

Deep sequencing as a probe of normal stem cell fate and preneoplasia in human epidermis

Benjamin D. Simons^{a,b,c,1}

^aCavendish Laboratory, Department of Physics, University of Cambridge, Cambridge CB3 0HE, United Kingdom; ^bWellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge CB2 1QN, United Kingdom; and ^cWellcome Trust–Medical Research Council Stem Cell Institute, University of Cambridge, Cambridge CB2 1QR, United Kingdom

Edited by Brigid L. M. Hogan, Duke University Medical Center, Durham, NC, and approved November 23, 2015 (received for review August 13, 2015)

Using deep sequencing technology, methods based on the sporadic acquisition of somatic DNA mutations in human tissues have been used to trace the clonal evolution of progenitor cells in diseased states. However, the potential of these approaches to explore cell fate behavior of normal tissues and the initiation of preneoplasia remain underexploited. Focusing on the results of a recent deep sequencing study of eyelid epidermis, we show that the quantitative analysis of mutant clone size provides a general method to resolve the pattern of normal stem cell fate and to detect and characterize the mutational signature of rare field transformations in human tissues, with implications for the early detection of preneoplasia.

stem cells | DNA sequencing | epidermis | cancer

Advances in genetic lineage tracing in transgenic animal models have provided important insights into the proliferative potential and fate behavior of stem and progenitor cell populations in normal tissues (1, 2). As well as providing constraints on the mechanisms that regulate stem cell self-renewal, these approaches have established a quantitative framework to address tumor initiation and progression (3–6). However, studies based on the clonal activation of oncogenes in animal models can fail to recapitulate the natural processes that lead to neoplasia in human tissues. In recent years, there has been increasing emphasis on the characterization of cancer genomes in human tissues and their potential to elucidate the pathways involved in tumor progression (7–14). Although these studies have revealed a range of cancer genes (15), the heterogeneity and evolutionary diversity of the tumor environment make the separation of driver and passenger mutations challenging.

Against the trend to focus on human tumor samples, a recent study has used ultradeep exome sequencing to determine the mutational profile of normal human eyelid epidermis (16). In the course of DNA replication, all dividing cells are subject to random SNPs. If the mutation rate is sufficiently low that their acquisition at a given locus in a cell subpopulation is typically associated with a single event, they confer a potentially unique hereditary label on cells, allowing the fate of their progeny to be traced over time. By resolving the mutant allele fraction in a biopsy using deep sequencing, the relative size of mutant clones can be inferred. A similar approach based on the spontaneous acquisition of mitochondrial DNA mutation has been used to address progenitor cell fate in human airways and intestinal epithelia (17, 18). To assess the selective growth advantage of different mutations in normal epidermis, Martincorena et al. (16) compared the dN/dS ratio and average size of clones derived from mutations in genes associated with cancer drivers with those associated with synonymous mutations in nondriver genes. Their analysis showed a significant increase in the abundance and average size of clones that bear mutations in NOTCH1 and tumor protein p53 (TP53) compared with the ensemble of apparently neutral mutations, whereas mutations in other drivers such as FAT1, NOTCH2, and NOTCH3 were not significantly increased. Based on these findings, the study reached the striking conclusion that cancer genes are under strong positive selection, even in physiologically normal skin. However, paradoxically, despite the apparent survival advantage, average

clone sizes even in TP53 mutant clones were only a modest factor of two larger than the ensemble average, suggesting that the degree of clonal dominance may be limited.

At first sight, one might expect that the relative abundance of gene-specific point mutations could reveal whether they confer a selective survival advantage. However, although variations in the observed frequency of SNPs will arise from the positive/negative selection of somatic mutations, they may also be intrinsic (germline-derived), making their functional significance at different sites difficult to assess (19, 20) (Fig. S1). Equally, the value of average mutant clone sizes is diminished by their sensitivity to tails of the size distribution, which can be compromised by the resolution limit of sequencing or statistical fluctuations due to rare events. Similar effects may compromise the dN/dS ratio, a measure of the relative abundance of nonsynonymous to synonymous mutations (21). However, by analyzing the full probability distribution of mutant clone sizes, and drawing upon knowledge of adult stem cell self-renewal strategies (1, 22), we show that quantitative insights can be gained into the dynamics and fate behavior of mutant clones, providing access to both the normal state properties of tissue-maintaining cells and their dynamics following premalignant transformation. In doing so, we offer a different perspective on the deep sequencing study of Martincorena et al. (16).

Deep Sequencing As a Clonal Marker in Human Epidermis

In mammals, skin is composed of a multilayered sheet of keratinocytes interspersed with hair follicles, sebaceous glands, and sweat glands (23). Lineage tracing studies using transgenic mouse models have revealed a surprising degree of compartmentalization, with the turnover of hair follicle, sebaceous gland, and interfollicular epidermis (IFE) maintained by independent stem-cell populations (24). In IFE, proliferation is confined to

Significance

The sporadic acquisition of somatic DNA mutation confers a hereditary label that can be used to trace the fate behavior of cells in normal and diseased states. Applied to human tumor samples, DNA deep sequencing methods have revealed the landscape of somatic mutations and have identified a repertoire of genes implicated in cancer. By adapting statistical methods used to analyze lineage tracing data in transgenic animal models, we use the example of epidermis to show how deep sequencing data can provide quantitative insight into the self-renewal properties of normal human tissues and can serve as a platform to define rare nonneutral field transformations.

Author contributions: B.D.S. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission.

¹Email: bds10@cam.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516123113/-DCSupplemental.

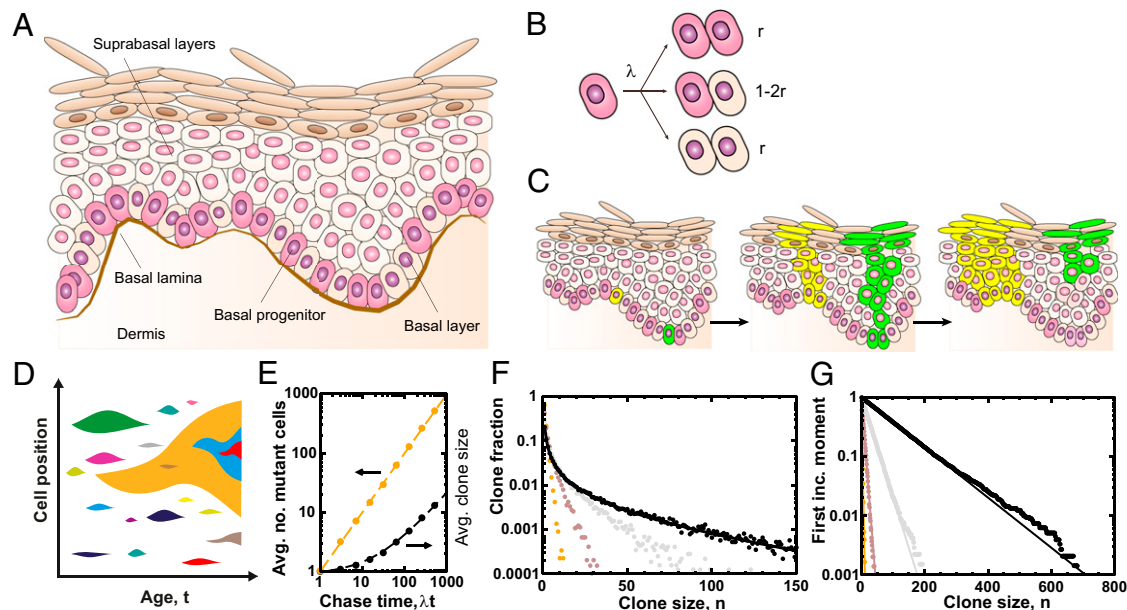


Fig. 1. Dynamics of clones in mammalian IFE. (A) Schematic depicting the cellular organization of human IFE. (B) In the paradigm of population asymmetry, IFE is maintained by basal progenitors that, following division, choose stochastically between symmetric and asymmetric fate with the probability of symmetric proliferation, $0 < r \leq 1/2$, perfectly balanced by terminal division. (C) When lineage-labeled, the clonal progeny of marked progenitors may expand through the replacement of neighbors (yellow clone) or may become lost through differentiation (green clone). (D) Following the chance acquisition of somatic mutation, clones (with colors denoting different SNPs) compete for survival. Over time, some clones become lost through differentiation whereas others expand. Clones that bear one mutation may acquire further mutations. (E) According to the model depicted in B, the average size of mutant clones (orange line) is predicted to increase as ωNt , where ω denotes the mutation rate per progenitor at a given locus, N is the size of the progenitor pool, and t is the age of the patient. The average mutant clone size is predicted to increase as $r\lambda t / \ln(r\lambda t)$ (black line). (F) The corresponding mutant clone size distribution, $P_n(t)$, and (G) first incomplete moment, $\mu_1(n, t)$, predicted by Eqs. 1 and 2, respectively (black lines). In E–G points show the results of stochastic simulation of the birth–death process depicted in B, where, in F and G, $\lambda t = 2^6$ (orange), 2^8 (brown), 2^{10} (gray), and 2^{12} (black) with $\omega N/\lambda = 1$ and $r = 1/2$.

cells in the basal layer that adhere to an underlying basement membrane (Fig. 1A). On commitment to terminal differentiation, basal cells detach from the basement membrane and transfer into the suprabasal layers before reaching the epidermal surface from which they are shed. In homeostasis, the progenitors that maintain IFE must undergo asymmetric self-renewal so that, following division, on average one cell remains in the self-renewing compartment, whereas the other commits to differentiation either directly or via a transit compartment with strictly limited proliferative potential. Such asymmetry may be invariant, enforced at the level of each and every cell division, or it may be achieved only at the level of the progenitor population (*SI Text*).

Beginning with the work of Mackenzie (25) and Potten et al. (26), early studies of IFE maintenance in mouse placed emphasis on a stem/transit-amplifying cell paradigm in which long-lived slow-cycling stem cells give rise to short-lived progenitors that undergo a limited series of symmetric division before terminal differentiation. Later, quantitative lineage tracing studies based on inducible genetic labeling revealed that murine epidermal maintenance relies instead upon the turnover of a basal progenitor pool that conforms to a process of “population asymmetry” in which their stochastic loss through terminal division is perfectly compensated by the duplication of neighbors (27–30) (Fig. 1B). Whether the repair of murine epidermis involves a transient adjustment in the fate behavior of the progenitor pool or is engineered by the activity of a second quiescent “reserve” stem cell population remains the subject of debate. In human, *in vitro* colony-forming assays, as well as transplantation and marker-based studies, point at engrained proliferative heterogeneity in the basal layer of IFE (31–34). However, in the absence of *in vivo* lineage tracing assays, the nature of stem cell self-renewal and tissue maintenance remains in question.

The resolution of cell fate behavior in mouse IFE relied upon the observation of “scaling” behavior of the clone size distribution following genetic pulse labeling (27–30, 35). According to their stochastic fate behavior (Fig. 1B), as progenitors compete neutrally for survival the density of clones (number per unit area) progressively diminishes, whereas the average size of surviving clones steadily increases (linearly with time) so that the overall number of marked cells remains constant over time (*SI Text*, Fig. 1C, and Fig. S2). However, despite their continual increase in size, the chance of finding a surviving clone larger than some multiple of the average remains constant and defined by an exponential distribution (Fig. S2). Combined with the overall conservation of labeled cell number, this phenomenon of scaling provides a robust, parameter-free signature of neutral cell competition and equipotency of the tissue-maintaining population (35). Although the exponential size dependence is particular to epithelial (and volumnar) tissues, the phenomenon of scaling applies generically to all cycling adult tissues supported by population asymmetry (*SI Text*). As a result, the same general approach has been used successfully to explore stem cell fate behavior in other tissues and organisms (1).

In contrast to genetic labeling approaches, where the induction frequency can be controlled through the dose dependence of the drug-inducing agent, clonal marking by somatic mutation involves a sequence of sporadic events masking the age of individual clones (Fig. 1D). Fortunately, under conditions of neutral competition, quantitative information on the fate behavior of the self-renewing population can still be recovered. In particular, if the pattern of stochastic progenitor cell fate observed in mouse IFE (Fig. 1B) were extrapolated to human, then, following the continual “induction” of clonally marked cells through the acquisition of somatic mutation, the probability of finding a mutant clone with $n > 0$ progenitors in a biopsy of a patient of age t would be independent of the (presumed unchanging) mutation rate and given by (17):

$$P_{n>0}(t) \approx \frac{1}{\ln(r\lambda t)} \frac{e^{-n/r\lambda t}}{n}, \quad [1]$$

where λ represents the division rate and $r\lambda$ denotes the loss/replacement rate of basal progenitors (Fig. 1 *B*, *E*, and *F* and *SI Text*). The “featureless” $1/n$ divergence of the distribution at small clone sizes is simply a manifestation of neutral dynamics that results in the largest fraction of surviving clones at any instant being ones that were “induced” in the recent past (Fig. 1*D*).

If rates of somatic mutation are sufficiently high, SNPs may arise independently at the same locus in different cells. Because estimates of mutant clone size using deep sequencing are based on measurements of the variable allele fraction (VAF), the multiplicity of induction events cannot be resolved. Fortunately, would such clone “merger” events occur, they would be signaled by a breakdown of the leading $1/n$ dependence, allowing their existence to be inferred indirectly (*SI Text*). However, while $\omega N/r\lambda \ll 1$, where N denotes the number of progenitors in a given biopsy and ω is the mutation rate associated with the given locus, the frequency of mutant clones derived from multiple induction events can be safely neglected (*SI Text*). To proceed, we will assume that this condition is met and look for consistency of the data with theory.

Although Eq. 1 provides a strong prediction with which to address deep sequencing data, the nonlinear dependence of $P_n(t)$ on clone size, n , makes comparison between experiment and theory cumbersome. Fortunately, a further straightforward manipulation of the size distribution provides a more convenient representation. Specifically, defining the average mutant clone size, $\langle n(t) \rangle \equiv \sum_{n=1}^{\infty} n P_n(t) = r\lambda t / \ln(r\lambda t)$, it follows that the “first incomplete moment” (36),

$$\mu_1(n, t) = \frac{1}{\langle n(t) \rangle} \sum_{m=n}^{\infty} m P_m(t) \approx e^{-n/r\lambda t}, \quad [2]$$

acquires a simple exponential dependence on clone size, n , with a decay constant $r\lambda t$, equivalent to the average size of a surviving clone induced at birth (i.e., at the time of the first exposure to mutation) (Fig. 1 *E–G* and *SI Text*). Moreover, by the nature of its definition, the first incomplete moment is conveniently insensitive to the smallest clones, where the resolution of the deep sequencing approach is likely to be compromised. In the context of epidermis, it therefore follows that the corresponding distribution of clone areas, A , is given by $\mu_1(A, t) = e^{-\rho A/r\lambda t}$, with ρ the areal progenitor density.

Eq. 2 provides an objective, parameter-free prediction with which population asymmetry and neutrality of clone dynamics can be assessed. For a given array of biopsies, mutant clone sizes can be inferred from the corresponding VAFs associated with individual SNPs. Then the first incomplete moment, $\mu_1(A, t)$, can be constructed directly from the data. Departure of the inferred distribution from the predicted exponential size dependence would indicate functional heterogeneity of mutant clones and evidence of nonneutral dynamics. Convergence onto exponential would indicate that clone dynamics is likely governed by the neutral competition of an equipotent progenitor pool. A fit to the exponent, $r\lambda t/\rho$, then provides access to the progenitor loss/replacement rate. Crucially, the exponential size dependence of $\mu_1(A, t)$ is sensitive only to the dynamics of the self-renewing (i.e., the active stem cell) population. As long as the dominant contribution to the measured mutant clone size distribution derives from clones associated with mutations that occurred on time-scales in excess of the transit time through any differentiation hierarchy, the exponential size dependence would be conserved.

Neutral Competition Between Keratinocyte Progenitors

In the study of Martincorena et al. (16), eyelid epidermis was derived from more than 200 biopsies of sizes ranging from 0.79 to

4.71 mm² harvested from four patients aged 55–73 y of age. In each case, coding exons were sequenced across 74 genes implicated in skin and other cancers to an average effective coverage of 500 × (*Dataset S1*). (For technical details on the sample preparation and sequencing approach refer to ref. 16.) As detailed in the study, because few mutations involve change in copy number, the areal contribution of individual mutant clones can be inferred as twice the product of the VAF with the area of the biopsy. (Events involving the mutation of both alleles at a given locus are considered to occur at a negligible frequency.) For clones of a size much smaller than that of the biopsy, intersection of the clone with the boundary is statistically improbable. For larger clone sizes, the estimated clonal area in a given biopsy may represent only a fraction of the true size. However, the exponential character of the predicted first complete moment is robust to such statistical fluctuations as well as errors in the accuracy of the sequencing approach (*SI Text*).

To gain insight into the relative abundance of clones that “spill” outside individual biopsies, because the mutation rate is low (16) we can explore the coincidence of common point mutations found in different biopsies. Taking as an exemplar patient PD18003, for which the largest volume of data was obtained, we find that, from 1,557 specific point mutations across 92 biopsies, some 102 (6.6%) are present in more than one biopsy. Of these, 90 point mutations are restricted to two biopsies, 10 span three, 1 spans four, and 1 spans five. Similar frequencies of clone dispersion are found for the three other patients (Table 1), with one clone spanning no fewer than 12 biopsies.

Because the total area of clones that occupy multiple biopsies cannot be reliably recovered, we first focused on the ensemble of clones that bear a point mutation contained within a single biopsy. Further, to assess the utility of the approach, we began by focusing on the subset of these clones that involve only synonymous mutation (Fig. 2*A* and Fig. *S3*). Because such mutations leave the associated protein sequence unchanged, it is expected that the dynamics of the corresponding clones remains neutral, providing a useful control to benchmark theory. Focusing on patient PD18003, for which there were a total of 257 synonymous point mutations restricted to a single biopsy, analysis of the first incomplete moment, $\mu_1(A, t)$, reveals a remarkably exponential size dependence (Fig. 2*B*), consistent with neutral competition of the constituent progenitors. As well as justifying the validity of the approach, this result establishes that, under conditions of normal homeostasis, the progenitors that maintain adult human IFE conform long-term to population asymmetric self-renewal.

With the size distribution of clones associated with synonymous mutations defined, we then considered the wider class of mutant clones including both synonymous and nonsynonymous (missense and nonsense) mutations. Once again, taking the 1,338 clones associated with a single point mutation, the size distribution, $\mu_1(A, t)$, shows only a small departure from exponential with the divergence impacting at the largest clone sizes (Fig. 2*C*, arrowhead). By fitting the data to the exponential clone size dependence (red curve), we can then use the predicted cumulative frequency to estimate the

Table 1. Frequency of point mutations shared by multiple biopsies

Patient ID	No. of biopsies*							
	1	2	3	4	5	6	7	12
PD13634	647	63	7	4	3		1	
PD18003	1,338	90	10	1	1			
PD20399	724	64	10	1		3		1
PD21910	181	14						

*Multiplicity of point mutations that span multiple biopsies. For example, in patient PD13634, 647 point mutations are found in only one biopsy, 63 are found in two biopsies, and so on.

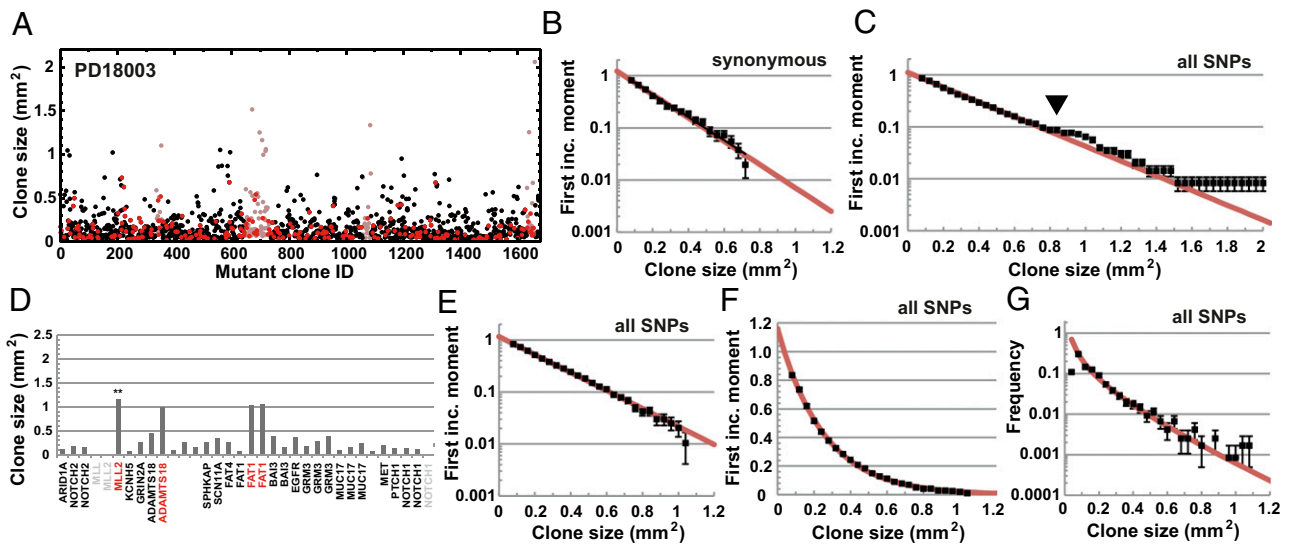


Fig. 2. Size distribution of mutant clones provides evidence of neutral dynamics. (A) Mutant clone sizes inferred from the analysis of the variable allele fractions of all ($n = 1,557$) point mutations for patient PD18003 (black points). Synonymous mutations are marked as red, and mutations belonging to the six biopsies considered outliers by size are shown in brown. (B) First incomplete moment, $\mu_1(A, t)$, associated with $n = 257$ synonymous mutations that belong to single biopsies (points) shows collapse onto the predicted exponential size dependence (Eq. 2, red line and Table 2). (C) First incomplete moment associated with all ($n = 1,338$) point mutations that belong to single biopsies. The departure of $\mu_1(A, t)$ (arrowhead) from exponential (red line) reflects contributions from seven mutant clones contained within six biopsies. (D) Clone sizes of one biopsy showing one of the outliers associated with mutation in MLL2 (**). Note that three other point mutations have clone sizes comparable to this outlier, suggesting that they belong to the same subclone. Nonpoint mutations are indicated as gray. When the six biopsies are excluded from the statistical ensemble, the distribution $\mu_1(A, t)$ (points) shown in E and F collapses onto the predicted exponential form (red line and Table 2). (G) The corresponding clone size distribution, $P(A, t)$, with points (black) showing data and the line (red) showing the theoretical prediction of Eq. 1. The departure of theory from experiment at clone sizes of 0.05 mm^2 and below indicates the resolution limit of the sequencing approach. With a basal cell density of $10,000$ cells per mm^2 , this suggests that a resolution limit of around 500 basal cells. Error bars denote SEM.

point of departure of the statistical distribution. Given the size of the ensemble of clones, we find that the observation of the seven largest clones with a size in excess of 1.08 mm^2 , a significant fraction of the size of the associated biopsies, would be statistically improbable within the framework of neutral dynamics (i.e., these clones would be predicted to occur with a frequency much less than 1 in 1,000). Furthermore, inspection of the mutational profile of the six biopsies containing the seven clones (Fig. S4) shows that five biopsies are associated with different point mutations (missense or frameshift deletion) in NOTCH1, whereas the sixth involves a missense mutation in MLL2. For the latter, three other mutations appear with a very similar VAF to MLL2 (Fig. 2D), suggesting that all four mutations belong to the same clone. Significantly, when these six biopsies are filtered out of the statistical cohort of 92, the first incomplete moment collapses onto a strikingly exponential size dependence (Fig. 2E and F). The coincidence of theory and experiment is further emphasized by comparison of the clone size distribution, $P(A, t)$, with the predicted size dependence (Fig. 2G).

Turning to patient PD13634, of the 725 discrete point mutations, 657 (89%) belong to a single biopsy with 159 of these associated with synonymous point mutations. Once again, their size distribution shows collapse onto an exponential dependence, consistent with neutral dynamics (Fig. S5A). Then, when combined with nonsynonymous mutations, the size distribution of all 657 mutant clones continues to collapse onto exponential with no apparent outliers by size (Figs. S2 and S5B). For patient PD20399, of the 803 discrete point mutations, 724 (90%) belong to a single biopsy. In this case, the size distribution of all 724 point mutations as well as the 154 mutant clones that bear a synonymous mutation also collapse onto exponential with no outliers by size (Figs. S2 and S5C and D). Finally, for patient PD21910, although the data are relatively sparse, of the 195 discrete mutations, 181 (93%) belong to a single biopsy. With just 35 of these mutant clones bearing a synonymous point mutation, the size distribution is noisy but consistent with exponential (Fig. S5E). Again, as expected,

comparison of all 181 mutations also reveals a collapse onto exponential with no outliers (Figs. S2 and S5F).

Nonneutral Expansion of Rare Mutant Clones

Although these results suggest that the vast majority of point mutations leave neutral dynamics unperturbed, the statistical method also provides a quantitative scheme to identify mutant clones that lie outside the normal (exponential) size distribution. Our results suggest that very few are associated with nonneutral dynamics—in one patient (PD18003), just six outlier clones were identified by size, whereas none were found in the other patients (Fig. 2 and Fig. S5). However, so far, we have excluded from our analysis mutant clones that span multiple biopsies. Because these clones are likely to be large, one might expect that they harbor the majority of cells that have undergone nonneutral transformation. Therefore, to gain insight into the nature of these dispersive clones, we explored the mutational profile of clones that spanned more than three biopsies.

Starting with patient PD18003, only one mutant clone bearing a missense mutation in SCN1A (C927S) spans more than three biopsies (Table 1), having an aggregate size of 2.52 mm^2 , more than a factor of two larger than the cutoff used to filter single-biopsy clones. Whether this outlier represents the chance expansion of a clone governed by neutral dynamics or derives from the proliferative advantage of mutant cells over their wild-type neighbors—the process of “field cancerization” (37)—driven by mutation of SCN1A is impossible to determine unambiguously. However, noting that the vast majority of the clone is limited to just one biopsy in which the point mutation in SCN1A is expressed with a VAF similar to that of point mutations associated with three other genes (Fig. 3A and B), it seems likely that expansion of this clone is driven by the chance acquisition of multiple point mutations. Indeed, by comparing the relative values of the VAFs, we can infer the likely order in which these point mutations were acquired (Fig. 3C).

Similarly, in patient PD20399, a clone bearing a missense point mutation in FGFR3 (R248C) spans no fewer than 12

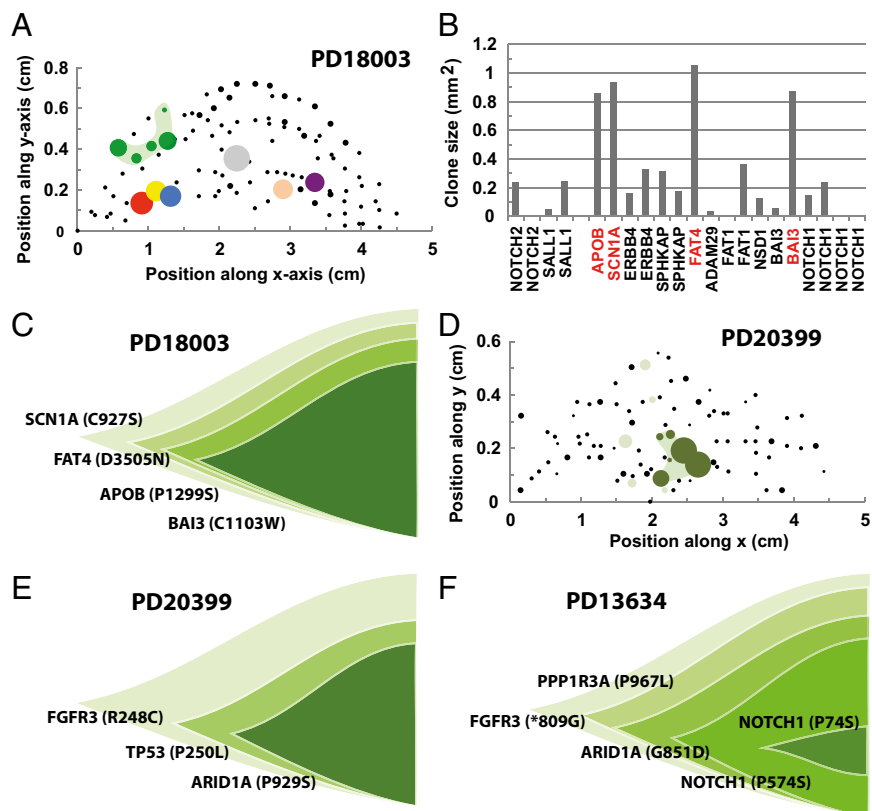


Fig. 3. Field transformation and nonneutral clone dynamics driven by multiple mutations. (A) Coordinates of biopsies of eyelid epidermis sampled for patient PD18003 with the size of points scaled by biopsy area. Colored points contained within single biopsies depict the six biopsies containing the seven mutant clones considered outliers by size. Green points mark five biopsies sharing the same point mutation in *SCN1A*. Size of colored dots scaled by the size of the outlier clones or the shared mutant clone. (B) Size of mutant clones belonging to one of the five biopsies with *SCN1A* mutation. Comparison of VAFs across all five biopsies suggests the clonal structure depicted in C. (D) Coordinates of biopsies from patient PD20399 with the size of dots scaled by the area of the biopsy. The 12 biopsies colored light or dark green share the same point mutation in *FGFR3* and the six biopsies marked in dark green also share point mutations in *TP53* and *ARID1A*. Green biopsies scaled by the size of the majority mutant clone. (E) Clonal organization implied by the mutational landscape of the clone depicted in D. (F) Analysis of mutational profile of mutant clones that span more than three biopsies in patient PD13634 reveal an oversized clone with the given clonal architecture.

biopsies covering an aggregate area of 7.41 mm². However, as noted by Martincorena et al. (16), in 6 of the 12 biopsies, this mutation appears alongside two other missense point mutations, one in *TP53* (P250L) and one in *ARID1A* (P929S), with all three bearing a very similar VAF (Fig. 3D). This coincidence suggests that it is the acquisition of these secondary mutations that drives nonneutral expansion of the clone. From the relative sizes of the three constituent mutations, we can infer the likely sequence of their acquisition (Fig. 3E). Interestingly, the large dispersion of the clone bearing the original mutation in *FGFR3* and its irregular spatial pattern (Fig. 3D) suggests that either mutation in *FGFR3* occurred independently at the same locus or this mutation may have a developmental origin. A second clone mutant for *NOTCH2* (P426S) spans five biopsies, but the majority lies within just two. In this case, its net aggregate size of 0.75 mm² suggests that it may belong to the ensemble of neutral mutations.

For patient PD13634, inspection of the mutational profile shows that one point mutation spans seven biopsies, three span five, and four span four. Inspection of the mutational profile shows these events can be traced to the expansion of just two clones and their subclones. Comparison of the VAFs of the constituent mutant clones suggests that, in one case, a consecutive sequence of five independent point mutations starting with *FGFR3* (*809G), followed by *PPP1R3A* (P967L), *ARID1A* (G851D), *NOTCH1* (P574S), and *NOTCH1* (P745), drives nonneutral expansion, leading to a clone with an aggregate size in excess of 8 mm² (Fig. 3F). A second independent clone, involving a synonymous mutation in *MUC17* (T3292T) followed by a nonsense mutation in *SPHKAP* (W308*), leads to a much smaller clone with an aggregate size of only 0.77 mm², well within the statistical ensemble of neutral mutations. Finally, for patient PD21910, there are no mutations that extend beyond two biopsies.

For consistency, we can further filter the ensemble of biopsies excluding those that contain the two oversized clones in patients

PD13634 and PD20399. Because these clones are subject to nonneutral expansion, they may affect neighbors by either suppressing their expansion or conveying them as passengers. Once removed, we find that the first incomplete moment maintains its exponential character, whereas the total clone size distribution falls onto the predicted size dependence (Fig. S6). Finally, to further challenge the hypothesis of neutrality, we determined the average mutant clone size across a range of genes. For all four patients, we found that departures of the average clone size associated with specific cancer drivers from that of the ensemble were not statistically significant (Fig. S7).

Although these findings suggest that the majority of mutations leave neutral dynamics unperturbed, it is important to consider what would emerge if the dynamics were nonneutral. If all point mutations conferred the same proliferative advantage, the first incomplete moment would also acquire an exponential size dependence, $\mu_1(n, t) \approx e^{-n/N(t)}$, with $N(t) = e^{vt}$ and v defining the net proliferative expansion rate of mutant progenitors (36). However, because such a size dependence would require all point mutations (synonymous and nonsynonymous) to confer precisely the same proliferative advantage, its relevance to the current

Table 2. Fit of the first incomplete moment to the predicted exponential size dependence

Fit parameter	Patient*			
	PD18003, 65 y old (F)	PD13634, 73 y old (M)	PD20399, 55 y old (F)	PD21910, 58 y old (F)
$r\lambda t/\rho$ (mm ² _{R2}) (syn.)	0.20 _{0.98}	0.22 _{0.98}	0.12 _{0.96}	0.31 _{0.87}
$r\lambda t/\rho$ (mm ² _{R2}) (all)	0.25 _{0.99}	0.24 _{0.94}	0.14 _{0.98}	0.32 _{0.96}
$r\lambda/\rho$ (mm ² ·y ⁻¹)	0.0039	0.0032	0.0025	0.0054

*Ratios $r\lambda t/\rho$ inferred from fits to $\mu_1(n, t)$ (red lines in Fig. 2 and Figs. S5E and S6). F, female; M, male.

study is unlikely. It is, however, important to note that, although the statistical approach provides the means to define clones that lie outside the normal size distribution, we cannot rule out the existence of a further subfraction of clones associated with nonneutral transformation that lie hidden within the bulk of the neutral distribution.

Discussion

These results demonstrate how analysis of deep sequencing data provides a general framework to study stem cell self-renewal of normal cycling adult human tissues. Applied to human IFE, we find that maintenance involves the turnover of a progenitor population following population asymmetry in which their stochastic loss through differentiation is compensated by duplication of neighbors. From a fit of the data to the exponential size dependence of $\mu_1(A, t)$, the inferred ratios $r\lambda t/\rho$ are found to be broadly consistent with the predicted linear increase with the age of the patient (Table 2). With an estimated basal cell density of $\rho = 10,000$ cells per mm^2 (38) and a progenitor fraction of basal cells of one in three [extrapolated from mouse (27)], a linear fit of the measured ratio suggests a loss/replacement rate of the self-renewing population of $r\lambda \approx 0.5$ per week. Although uncertainty in both the progenitor fraction and the relative frequency of divisions leading to symmetric or asymmetric fate undermine the predictive value of this rate, a loss/replacement time measured in weeks is broadly consistent with the expected proliferative activity of cycling keratinocyte progenitors in normal homeostasis which, on the basis of BrdU incorporation, points at an average cell division rate in human scalp epidermis of around two per week (38).

Significantly, the deep sequencing approach also provides a quantitative assay to expose rare mutant clones that have undergone field transformation and to assess their mutational

profile. Application of this approach to human epidermis shows that, despite evidence for positive selection (16), population asymmetry and neutrality of epidermal progenitor cell fate may be surprisingly robust to the acquisition of somatic point mutations, even in genes associated with cancer drivers. Indeed, the multiplicity of mutations in the minority of clones (*ca.* 0.1% or less) that lie outside the normal size distribution suggests that proliferative advantage may typically rely on epistasis, requiring the acquisition of multiple mutations across a range of genes.

Under conditions of normal homeostasis, clonal evolution in IFE is constrained to two dimensions, with clones expanding in cohesive clusters across the basal and suprabasal layers (Fig. 1C). Applied to higher-dimensional (volumar) tissues, as well as other epithelial tissues, the same clone size dependence is predicted to apply without further revision (*SI Text*) (35). However, if occupancy of the self-renewing compartment is constrained to lower dimension, or if stem cells are restricted to closed niche domains, the same general technology applies, but the predicted mutant clone size distribution must be appropriately revised (*SI Text*). Therefore, applied to deep sequencing studies, the current theoretical scheme provides a general method to probe stem cell fate behavior in normal cycling adult human tissues and to identify the existence and mutational signature of rare field transformations driven by the nonneutral dynamics of mutant cells, with potential applications to the early detection of preneoplasia.

ACKNOWLEDGMENTS. I thank Peter Campbell, Phil Jones, and Inigo Martincorena for sharing information on the sizes of the biopsies used in their study and for making their sequencing data publically available. I also thank Trevor Graham, Philip Greulich, and Anna Philpott for valuable discussions. This work was supported by Wellcome Trust Grant 098357/Z/12/Z.

1. Simons BD, Clevers H (2011) Strategies for homeostatic stem cell self-renewal in adult tissues. *Cell* 145(6):851–862.
2. Van Keymeulen A, Blanpain C (2012) Tracing epithelial stem cells during development, homeostasis, and repair. *J Cell Biol* 197(5):575–584.
3. Driessens G, Beck B, Caauwe A, Simons BD, Blanpain C (2012) Defining the mode of tumour growth by clonal analysis. *Nature* 488(7412):527–530.
4. Blanpain C (2013) Tracing the cellular origin of cancer. *Nat Cell Biol* 15(2):126–134.
5. Vermeulen L, et al. (2013) Defining stem cell dynamics in models of intestinal tumor initiation. *Science* 342(6161):995–998.
6. Ellenbroek SJ, van Rheeunen J (2014) Imaging hallmarks of cancer in living mice. *Nat Rev Cancer* 14(6):406–418.
7. Merlo LMF, Pepper JW, Reid BJ, Maley CC (2006) Cancer as an evolutionary and ecological process. *Nat Rev Cancer* 6(12):924–935.
8. Bozic I, et al. (2010) Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci USA* 107(43):18545–18550.
9. Navin N, et al. (2011) Tumour evolution inferred by single-cell sequencing. *Nature* 472(7341):90–94.
10. Gerlinger M, et al. (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 366(10):883–892.
11. Greaves M, Maley CC (2012) Clonal evolution in cancer. *Nature* 481(7381):306–313.
12. Vogelstein B, et al. (2013) Cancer genome landscapes. *Science* 339(6127):1546–1558.
13. Sottoriva A, et al. (2015) A Big Bang model of human colorectal tumor growth. *Nat Genet* 47(3):209–216.
14. Martincorena I, Campbell PJ (2015) Somatic mutation in cancer and normal cells. *Science* 349(6255):1483–1489.
15. Stratton MR, Campbell PJ, Futreal PA (2009) The cancer genome. *Nature* 458(7239):719–724.
16. Martincorena I, et al. (2015) Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 348(6237):880–886.
17. Teixeira VH, et al. (2013) Stochastic homeostasis in human airway epithelium is achieved by neutral competition of basal cell progenitors. *eLife* 2:e00966.
18. Baker A-M, et al. (2014) Quantification of crypt and stem cell evolution in the normal and neoplastic human colon. *Cell Reports* 8(4):940–947.
19. Martincorena I, Seshasayee ASN, Luscombe NM (2012) Evidence of non-random mutation rates suggests an evolutionary risk management strategy. *Nature* 485(7396):95–98.
20. Dees ND, et al. (2012) MuSiC: Identifying mutational significance in cancer genomes. *Genome Res* 22(8):1589–1598.
21. Greenman C, Wooster R, Futreal PA, Stratton MR, Easton DF (2006) Statistical analysis of pathogenicity of somatic mutations in cancer. *Genetics* 173(4):2187–2198.
22. Krieger T, Simons BD (2015) Dynamic stem cell heterogeneity. *Development* 142(8):1396–1406.
23. Hsu Y-C, Li L, Fuchs E (2014) Emerging interactions between skin stem cells and their niches. *Nat Med* 20(8):847–856.
24. Page ME, Lombard P, Ng F, Göttgens B, Jensen KB (2013) The epidermis comprises autonomous compartments maintained by distinct stem cell populations. *Cell Stem Cell* 13(4):471–482.
25. Mackenzie IC (1970) Relationship between mitosis and the ordered structure of the stratum corneum in mouse epidermis. *Nature* 226(5246):653–655.
26. Potten CS, Kovacs L, Hamilton E (1974) Continuous labelling studies on mouse skin and intestine. *Cell Tissue Kinet* 7(3):271–283.
27. Clayton E, et al. (2007) A single type of progenitor cell maintains normal epidermis. *Nature* 446(7132):185–189.
28. Doupé DP, Klein AM, Simons BD, Jones PH (2010) The ordered architecture of murine ear epidermis is maintained by progenitor cells with random fate. *Dev Cell* 18(2):317–323.
29. Mascré G, et al. (2012) Distinct contribution of stem and progenitor cells to epidermal maintenance. *Nature* 489(7415):257–262.
30. Lim X, et al. (2013) Interfollicular epidermal stem cells self-renew via autocrine Wnt signaling. *Science* 342(6163):1226–1230.
31. Rheinwald JG, Green H (1975) Serial cultivation of strains of human epidermal keratinocytes: The formation of keratinizing colonies from single cells. *Cell* 6(3):331–343.
32. Barrandon Y, Green H (1987) Three clonal types of keratinocyte with different capacities for multiplication. *Proc Natl Acad Sci USA* 84(8):2302–2306.
33. Jones PH, Watt FM (1993) Separation of human epidermal stem cells from transit amplifying cells on the basis of differences in integrin function and expression. *Cell* 73(4):713–724.
34. Watt FM (2014) Mammalian skin cell biology: At the interface between laboratory and clinic. *Science* 346(6212):937–940.
35. Klein AM, Simons BD (2011) Universal patterns of stem cell fate in cycling adult tissues. *Development* 138(15):3103–3111.
36. Klein AM, Brash DE, Jones PH, Simons BD (2010) Stochastic fate of p53-mutant epidermal progenitor cells is tilted toward proliferation by UV B during preneoplasia. *Proc Natl Acad Sci USA* 107(1):270–275.
37. Slaughter DP, Southwick HW, Smejkal W (1953) Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer* 6(5):963–968.
38. Jones PH, Harper S, Watt FM (1995) Stem cell patterning and fate in human epidermis. *Cell* 80(1):83–93.
39. Bailey NTJ (1964) *The Elements of Stochastic Processes with Applications to the Natural Sciences* (Wiley, New York).
40. Zhang MQ (1998) Statistical features of human exons and their flanking regions. *Hum Mol Genet* 7(5):919–932.