA) for mouse exome capture and sequencing. Isolation of genomic DNA (gDNA), library preparation for sequencing, and exome sequencing were performed by Otogenetics. gDNA was isolated from tumor cell lines as well as from two strain-matched mouse

tails using DNeasy Blood & Tissue kit (Qiagen #69506). Briefly, gDNA was subjected to agarose gel and OD ratio tests to confirm the purity and concentration prior to Bioruptor (Diagenode, Inc., Denville, NJ USA) fragmentation. Fragmented gDNAs were tested for size distribution and concentration using an Agilent Tapestation 2200 and Nanodrop. Illumina libraries were made from qualified fragmented gDNA using SPRIworks HT Reagent Kit (Beckman Coulter, Inc. Indianapolis, IN USA, catalog# B06938) and the resulting libraries were subjected to exome enrichment using SureSelectXT Mouse All Exon (Agilent Technologies, Wilmington, DE USA, catalog# 5190-4641) following manufacturer's instructions. Enriched libraries were tested for enrichment by qPCR and for size distribution and concentration by an Agilent Bioanalyzer 2100. The samples were then sequenced on an Illumina HiSeq2500 using Rapid v2 SBS chemistry which generated paired-end reads of 106 nucleotides (nt).

Sequence alignment, processing and quality control. Reads were mapped to the GRCm38/mm10 version of the *Mus musculus* genome using BWA (version 0.7.12)[75] with default parameters. The Picard MarkDuplicated module was used to remove duplicates from the data (version 1.119; **http://broadinstitute.github.io/picard**). The Genome Analysis Tool Kit (GATK-Lite) toolkit (version 2.3-9)[117] module IndelRealigner and BaseRecalibrator were used to preprocess the alignments. During base quality recalibration, dbSNP variants were used as known sites, according to GATK Best Practices recommendations[76]. Finally, alignment and coverage metrics were collected using Picard. We sequenced an average of 47 million unique reads per sample. Targeted bases were sequenced to a mean depth of 62, and more than 88% of targeted bases were sequenced to 20× coverage or greater.

Identification of SNVs and annotation. SNVs were called using somatic variant detection program MuTect (version 1.1.7)[77]. Each tumor was called against a tail from a strain-matched mouse. Calls were filtered against a database of known *Mus musculus* germline SNPs available

at ftp.ncbi.nih.gov/snp/organisms/mouse_10090/ VCF/genotype, as well as against both normal tails from this experiment. Results were further filtered to calls with a minimum read depth of 10 at the locus for both tumor and matched normal, and to calls where at least one alternate read had a mapping quality score of 60 or higher. Variants were annotated using Annovar (downloaded on 2/4/2016)[78], and these annotations were used as the basis for assessing exonic variants as synonymous, nonsynonymous, stopgain, or stoploss.

Concluding Thoughts

Tumors are incredibly complex entities, and this complexity has proved to be a significant roadblock to the treatment and cure of cancer. Not only do tumors display genetic and cellular heterogeneity, but this heterogeneity is fluid, with tumors evolving over time. By combining next generation sequencing and multi-color lineage tracing approaches, it has been possible to begin to dissect this heterogeneity in a chemical model of skin cancer, identifying common patterns that take place during the progression of an initiated cell to a metastatic disease.

We have learned that a benign tumor arises from a single initiated cell in this model, as is believed to be the case in most human tumors, but even in the case of this clonal initiation, the benign tumor can recruit additional populations of “normal” skin cells into the tumor. We see from this that the presence of multiple cellular populations in the tumor is not in itself evidence of polyclonal initiation, but that even monoclonally initiated benign tumors can exhibit both genetic and cellular heterogeneity from a very early timepoint.

This early heterogeneity, however, is displaced at the time of progression to malignancy, with the emergence of a dominant progressing clone. It is possible that this progressing clone has acquired a specific additional genetic event that enables progression, although the specific event that would facilitate this is presently unknown, and very well may vary tumor to tumor. Loss of *Trp53*, loss of *Cdkn2a*, and copy gain of chromosome 1 are potential candidates, and may be responsible for facilitating progression in at least a subset of skin tumors.

Following progression, the tumor landscape again continues to gain complexity, with multiple subclones proliferating, intermixing, and evolving independently. We saw that in some cases these clones occupy anatomically distinct regions, such that multi-region sampling of the tumor would reveal their divergence. In other cases however, such as the “speckled”

carcinomas we observed, separate clones are thoroughly intermixed, and their identification would necessitate single-cell resolution.

Finally, we have found evidence suggesting that metastasis is a fundamentally distinct process from initiation and progression. While initiation and progression were both observed to be clonal events, we observed that many cells in the primary tumor could gain metastatic capability, including cells from distinct cellular populations. This suggests that metastasis may be linked to a change in cellular state, perhaps from microenvironmental cues, rather than the acquisition of a particular mutation. We further found that cells metastasizing to distinct anatomical sites did not necessarily travel via a lymph node, but rather departed from the primary tumor more or less synchronously and spread in parallel to the lymph nodes and other distant sites, where they continued to evolve separately.

The complexity of chemically induced skin tumors recapitulates many aspects of the complexity of human tumors. These parallels make the chemically induced skin tumor model well suited for studying tumor evolution and heterogeneity, as well as for testing therapeutic strategies to combat heterogeneous tumors going forward. The responsiveness of high mutation burden DMBA-induced tumor cell lines to α-PD-1 further suggest it will be a suitable model for studying checkpoint blockades and other immunotherapy strategies.

Currently, there is limited understanding of the relationship between tumor heterogeneity and immunotherapy, but it is expected that heterogeneity will have an impact on all T cell based strategies, including checkpoint blockades. T cells rely on tumor neoantigens to identify tumor cells and kill them. Neoantigens, however, are the product of mutations, and so show the same intratumoral heterogeneity that mutations do. Having a model system in which both heterogeneity and immunotherapy can be studied will be a great advantage in investigating immunotherapy in the context of a heterogeneous tumor, as well as for studying the impact of heterogeneity on the anti-tumor immune response.

REFERENCES

1. Parada, L. F., Tabin, C. J., Shih, C. & Weinberg, R. A. Human EJ bladder carcinoma oncogene is homologue of Harvey sarcoma virus ras gene. *Nature* **297,** 474–8 (1982).
2. Taparowsky, E. *et al.* Activation of the T24 bladder carcinoma transforming gene is linked to a single amino acid change. *Nature* **300,** 762–5 (1982).
3. Varmus, H. How Tumor Virology Evolved into Cancer Biology and Transformed Oncology. *Annu. Rev. Cancer Biol.* **1,** 1–18 (2017).
4. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194,** 23–8 (1976).
5. Crawford, L. V *et al.* Characterization of the complex between SV40 large T antigen and the 53K host protein in transformed mouse cells. *Cold Spring Harb. Symp. Quant. Biol.* **44 Pt 1,** 179–187 (1980).
6. Nowell, P. C. & Hungerford, D. A. Chromosome Changes in Human Leukemia and a Tentative Assessment of Their Significance. *Ann. N. Y. Acad. Sci.* **113,** 654–62 (1964).
7. Gottlieb, S. K. Chromosomal abnormalities in certain human malignancies. A review. *J. Am. Med. Assoc.* **209,** 1063–66 (1969).
8. Knudson, A. G. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proc. Natl. Acad. Sci.* **68,** 820–823 (1971).
9. Burns, P. A. *et al.* Loss of heterozygosity and mutational alterations of the p53 gene in skin tumours from interspecific hybrid mice. *Oncogene* **6,** 2363–9 (1991).
10. Ruggeri, B. *et al.* Alterations of the p53 tumor suppressor gene during mouse skin tumor progression. *Cancer Res.* **51,** 6615–21 (1991).
11. 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).
12. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481,** 306–313 (2012).
13. Shain, A. H. *et al.* The Genetic Evolution of Melanoma from Precursor Lesions. *N. Engl.*

J. Med. **373,** 1926–1936 (2015).

14. Gerlinger, M. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* **366,** 883–92 (2012).
15. Sottoriva, A. *et al.* A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **47,** 209–216 (2015).
16. Alizadeh, A. A. *et al.* Toward understanding and exploiting tumor heterogeneity. *Nat. Med.* **21,** 846–853 (2015).
17. McGranahan, N. & Swanton, C. Biological and Therapeutic Impact of Intratumor Heterogeneity in Cancer Evolution. *Cancer Cell* **27,** 15–26 (2015).
18. Turke, A. B. *et al.* Pre-existence and clonal selection of MET amplification in EGFR mutant NSCLC. *Cancer Cell* **17,** 77–88 (2010).
19. Parsons, B. L. Many different tumor types have polyclonal tumor origin: Evidence and implications. *Mutat. Res.* **659,** 232–247 (2008).
20. Reddy, A. L. & Fialkow, P. J. Influence of dose of initiator on two-stage skin carcinogenesis in BALB/c mice with cellular mosaicism. *Carcinogenesis* **9,** 751–754 (1988).
21. Winton, D. J., Blount, M. A. & Ponder, B. A. Polyclonal origin of mouse skin papillomas. *Br. J. Cancer* **60,** 59–63 (1989).
22. Kaplan, R. N. *et al.* VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature* **438,** 820–7 (2005).
23. Harney, A. S. *et al.* Real-Time Imaging Reveals Local, Transient Vascular Permeability, and Tumor Cell Intravasation Stimulated by TIE2hi Macrophage-Derived VEGFA. *Cancer Discov.* **5,** 932–43 (2015).
24. Aceto, N. *et al.* Circulating Tumor Cell Clusters Are Oligoclonal Precursors of Breast Cancer Metastasis. *Cell* **158,** 1110–1122 (2014).
25. Klein, C. A. Parallel progression of tumour and metastases. *Nat. Rev. Cancer* **9,** 302–12

(2009).

26. Quintanilla, M., Brown, K., Ramsden, M. & Balmain, A. Carcinogen-specific mutation and amplification of Ha-ras during mouse skin carcinogenesis. *Nature* **322,** 78–80 (1986).

27. Wong, C. E. *et al.* Inflammation and Hras signaling control epithelial-mesenchymal transition during skin tumor progression. *Genes Dev.* **27,** 670–82 (2013).

28. McCreery, M. Q. *et al.* Evolution of metastasis revealed by mutational landscapes of chemically induced skin cancers. *Nat. Med.* **21,** 1514–1520 (2015).

29. Snippert, H. J. *et al.* Intestinal Crypt Homeostasis Results from Neutral Competition between Symmetrically Dividing Lgr5 Stem Cells. *Cell* **143,** 134–144 (2010).

30. McCreery, M. Q. & Balmain, A. Chemical Carcinogenesis Models of Cancer: Back to the Future. *Annu. Rev. Cancer Biol.* **1,** 285–312 (2017).

31. Rizvi, N. A. *et al.* Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer. *Science* **348,** 124–9 (2015).

32. Chang, K. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45,** 1113–1120 (2013).

33. Ciriello, G. *et al.* Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* **45,** 1127–1133 (2013).

34. Brabletz, T., Lyden, D., Steeg, P. S. & Werb, Z. Roadblocks to translational advances on metastasis research. *Nat. Med.* **19,** 1104–9 (2013).

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

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

37. Ise, K. *et al.* Targeted deletion of the H-ras gene decreases tumor formation in mouse skin carcinogenesis. *Oncogene* **19,** 2951–1956 (2000).

38. Bos, J. L. ras Oncogenes in Human Cancer: A Review. *Cancer Res.* **49,** 4682–4689

(1989).

39. Balmain, A. & Pragnell, I. Mouse skin carcinomas induced in vivo by chemical carcinogens have a transforming Harvey-ras oncogene. *Nature* **303,** 72–4 (1983).

40. Finch, J. S., Albino, H. E. & Bowden, G. T. Quantitation of early clonal expansion of two mutant 61st codon c-Ha-ras alleles in DMBA/TPA treated mouse skin by nested PCR/RFLP. *Carcinogenesis* **17,** 2551–2557 (1996).

41. Loehrke, H. *et al.* On the persistence of tumor initiation in two-stage carcinogenesis on mouse skin. *Carcinogenesis* **4,** 771–775 (1983).

42. Nagase, H., Mao, J. & Balmain, A. Allele-specific Hras Mutations and Genetic Alterations at Tumor Susceptibility Loci in Skin Carcinomas from Interspecific Hybrid Mice. *Cancer Res.* **63,** 4849–4853 (2003).

43. Chen, B., You, L., Wang, Y., D.Stoner, G. & You, M. Allele-specific activation and expression of the K-ras gene in hybrid mouse lung tumors induced by chemical carcinogens. *Carcinogenesis* **15,** 2031–2035 (1994).

44. Samuels, Y. *et al.* High Frequency of Mutations of the PIK3CA Gene in Human Cancers. *Science* **304,** 554 (2004).

45. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–8 (2013).

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

47. Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467,** 1109–13 (2010).

48. Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoediting: from immuno-surveillance to tumor escape. *Nat. Immunol.* **3,** 991–998 (2002).

49. Giancotti, F. G. Mechanisms Governing Metastatic Dormancy and Reactivation. *Cell* **155,** 750–764 (2013).

50. Lin, Z. *et al.* LMP1 regulates periodontal ligament progenitor cell proliferation and differentiation. *Bone* **47,** 55–64 (2010).

51. Sadej, R., Grudowska, A., Turczyk, L., Kordek, R. & Romanska, H. M. CD151 in cancer progression and metastasis: a complex scenario. *Lab. Invest.* **94,** 41–51 (2014).

52. Takeda, Y. *et al.* Diminished metastasis in tetraspanin CD151–knockout mice. *Blood* **118,** 464–472 (2011).

53. Quigley, D. A. *et al.* Network analysis of skin tumor progression identifies a rewired genetic architecture affecting inflammation and tumor susceptibility. *Genome Biol.* **12,** R5 (2011).

54. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2,** 401–4 (2012).

55. Gao, J. *et al.* Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* **6,** pl1 (2013).

56. Tsai, J. H., Donaher, J. L., Murphy, D. a, Chau, S. & Yang, J. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell* **22,** 725–36 (2012).

57. The Cancer Genome Atlas. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489,** 519–25 (2012).

58. Aldaz, C., Trono, D., Larcher, F., Slaga, T. & Conti, C. Sequential trisomization of chromosomes 6 and 7 in mouse skin premalignant lesions. *Mol. Carcinog.* **2,** 22–6 (1989).

59. Namiki, T. *et al.* AMP kinase-related kinase NUAK2 affects tumor growth, migration, and clinical outcome of human melanoma. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 6597–602 (2011).

60. The Cancer Genome Atlas. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511,** 543–550 (2014).

61. Linardopoulos, S. *et al.* Deletion and altered regulation of p16INK4a and p15INK4b in undifferentiated mouse skin tumors. *Cancer Res.* **55,** 5168–5172 (1995).

62. Huber, K. V. M. *et al.* Stereospecific targeting of MTH1 by (S)-crizotinib as an anticancer strategy. *Nature* **508,** 222–7 (2014).

63. Gad, H. *et al.* MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool. *Nature* **508,** 215–21 (2014).

64. Westcott, P. M. K. *et al.* The mutational landscapes of genetic and chemical models of Kras-driven lung cancer. *Nature* **517,** 489–492 (2014).

65. McFadden, D. G. *et al.* Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. *Cell* **156,** 1298–311 (2014).

66. Klein, C. A. Selection and adaptation during metastatic cancer progression. *Nature* **501,** 365–72 (2013).

67. Peinado, H., Lavotshkin, S. & Lyden, D. The secreted factors responsible for pre-metastatic niche formation: Old sayings and new thoughts. *Semin. Cancer Biol.* **21,** 139–146 (2011).

68. Cox, T. R. *et al.* The hypoxic cancer secretome induces pre-metastatic bone lesions through lysyl oxidase. *Nature* **522,** 106–10 (2015).

69. Mittal, D., Gubin, M. M., Schreiber, R. D. & Smyth, M. J. New insights into cancer immunoediting and its three component phases-elimination, equilibrium and escape. *Curr. Opin. Immunol.* **27,** 16–25 (2014).

70. Nassar, D., Latil, M., Boeckx, B., Lambrechts, D. & Blanpain, C. Genomic landscape of carcinogen-induced and genetically induced mouse skin squamous cell carcinoma. *Nat. Med.* **21,** (2015).

71. Engel, J., Emeny, R. T. & Hölzel, D. Positive lymph nodes do not metastasize. *Cancer Metastasis Rev.* **31,** 235–246 (2012).

72. Fisher, B. *et al.* Twenty-five-year follow-up of a randomized trial comparing radical

mastectomy, total mastectomy, and total mastectomy followed by irradiation. *N. Engl. J. Med.* **347,** 567–575 (2002).

73. Veronesi, U., Marubini, E., Mariani, L., Valagussa, P. & Zucali, R. The dissection of internal mammary nodes does not improve the survival of breast cancer patients. 30-Year results of a randomised trial. *Eur. J. Cancer* **35,** 1320–1325 (1999).
74. Varela, I. *et al.* Exome sequencing identifies frequent mutaiton of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* **469,** 539–542 (2011).
75. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–60 (2009).
76. DePristo, M. a *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–8 (2011).
77. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31,** 213–9 (2013).
78. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).
79. Forbes, S. a *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43,** 805–811 (2014).
80. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science* **339,** 1546–58 (2013).
81. Talevich, E., Shain, A. H. & Bastian, B. C. CNVkit: Copy number detection and viziualization for targeted sequencing using off-target reads. (2014). doi:10.1101/010876
82. Untergasser, A. *et al.* Primer3--new capabilities and interfaces. *Nucleic Acids Res.* **40,** e115 (2012).
83. Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23,** 1289–91 (2007).
84. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10 (2011).

85. Li, H. *et al.* The Sequence Alignment / Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

86. Roth, A. *et al.* PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11,** 396–8 (2014).

87. Carvalho, B. S. & Irizarry, R. a. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26,** 2363–7 (2010).

88. Quigley, D. Equalizer reduces SNP bias in Affymetrix microarrays. *BMC Bioinformatics* **16,** 238 (2015).

89. Calbo, J. *et al.* A Functional Role for Tumor Cell Heterogeneity in a Mouse Model of Small Cell Lung Cancer. *Cancer Cell* **19,** 244–56 (2011).

90. Cleary, A. S., Leonard, T. L., Gestl, S. a & Gunther, E. J. Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. *Nature* **508,** 113–7 (2014).

91. Jonason, A. S. *et al.* Frequent clones of p53-mutated keratinocytes in normal human skin. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 14025–14029 (1996).

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

93. Goruppi, S. & Dotto, G. P. Mesenchymal stroma: primary determinant and therapeutic target for epithelial cancer. *Trends Cell Biol.* **23,** 593–602 (2013).

94. Li, S. *et al.* A keratin 15 containing stem cell population from the hair follicle contributes to squamous papilloma development in the mouse. *Mol. Carcinog.* **52,** 751–9 (2013).

95. Driessens, G., Beck, B., Caauwe, A., Simons, B. D. & Blanpain, C. Defining the mode of tumour growth by clonal analysis. *Nature* **488,** 527–31 (2012).

96. Cheung, K. J., Gabrielson, E., Werb, Z. & Ewald, A. J. Collective invasion in breast cancer requires a conserved Basal epithelial program. *Cell* **155,** 1639–51 (2013).

97. Maddipati, R. & Stanger, B. Z. Pancreatic Cancer Metastases Harbor Evidence of

Polyclonality. (2015). doi:10.1158/2159-8290.CD-15-0120

98. Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520,** 353–7 (2015).

99. Sanborn, J. Z. *et al.* Phylogenetic analyses of melanoma reveal complex patterns of metastatic dissemination. *Proc. Natl. Acad. Sci.* **112,** 10995–11000 (2015).

100. Swanton, C. Intratumor Heterogeneity: Evolution through Space and Time. *Cancer Res.* **72,** 4875–4882 (2012).

101. Ito, M. *et al.* Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nat. Med.* **11,** 1351–1354 (2005).

102. Schepers, A. G. *et al.* Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. *Science* **337,** 730–5 (2012).

103. Bigger, C. A., Sawicki, J. T., Blake, D. M., Raymond, L. G. & Dipple, A. Products of binding of 7,12-dimethylbenz(a)anthracene to DNA in mouse skin. *Cancer Res.* **43,** 5647–51 (1983).

104. Hennings, H., Shores, R., Mitchell, P., Spangler, E. F. & Yuspa, S. H. Induction of papillomas with a high probability of conversion to malignancy. *Carcinogenesis* **6,** 1607–10 (1985).

105. Brown, K., Strathdee, D., Bryson, S., Lambie, W. & Balmain, A. The malignant capacity of skin tumours induced by expression of a mutant H-ras transgene depends on the cell type targeted. *Curr. Biol.* **8,** 516–24 (1998).

106. Nguyen, D. X., Bos, P. D. & Massagué, J. Metastasis: from dissemination to organ-specific colonization. *Nat. Rev. Cancer* **9,** 274–84 (2009).

107. McCann, J., Spingarn, N. E., Kobori, J. & Ames, B. N. Detection of carcinogens as mutagens: bacterial tester strains with R factor plasmids. *Proc. Natl. Acad. Sci.* **72,** 979–83 (1975).

108. Fialkow, P. J., Gartler, S. M. & Yoshida, A. Clonal origin of chronic myelocytic leukemia in

man. *Proc. Natl. Acad. Sci.* **58,** 1468–71 (1967).

109. Novelli, M. R. *et al.* Polyclonal origin of colonic adenomas in an XO/XY patient with FAP. *Science* **272,** 1187–90 (1996).

110. Dotto, G. P. Multifocal epithelial tumors and field cancerization: Stroma as a primary determinant. *J. Clin. Invest.* **124,** 1446–53 (2014).

111. Dvorak, H. F. Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. *N. Engl. J. Med.* **315,** 1650–9 (1986).

112. Plikus, M. V. *et al.* Epithelial stem cells and implications for wound repair. *Semin. Cell Dev. Biol.* **23,** 946–53 (2012).

113. Shou, M. *et al.* Specificity of cDNA-expressed human and rodent cytochrome P450s in the oxidative metabolism of the potent carcinogen 7,12-dimethylbenz[a]anthracene. *Mol. Carcinog.* **17,** 241–9 (1996).

114. Meyer, K. C. *et al.* Evidence that the bulge region is a site of relative immune privilege in human hair follicles. *Br. J. Dermatol.* **159,** 1077–85 (2008).

115. Clayton, E. *et al.* A single type of progenitor cell maintains normal epidermis. *Nature* **446,** 185–9 (2007).

116. Kwon, M. & Berns, A. Tumor heterogeneity-induced signaling regulates SCLC metastasis. *Cell Cycle* **14,** 3347–8 (2015).

117. McKenna, A. *et al.* The Genome Analysis Toolkit : A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).

118. Brown, J. R. & Thornton, J. L. Percivali Pott (1714-1788) and Chimney Sweepers' Cancer of the Scrotum. *Br. J. Ind. Med.* **14,** 68–70 (1957).

119. Yamagiwa, K. & Ichikawa, K. Experimental study of the pathogenesis of carcinoma. *J. Cancer Res.* **27,** 123–81 (1918).

120. Berenblum, I. & Shubik, P. A new, quantitative, approach to the study of the stages of chemical cartinogenesis in the mouse's skin. *Br. J. Cancer* **1,** 383–91 (1947).

121. Hecker, E. Cocarcinogenic principles from the seed oil of Croton tiglium and from other Euphorbiaceae. *Cancer Res.* **28,** 2338–2349 (1968).

122. Takigawa, M., Verma, A. K., Simsiman, R. C. & Boutwell, R. K. Inhibition of mouse skin tumor promotion and of promoter-stimulated epidermal polyamine biosynthesis by alpha-Difluoromethylornithine. *Carcinogenesis* **43,** 3732–8 (1983).

123. Lindahl, T. The Intrinsic Fragility of DNA (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* (2016).

124. Engelman, J. A. & Settleman, J. Acquired resistance to tyrosine kinase inhibitors during cancer therapy. *Curr. Opin. Genet. Dev.* **18,** 73–9 (2008).

125. Petljak, M. & Alexandrov, L. B. Understanding mutagenesis through delineation of mutational signatures in human cancer. *Carcinogenesis* **37,** 531–540 (2016).

126. Viros, A. *et al.* Ultraviolet radiation accelerates BRAF-driven melanomagenesis by targeting TP53. *Nature* **511,** 478–482 (2014).

127. Pfeifer, G. P., You, Y.-H. & Besaratinia, A. Mutations induced by ultraviolet light. *Mutat. Res.* **571,** 19–31 (2005).

128. Kucab, J. E. *et al.* TP53 mutations induced by BPDE in Xpa-WT and Xpa-Null human TP53 knock-in (Hupki) mouse embryo fibroblasts. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **773,** 48–62 (2015).

129. Forkert, P. G. Mechanisms of lung tumourigenesis by ethyl carbamate and vinyl carbamate. *Drug Metab. Rev.* **42,** 355–78 (2010).

130. Kurowska, M., Labocha-Pawlowska, A., Gnizda, D., Maluszynski, M. & Szarejko, I. Molecular analysis of point mutations in a barley genome exposed to MNU and gamma rays. *Mutat. Res.* **738–739,** 52–70 (2012).

131. Shih, C., Shilo, B. Z., Goldfarb, M. P., Dannenberg, A. & Weinberg, R. a. Passage of phenotypes of chemically transformed cells via transfection of DNA and chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 5714–8 (1979).

132. Shilo, B. Z. & Weinberg, R. a. Unique transforming gene in carcinogen-transformed mouse cells. *Nature* **289,** 607–9 (1981).

133. Perucho, M. *et al.* Human-tumor-derived cell lines contain common and different transforming genes. *Cell* **27,** 467–76 (1981).

134. Santos, E., Tronick, S. R., Aaronson, S. A., Pulciani, S. & Barbacid, M. T24 human bladder carcinoma oncogene is an activated form of the normal human homologue of BALB- and Harvey-MSV transforming genes. *Nature* **298,** 343–7 (1982).

135. Balmain, A. & Yuspa, S. H. Milestones in skin carcinogenesis: the biology of multistage carcinogenesis. *J Invest Dermatol* **134,** E2-7 (2014).

136. Bizub, D., Wood, A. W. & Skalka, A. M. Mutagenesis of the Ha-ras oncogene in mouse skin tumors induced by polycyclic aromatic hydrocarbons. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 6048–52 (1986).

137. Brown, K., Buchmann, A. & Balmain, A. Carcinogen-induced mutations in the mouse c-Ha-ras gene provide evidence of multiple pathways for tumor progression. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 538–42 (1990).

138. Zarbl, H., Sukumar, S., Arthur, a V, Martin-Zanca, D. & Barbacid, M. Direct mutagenesis of Ha-ras-1 oncogenes by N-nitroso-N-methylurea during initiation of mammary carcinogenesis in rats. *Nature* **315,** 382–385 (1985).

139. Sikpi, M. O., Waters, L. C., Kraemer, K. H., Preston, R. J. & Mitra, S. N-methyl-N-nitrosourea-induced mutations in a shuttle plasmid replicated in human cells. *Mol. Carcinog.* **3,** 30–36 (1990).

140. You, M., Candrian, U., Maronpot, R. R., Stoner, G. D. & Anderson, M. W. Activation of the Ki-ras protooncogene in spontaneously occurring and chemically induced lung tumors of the strain A mouse. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 3070–3074 (1989).

141. Ozturk, M. p53 mutation in hepatocellular carcinoma after aflatoxin exposure. *Lancet* **338,** 1356–9 (1991).

142. Hsu, I. C. *et al.* Mutational hotspot in the p53 gene in human hepatocellular carcinomas. *Nature* **350,** 427–8 (1991).

143. Schulze, K. *et al.* Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat. Genet.* **47,** 505–11 (2015).

144. Schmeiser, H. H. *et al.* Aristolochic acid activates ras genes in rat tumors at deoxyadenosine residues. *Cancer Res.* **50,** 5464–9 (1990).

145. Poon, S. L. *et al.* Genome-wide mutational signatures of aristolochic acid and its application as a screening tool. *Sci. Transl. Med.* **5,** 197ra101 (2013).

146. Polak, P. *et al.* Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **518,** 360–364 (2015).

147. Nik-Zainal, S. *et al.* The genome as a record of environmental exposure. *Mutat. Res.* **30,** 763–70 (2015).

148. Keller, R. R. *et al.* Carcinogen-specific mutations in preferred Ras-Raf pathway oncogenes directed by strand bias. *Carcinogenesis* **Epub ahead,** (2016).

149. Skopek, T. R., Walker, V. E., Cochrane, J. E., Craft, T. R. & Cariello, N. F. Mutational spectrum at the Hprt locus in splenic T cells of B6C3F1 mice exposed to N-ethyl-N-nitrosourea. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 7866–70 (1992).

150. Buchmann, A., Karcier, Z., Schmid, B., Strathmann, J. & Schwarz, M. Differential selection for B-raf and Ha-ras mutated liver tumors in mice with high and low susceptibility to hepatocarcinogenesis. *Mutat. Res.* **638,** 66–74 (2008).

151. Friedewald, W. F. & Rous, P. The initiating and promoting elements in tumor production: An analysis of the effects of tar, benzpyrene, and methylcholanthrene on rabbit skin. *1944* **80,** 101–26

152. Berenblum, I. & Shubik, P. The role of croton oil applications, associated with a single painting of a carcinogen, in tumour induction of the mouse's skin. *Br. J. Cancer* **1,** 379–82 (1947).

153. Boutwell, R. K., Takigawa, M., Verma, A. K. & Ashendel, C. L. Observations on the mechanism of skin tumor promotion by phorbol esters. *Princess Takamatsu Symp.* **14,** 177–93 (1983).

154. Yuspa, S. H. The pathogenesis of squamous cell cancer: lessons learned from studies of skin carcinogenesis--thirty-third G. H. A. Clowes Memorial Award Lecture. *Cancer Res.* **54,** 1178–89 (1994).

155. Delclos, K. B., Nagle, D. S. & Blumberg, P. M. Specific binding of phorbol ester tumor promoters to mouse skin. *Cell* **19,** 1025–32 (1980).

156. Castagna, M. *et al.* Direct activation of calcium-activated, phospholipid-dependent protein kinase by tumor-promoting phorbol esters. *J. Biol. Chem.* **257,** 7847–51 (1982).

157. Ushmorov, A. G., Furstenberger, G., Faissner, A. & Marks, F. Effects of complete and incomplete tumor promoters on hair growth, angiogenesis, and tenascin expression in the skin of NMRI mice. *Carcinogenesis* **15,** 2739–45 (1994).

158. Rehman, I. *et al.* Frequent codon 12 Ki-ras mutations in mouse skin tumors initiated by N-methyl-N'-nitro-N-nitrosoguanidine and promoted by mezerein. *Mol. Carcinog.* **27,** 298–307 (2000).

159. Megosh, L., Halpern, M., Farkash, E. & O'Brien, T. G. Analysis of ras gene mutational spectra in epidermal papillomas from K6/ODC transgenic mice. *Mol. Carcinog.* **22,** 145–9 (1998).

160. Jansen, A. P. *et al.* Relation of the induction of epidermal ornithine decarboxylase and hyperplasia to the different skin tumor-promotion susceptibilities of protein kinase C alpha, -delta and -epsilon transgenic mice. *Int. J. cancer* **93,** 635–43 (2001).

161. Zayed, S., Sorg, B. & Hecker, E. Structure activity relations of polyfunctional diterpenes of the tigliane type, VI. Irritant and tumor promoting activities of semisynthetic mono and diesters of 12-deoxyphorbol. *Planta Med.* **50,** 65–9 (1984).

162. Szallasi, Z., Krsmanovic, L. & Blumberg, P. M. Nonpromoting 12-deoxyphorbol 13-esters

inhibit phorbol 12-myristate 13-acetate induced tumor promotion in CD-1 mouse skin. *Cancer Res.* **53,** 2507–12 (1993).

163. Wang, M. T. *et al.* K-Ras Promotes Tumorigenicity through Suppression of Non-canonical Wnt Signaling. *Cell* **163,** 1237–51 (2015).

164. Aydinlik, H., Ngugen, T. D., Moennikes, O., Buchmann, A. & Schwarz, M. Selective pressure during tumor promotion by phenobarbital leads to clonal outgrowth of beta-catenin-mutated mouse liver tumors. *Oncogene* **20,** 7812–6 (2001).

165. Loktionov, A. *et al.* Tissue-specific activating mutations of Ha- and Ki-ras oncogenes in skin, lung, and liver tumors induced in mice following transplacental exposure to DMBA. *Mol. Carcinog.* **3,** 134–40 (1990).

166. Sukumar, S. & Barbacid, M. Specific patterns of oncogene activation in transplacentally induced tumors. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 718–22 (1990).

167. Su, F. *et al.* RAS mutations in cutaneous squamous-cell carcinomas in patients treated with BRAF inhibitors. *N. Engl. J. Med.* **366,** 207–15 (2012).

168. Poulikakos, P. I., Zhang, C., Bollag, G., Shokat, K. M. & Rosen, N. RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. *Nature* **464,** 427–30 (2010).

169. Robert, C., Arnault, J. P. & Mateus, C. RAF inhibition and induction of cutaneous squamous cell carcinoma. *Curr. Opin. Oncol.* **23,** 177–82 (2011).

170. To, M. D. *et al.* A functional switch from lung cancer resistance to susceptibility at the Pas1 locus in Kras2LA2 mice. *Nat. Genet.* **38,** 926–930 (2006).

171. Song, I. Y. & Balmain, A. Cellular reprogramming in skin cancer. *Semin. Cancer Biol.* **32,** 32–9 (2015).

172. Popova, N. V *et al.* Independent inheritance of genes regulating two subpopulations of mouse clonogenic keratinocyte stem cells. *J. Investig. dermatology symoposium Proc.* **9,** 253–60 (2004).

173. Wakabayashi, Y., Mao, J.-H., Brown, K., Girardi, M. & Balmain, A. Promotion of Hras-induced squamous carcinomas by a polymorphic variant of the Patched gene in FVB mice. *Nature* **445,** 761–5 (2007).

174. Nebert, D. W., Dalton, T. P., Okey, A. B. & Gonzalez, F. J. Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. *J. Biol. Chem.* **279,** 23847–50 (2004).

175. Quigley, D. A. *et al.* Genetic architecture of mouse skin inflammation and tumour susceptibility. *Nature* **458,** 505–8 (2009).

176. Ewart-Toland, A. *et al.* Identification of Stk6/STK15 as a candidate low-penetrance tumor-susceptibility gene in mouse and human. *Nat. Genet.* **34,** 403–12 (2003).

177. To, M. D. *et al.* Progressive genomic instability in the FVB/Kras(LA2) mouse model of lung cancer. *Mol. Cancer Res.* **9,** 1339–45 (2011).

178. Turajlic, S. & Swanton, C. Metastasis as an evolutionary process. *Science* **352,** 169–75 (2016).

179. Takahashi, M., Hori, M., Mutoh, M., Wakabayashi, Y. & Nakagama, H. Experimental animal models of pancreatic carcinogenesis for prevention studies and their relevance to human disease. *Cancers (Basel)* **3,** 582–602 (2011).

180. Hoenerhoff, M. J., Hong, H. H., Ton, T. V, Lanhousse, S. A. & Sills, R. C. A review of the molecular mechanisms of chemically induced neoplasia in rat and mouse models in National Toxicology Program bioassays and their relevance to human cancer. *Toxicol. Pathol.* **37,** 835–48 (2009).

181. Nagao, M. *et al.* Human exposure to carcinogenic heterocyclic amines and their mutational fingerprints in experimental animals. *Environ. Health Perspect.* **104,** 497–501 (1996).

182. Nakagama, H., Nakanishi, M. & Ochiai, M. Modeling human colon cancer in rodents using a food-borne carcinogen, PhIP. *Cancer Sci.* **96,** 627–36 (2005).

183. Hanahan, D., Wagner, E. F. & Palmiter, R. D. The origins of oncomice: a history of the first transgenic mice genetically engineered to develop cancer. *Genes Dev.* **21,** 2258–70 (2007).

184. Kwon, M. C. & Berns, A. Mouse models for lung cancer. *Mol. Oncol.* **7,** 165–77 (2013).

185. Jacks, T. Modeling cancer in the mouse. *Harvery Lect.* **101,** 1–19 (2005).

186. Nardella, C., Lunardi, A., Patnaik, A., Cantley, L. C. & Pandolfi, P. P. The APL paradigm and the 'co-clinical trial' project. *Cancer Discov.* **1,** 108–16 (2011).

187. Clohessy, J. G. & Pandolfi, P. P. Mouse hospital and co-clinical trial project--from bench to bedside. *Nat. Rev. Clin. Oncol.* **12,** 491–8 (2015).

188. Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457,** 608–11 (2009).

189. Peterson, S. C. *et al.* Basal cell carcinoma preferentially arises from stem cells within hair follicle and mechanosensory niches. *Cell Stem Cell* **16,** 400–12 (2015).

190. Youssef, K. K. *et al.* Identification of the cell lineage at the origin of basal cell carcinoma. *Nat. Cell Biol.* **12,** 299–305 (2010).

191. Wang, G. Y., Wang, J., Mancianti, M. L. & Epstein, E. H. Basal cell carcinomas arise from hair follicle stem cells in Ptch1(+/-) mice. *Cancer Cell* **19,** 114–24 (2011).

192. Mitsui, J. *et al.* Two distinct mechanisms of augmented antitumor activity by modulation of immunostimulatory/inhibitory signals. *Clin. cancer Res.* **16,** 2781–91 (2010).

193. Gubin, M. M. *et al.* Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515,** 577–581 (2014).

194. Aleksic, K. *et al.* Evolution of Genomic Instability in Diethylnitrosamine- Induced Hepatocarcinogenesis in Mice. *Hepatology* **53,** 895–904 (2011).

195. Vesselinovitch, S. D. & Mihailovich, N. Kinetics of diethylnitrosamine hepatocarcinogenesis in the infant mouse. *Cancer Res.* **43,** 4253–9 (1983).

196. Da Costa, R. M. G. *et al.* The N-Nitrosodiethylamine Mouse Model: Sketching a Timeline

of Evolution of Chemically-induced Hepatic Lesions. *Anticancer Res.* **7038,** 7029–7038 (2014).

197. Verna, L., Whysner, J. & Williams, G. M. N-Nitrosodiethylamine Mechanistic Data and Risk Assessment: Bioactivation, DNA-Adduct Formation, Mutagenicity, and Tumor Initiation. *Pharmacol. Ther.* **71,** 57–81 (1996).

198. Malkinson, A. M. Genetic studies on lung tumor susceptibility and histogenesis in mice. *Environ. Health Perspect.* **93,** 149–59 (1991).

199. Foley, E. J. Antigenic Properties of Methylcholanthrene-induced Tumors in Mice of the Strain of Origin. *Cancer Res.* **13,** 835–837 (1953).

200. Riggins, R. S. & Pilcht, Y. Immunity to Spontaneous and Methylcholanthrene-induced Tumors in Inbred Mice. *Cancer Res.* **24,** 1994–1997 (1994).

201. Algarra, I., Perez, M., Serrano, M. J., Garrido, F. & Gaforio, J. J. c-K-ras overexpression is characteristic for metastases derived from a methylcholanthrene-induced fibrosarcoma. *Invasion Metastasis* **18,** 261–70 (1998).

202. Borrello, M. G., Carbone, G., Pierotti, M. A., Molla, A. & Della Porta, G. Activated c-K-ras and c-N-ras oncogenes in 3-methylcholanthrene-induced BALB/c fibrosarcomas. *Carcinogenesis* **9,** 1517–9 (1988).

203. Ogawa, K. *et al.* Frequent p53 mutations and occasional loss of chromosome 4 in invasive bladder carcinoma induced by N-butyl-N-(4-hydroxybutyl)nitrosamine in B6D2F1 mice. *Mol. Carcinog.* **21,** 70–9 (1998).

204. He, Z., Kosinska, W., Zhao, Z., Wu, X. & Guttenplan, J. B. Mutagenesis Tissue-specific mutagenesis by N-butyl-N- (4-hydroxybutyl) nitrosamine as the basis for urothelial carcinogenesis. *Mutat. Res.* **742,** 92–5 (2012).

205. Vasconcelos-Nobrega, C., Colaco, A., Lopes, C. & Oliveira, P. A. Review: BBN as an urothelial carcinogen. *In Vivo (Brooklyn).* **26,** 727–39 (2012).

206. Nishikawa, A. *et al.* Induction of colon tumors in C57BL/6J mice fed MeIQx, IQ, or PhIP

followed by dextran sulfate sodium treatment. *Toxological Sci.* **84,** 243–8 (2005).

207. Dias, M. *et al.* Benign and malignant mammary tumors induced by DMBA in female Wistar rats. *Eur. J. Gynaecol. Oncol.* **20,** 285–8 (1999).

208. Faustino-Rocha, A. I., Ferreira, R., Oliveira, P. A., Gama, A. & Ginja, M. N-Methyl-N-nitrosourea as a mammary carcinogenic agent. *Tumor Biol.* **36,** 9095–9117 (2015).

209. Medina, D. & Warner, M. R. Mammary tumorigenesis in chemical carcinogen-treated mice. IV. Induction of mammary ductal hyperplasias. *J. Natl. Cancer Inst.* **57,** 331–7 (1976).

210. Pour, P. M., Salmasi, S. Z. & Runge, R. G. Selective induction of pancreatic ductular tumors by single doses of N-nitrosobis(2-oxopropyl)amine in Syrian golden hamsters. *Cancer Lett.* **4,** 317–23 (1978).

211. Fujii, H., Egami, H., Chaney, W., Pour, P. & Pelling, J. Pancreatic ductal adenocarcinomas induced in Syrian hamsters by N-nitrosobis(2-oxopropyl)amine contain a c-Ki-ras oncogene with a point-mutated codon 12. *Mol. Carcinog.* **3,** 296–301 (1990).

212. Matsushita, H. *et al.* Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. *Nature* **482,** 400–404 (2012).

213. McFadden, D. G. *et al.* Mutational landscape of EGFR-, MYC-, and Kras-driven genetically engineered mouse models of lung adenocarcinoma. *Proc. Natl. Acad. Sci.* **113,** E6409-17 (2016).

214. Cha, E. *et al.* Improved Survival with T Cell Clonotype Stability After Anti – CTLA-4 Treatment in Cancer Patients. *Sci. Transl. Med.* **6,** 238ra70 (2014).

215. To, M. D., Rosario, R. D., Westcott, P. M. K., Banta, K. L. & Balmain, A. Interactions between wild-type and mutant Ras genes in lung and skin carcinogenesis. *Oncogene* **32,** 4028–33 (2013).

216. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* **12,** 3243–8 (2001).

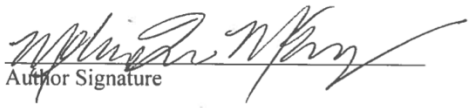
217. Jaks, V. *et al.* Lgr5 marks cycling, yet long-lived, hair follicle stem cells. *Nat. Genet.* **40,** 1291–9 (2008).

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

6/30/17

Date