Temporal Dissection of Tumorigenesis in Primary Cancers

Steffen Durinck1,*, Christine Ho2,*, Nicholas J. Wang1,*, Wilson Liao3, Lakshmi R. Jakkula1, Eric A. Colllisson1, Jennifer Pons3, Sai-Wing Chan3, Ernest T. Lam3, Catherine Chu3, Kyunghee Park5, Sung-woo Hong5, Joe S. Hur6, Nam Huh5, Isaac M. Neuhaus3, Siegrit S. Yu7, Roy C. Grekin3, Theodora M. Mauro3, James E. Cleaver3, Pui-Yan Kwok3, Philip E. LeBoit4, Gad Getz2, Kristian Cibulskis7, Jon C. Aster8, Haiyan Huang2, Elizabeth Purdom2, Jian Li9, 10, Lars Bolund9, 10, Sarah T. Arron3, Joe W. Gray1, 11, Paul T. Spellman1,†, and Raymond J. Cho3,†

ABSTRACT

Timely intervention for cancer requires knowledge of its earliest genetic aberrations. Sequencing of tumors and their metastases reveals numerous abnormalities occurring late in progression. A means to temporally order aberrations in a single cancer, rather than inferring them from serially acquired samples, would define changes preceding even clinically evident disease. We integrate DNA sequence and copy number information to reconstruct the order of abnormalities as individual tumors evolve for 2 separate cancer types. We detect vast, unreported expansion of simple mutations sharply demarcated by recombinative loss of the second copy of TP53 in cutaneous squamous cell carcinomas (cSCC) and serous ovarian adenocarcinomas, in the former surpassing 50 mutations per megabase. In cSCCs, we also report diverse secondary mutations in known and novel oncogenic pathways, illustrating how such expanded mutagenesis directly promotes malignant progression. These results reframe paradigms in which TP53 mutation is required later, to bypass senescence induced by driver oncogenes.


© 2011 AACR.

INTRODUCTION

Molecular characterization of human cancers usually profiles a single point in time, yielding a catalog of genomic and epigenetic abnormalities reflecting years of somatic change. Although recent efforts reveal some mutations associated with metastasis and recurrence (1), timing of events early in tumorigenesis remains difficult. Precursor dysplastic lesions may be sampled and compared against invasive malignancies, but in many cancer types, early lesions are not clinically identifiable, nor is it obvious which lesions will actually progress. In addition, the increasingly apparent heterogeneity of human cancers suggests such comparisons will require very large sample sizes to reconstruct progression (2).

Primary cutaneous squamous cell carcinomas (cSCC) rank among the most common human malignancies, with an annual incidence in Caucasians of >150 in 100,000 individuals (3). These tumors arise in anatomic sites and with a demographic proportionate to sunlight exposure and acquire a mutational spectrum reflecting significant UV radiation damage (4). Although many are excised without
complication, cSCCs sometimes behave aggressively, with recurrence and regional spread, especially in immunosuppressed and repair-deficient genetic backgrounds (3). We hypothesized that the long-term mutational stress on these tumors might offer unique insight into the progressive events determining a cancer’s individuality.

RESULTS

Exome-level sequencing of 8 primary cSCCs and matched normal tissue revealed a very large mutation burden of approximately 1,300 somatic single-nucleotide variants per cSCC exome (1 per ~30,000 bp of coding sequence; Supplementary Fig. S1), making cSCCs among the most highly mutated human malignancies. Of mutations assessed by capillary sequencing across our series, 75 of 75 (100%) confirmed the originally detected mutations, including 4 instances of dinucleotide substitution. C > T transition base substitutions at dipyrimidine sites were by far the most common change (>85%), consistent with UV damage. Past analysis of selected TP53 exons suggests that the gene is mutated in 50–90% of cSCCs (5). Our study identified TP53 mutations in 7 of 8 skin cancers, all coinciding with previously reported changes in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (6). Known changes were also found in CDKN2A, encoding the p16/p14*^RB^ bifunctional tumor suppressor, the HRAS small GTPase, and instances of COSMIC mutations not previously described in cSCCs (Table 1; full list of base substitutions provided in Supplementary Table S1). We also detected 42 discrete chromosomal abnormalities, about 5 per sample (Supplementary Table S2).

The ability to temporally order successive molecular changes within an individual tumor, beginning in the initial stages of tumorigenesis, would allow discrimination between mutations forming precancerous lesions from those producing invasive carcinomas. The high prevalence of both simple mutations and copy number abnormalities in cSCCs and ovarian cancers enabled us to reconstruct the evolutionary order of some somatic changes based on the following idea: If a mutation precedes a regional duplication, its copy number is doubled, whereas mutations following a duplication event appear in haploid copy number (7). Therefore, (1) simple mutations preceding a chromosomal duplication event show discretely higher copy numbers compared with those occurring after duplication (Fig. 1); and (2) the ratio of heterozygous to homozygous mutations, in a region of copy-neutral LOH (CN-LOH), directly measures the age of the duplication in evolutionary time (Fig. 2; Supplementary Fig. S2).

We first used this principle to investigate the specific temporal order of mutations in areas of CN-LOH, in which a regional chromosomal duplication replaces the matching portion of the paired chromosome (8, 9). Of the cSCCs in our series, 4 of 8 showed CN-LOH at chromosome 17p, all harboring TP53 mutations reported in COSMIC. Remarkably, all 4 TP53 mutations were present at high allelic abundance in the CN-LOH region, compared with other somatic mutations, indicating that TP53 mutations occurred and were duplicated before other mutations arose (Fig. 1B and C). In aggregate, 59 of 63 mutations in 17p appear after loss of the second TP53 wild-type allele, 15-fold greater than those preceding loss. CN-LOH events at 17p represent 2% of coding sequence and show normalized mutation frequencies reflective of the remainder of the exome (Supplementary Fig. S3). Although studies establish some p53 mutations as gain-of-function with respect to cancer type (10), or biochemically dominant negative, ours is the first to report that the vast majority of simple mutations—tens of thousands genome-wide in the case of cSCCs—appear sharply gated by elimination of the second copy of TP53. We further detect at least partial persistence of active DNA repair, suggesting that a profound loss of damage surveillance contributes to the high number of observed mutations (Supplementary Fig. S1; ref. 11).

Three samples without CN-LOH at 17p show at least 2 distinct TP53 mutations, presumably causing biallelic mutation. In the sample in which TP53 mutations were not detected, a regionally duplicated mutation in the ATM kinase domain was observed, suggesting an alternative means of escaping damage surveillance mechanisms during telomere crisis (12).

We sought to validate our observations in an additional cancer type. Recently, full genomic sequence and copy number changes were determined for 10 ovarian serous adenocarcinomas by The Cancer Genome Atlas Project. Ovarian cancers generally show more complex karyotypic abnormalities than do cSCCs (13). In the 3 samples with a clear, informative CN-LOH event at 17p, we again found solid evidence for complete loss of TP53 as the earliest event (Fig. 1D). These initial events in ovarian tumorigenesis could not have been determined through sequencing of precursor lesions and invasive cancers (1, 14), as the asymptomatic nature of early disease precludes tissue collection.

Integrative analyses of copy number and exome sequence also reveal information about the temporal order of chromosomal abnormalities within an individual cancer (7). As described above, the ratio of heterozygous to homozygous mutations ρ, in a given region of CN-LOH, provides a direct measure of the relative age of the duplication (Fig. 2A). In other words, duplications with higher ρ occur earlier than those with lower ρ. We found that ρ varied widely among regions of CN-LOH (Figs. 2B–D and 3) and could statistically distinguish the temporal order of aberrations within a sample (Fig. 2). Overall, 7 informative duplications co-occurring with 17p CN-LOH all showed a substantially lower relative ρ (Supplementary Table S2) and thus likely occurred after 17p duplication. Therefore, loss of the second TP53 allele appears to precede not only a vast expansion of simple mutations but also the development of chromosomal aberrations. As a general principle, any regional copy gain acquires a heterozygote mutation frequency uniquely reflective of the time of gain. For selected instances in our series, extension of this principle enabled temporal dissection of more complex copy gains (Fig. 3), revealing that these alterations also follow complete TP53 loss. In cSCCs, we found 486 nonsynonymous mutations that were sequenced deeply enough to determine copy number (>50 independent reads) and that fell at least once in a region of CN-LOH. These included known mutations in CDKN2A, WT1, and HRAS (Table 1), each of which showed multiple instances of either wild-type allele loss or biallelic mutation, as seen for TP53. Of interest, this pattern of recurrent biallelic inactivation was also detected at high prevalence for the suspected epithelial tumor suppressors NOTCH1 and
NOTCH2 and the polycystic kidney disease gene PKHD1 (See Supplementary Methods). NOTCH1 shows 3 instances of early truncation, 2 of which show wild-type loss; 1 case of multiple mutation; and 2 other mutations, 1 of which occurs in a splice site (Table 1). NOTCH2 shows multiple mutations in 4 of 8 samples, and 3 of these contain at least 1 truncating mutation.

**DISCUSSION**

We trace the mutational evolution of individual tumors, using a novel, sequence-based assessment strategy and, in doing so, provide a patient-centric complement to more traditional "mutation-by-stage" approaches (2, 14). Our results illuminate key aspects of timing in cancer evolution without requiring large sample series, for which precursor lesions are often inaccessible. TP53 is often mutated in precursor lesions, but paradigms of oncogenesis propose p53 loss as a late requirement, overcoming senescence programs activated by prior activation of driver oncogenes (15, 16) and enabling survival through telomere crisis (17). Furthermore, biallelic TP53 loss occurs frequently, despite evidence that p53 mutants behave dominantly both structurally and functionally with respect to phenotypes such as tumor formation (10, 18).

Our data reveal that decades of UV damage and inactivation of a single TP53 allele result in only about 100 mutations in the epithelial exome. This tenacious genetic stability explains the benign behavior of clonal keratinocyte proliferations, harboring heterozygous TP53 mutation, that commonly form in sun-exposed skin (19; 20). Subsequent elimination of the second TP53 allele, often through recombination, sharply demarcates a vast expansion in simple mutations, in cSCCs reaching 50 per megabase (150,000 per genome) and making them the most mutagenized human cancers known. Because DNA repair remains at least partially active, this vast mutation burden might result from the collabroative effects of ongoing DNA damage (from intrinsic and exogenous insults) coupled with disabled DNA damage-induced apoptosis. The reproduction of this phenomenon in ovarian adenocarcinoma suggests that the vast majority of mutations follow second TP53 allele loss, irrespective of mode of DNA damage or tissue of origin.

Classic studies report that precursor lesions and invasive cancers both carry mutated driver oncogenes, but find TP53 inactivation more frequent in invasive disease (15, 21), suggesting p53 inactivation to be a late event. Activation of a key oncogene prior to biallelic TP53 loss in our series is formally possible, but few coding mutations precede 17p duplication, and none recur in established oncogenes. Although apparently contradictory, these findings could be reconciled by a temporal requirement that TP53 mutation precede driver oncogene mutation in precursor lesions destined to progress to invasive cancer. In this model, precursor lesions that activate oncogenes first (before TP53 inactivation) fail to progress, but would nonetheless be detected in "mutation frequency-by-stage" surveys. Alternatively, different cancer types might exhibit distinct temporal ordering of key mutations. Application of our approach to sequence data from

---

**Table 1. Description of sample cases and selected mutations of interest**

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age</th>
<th>Site</th>
<th>Immune status</th>
<th>TP53</th>
<th>CDKN2A</th>
<th>NOTCH1</th>
<th>NOTCH2</th>
<th>NOTCH3</th>
<th>NOTCH4</th>
<th>Other known COSMIC mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. M</td>
<td>87</td>
<td>Scalp</td>
<td>+</td>
<td>E285K</td>
<td>P1771S, R1595Q</td>
<td>P2265</td>
<td></td>
<td></td>
<td></td>
<td>WT1</td>
</tr>
<tr>
<td>3. M</td>
<td>84</td>
<td>Left dorsal hand</td>
<td>+</td>
<td>E224</td>
<td>C478F</td>
<td></td>
<td>R1333C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. F</td>
<td>61</td>
<td>Left cheek</td>
<td>+</td>
<td>Y220N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. M</td>
<td>83</td>
<td>Left cheek</td>
<td>+</td>
<td>H179Y,</td>
<td>W1769*</td>
<td>Q1634*, T2278I</td>
<td></td>
<td></td>
<td></td>
<td>S1602F</td>
</tr>
<tr>
<td>7. M</td>
<td>58</td>
<td>Left aural helix</td>
<td>-</td>
<td>E286K, T329I, E349*</td>
<td>Q1924*, Q1616*, G480D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HRAS</td>
</tr>
<tr>
<td>8. M</td>
<td>63</td>
<td>Lower lip</td>
<td>+</td>
<td>None*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HSPB2</td>
</tr>
</tbody>
</table>

**NOTE:** Boldface denotes loss of wild-type allele; * denotes nonsense mutations; shaded mutations are found in the COSMIC database.

In sample 8, no mutation in TP53 was seen, but early, duplicated mutation was detected in ATM at G2024R, in the PI kinase-related domain. Given their sheer number, some TP53 mutations may act in true dominant negative fashion but were not detected in our series. COSMIC Notch family mutations are activating by prior activation of driver oncogenes (15, 16) and enabling survival through telomere crisis (17). Furthermore, biallelic TP53 loss occurs frequently, despite evidence that p53 mutants behave dominantly both structurally and functionally with respect to phenotypes such as tumor formation (10, 18).
Figure 1. Loss of second TP53 wild-type allele precedes simple mutations in cSCCs and ovarian cancers. A, schematic for acquisition of discretely higher copy number for mutations preceding a regional duplication, enabling temporal ordering of mutations during tumorigenesis. For representative samples, mutant allele frequency (y-axis) is shown plotted against physical position on chromosome 17 (x-axis) for cSCCs (B and C) and ovarian cancers (D). More than 50 independent reads were required for inclusion of a mutation. Estimated allele frequencies are shown as lines for heterozygous (solid) and homozygous (dotted) mutations. Regions without CN-LOH in ovarian cancers often show complex ploidy, generating greater variance in mutation frequency; these simple frequency estimates were therefore not applied. Mutation frequencies are centered below 1.0 for homozygotes and 0.5 for heterozygotes because of the effects of partial dilution with nontumorous cells (see Supplementary Methods for more details).
**Temporal Dissection of Tumorigenesis in Primary Cancers**

**Figure 2.** Conceptual framework for temporal ordering of chromosomal aberrations. **A**, schematic for differential mutant allele frequencies in areas of CN-LOH, based on relative age of chromosomal event during tumorigenesis. Earlier events allow greater time for accumulation of heterozygous mutations, whereas later events are predominated by homozygous mutations. Three events are ordered in a single sample, with the earliest on chromosome 17 (**B**), later on chromosome 2 (**C**), and last on chromosome 14 (**D**). Relative mutant allele frequency (y-axis) is plotted against chromosomal physical position (x-axis) for cSCCs. To the right of each panel, a histogram of mutant allele frequencies for each CN-LOH event shows density (y-axis) plotted against allele frequency (x-axis). A modified binomial distribution function (see Supplementary Methods) distinguishes the proportion of heterozygote and homozygote mutant frequencies, ordering duplications in chromosome 17 (0.00–0.12), chromosome 2 (0.17–0.45), and chromosome 14 (0.65–0.94).
other cancer types, such as colon adenocarcinoma, should help distinguish these possibilities.

In selected instances, we are able to show mutant TP53 duplication occurring before dosage changes in mutant alleles, such as for CDKN2A and WT1 (Fig. 3). The consequences of such expanded mutagenesis and chromosomal instability emerge dramatically in the Notch signaling pathway, in which multiple family members develop mutations and wild-type alleles are lost frequently. Constitutive Notch protein activation drives subsets of acute leukemias (22), but a clear tumor suppressor phenotype has also been established in keratinocytes (23), with attenuated expression producing proliferation and invasive morphologies. Further study should clarify the breadth of epithelial cancers harboring these somatic changes, as well as their specific functional effects. Our data also confirm low-prevalence activation of known oncogenes such as Ras in human cSCCs, raising the possibility that numerous mutations in other pathways may serve as the functional oncogene in this setting (24).

Taken together, these insights imply that targeting activated oncogenes (e.g., those with small molecule inhibitors) fails to address a fundamental, detectable abnormality in cancer genomes that accelerates evolution toward clinical resistance. Temporal dissection of tumorigenesis provides early, assayable diagnostic markers and illuminates the specific biological consequences of these aberrations. We show the utility of this method for the CN-LOH and copy gains that constitute most chromosomal aberrations. Because many cancer types carry rearrangement of substantial proportions of the genome, especially those spanning key oncogenes, extension of this method should rapidly reveal additional ordered events. The described reconstruction of genomic aberration history can be applied immediately to any cancer for which sequence data and copy number are available.

**METHODS**

We obtained 8 matched cSCCs and normal tissue samples as part of a skin cancer study protocol, with all subjects providing informed consent according to procedures approved by the University of California, San Francisco (UCSF), Committee on Human Research, San Francisco, California. Diagnosis of cSCCs was confirmed for all tumors via histologic examination of a standard biopsy specimen by a board-certified dermatopathologist. DNA was extracted from tumor and control samples, and allele-specific copy number analysis was performed using Affymetrix Genome-Wide Human SNP Array 6.0 chips. Approximately 40 megabases of coding region were isolated from each sample, using oligonucleotide-based hybrid capture and sequenced with the Illumina sequencing-by-synthesis platform (Supplementary Fig. S4). Mutation detection was performed as previously described (see Supplementary Methods for detail). Seventy-five mutations were independently validated using Sanger sequencing; 100% confirmed the originally identified somatic change. Sequences for all nonsynonymous mutations will be deposited in the database of Genotypes and Phenotypes. Patient information and genomic profiling for ovarian serous adenocarcinomas analyzed in this study have been described previously (25).

Chromosomal regions with aberrant copy number, including CN-LOH and simple copy gains and losses, were identified based on discrete shifts in single nucleotide polymorphism (SNP) copy number from both SNP array and exome sequencing data. The type of aberration was further confirmed by assessing raw copy number depth from SNP array data. Mutations were called after alignment of sequence reads to a reference genome (NCBI36). The fraction of chromosomal copies carrying a mutation was estimated as the fraction of all independent sequence reads containing
Temporal Dissection of Tumorigenesis in Primary Cancers

that mutation. (A detailed description of patient consent, methods, and reagents used for tissue acquisition, genomic profiling, and statistical analysis of mutational evolution is provided in the Supplementary Methods.)

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Allan Balmain, Boris Bastian, Jeffrey Cheng, Douglas Brash, and Dennis Oh for early support and helpful discussions; and Henrik Bengtsson, Pierre Neuvial, Hubert Stoppler, Matthew Akana, Connie Ha, Lauren Lee, Annie Poon, and Eric Dybko for technical assistance.

Grant Support

W. Liao is supported by a Dermatology Foundation Poirot's Career Development Award and the National Institute of Arthritis, Musculoskeletal, and Skin Diseases (K08AR057763); E.A. Collisson by NIH/NCI K08 CA137153; J.E. Cleaver by the University of California Cancer Coordinating Committee; S.T. Arroyo by NIH/National Center for Research Resources/OD UCSF–Clinical & Translational Science Institute Grant KL2 RR024410, a Canary Foundation/American Cancer Society Postdoctoral Fellowship for the Early Detection of Cancer, and a Dermatology Foundation Career Development Award in Dermatologic Surgery; and R.J. Cho by a Dermatology Foundation Career Development Award and as a Samsung Biotechnology Scholar-in-Residence.

This research was supported under J.W. Gray [by the Director, Office of Science, Office of Biological and Environmental Research, U.S. Department of Energy, under Contract DE-AC02-05CH11231]; by NIH, National Cancer Institute (NCI) Grants P50 CA 58207, U54 CA 112970, and NHGRI U24 CA 126551; by the Department of the Army, Award W81XWH-07-1-0663 (The U.S. Army Medical Research Acquisition Activity, Fort Detrick, Maryland, is the awarding agency); and by the Stand Up To Cancer–American Association for Cancer Research Dream Team Translational Cancer Research Grant SU2C-AACR-DT0409. The content of this information does not necessarily reflect the position or the policy of the federal government, and no official endorsement should be inferred]; P.T. Spellman (by NIH/NCI U24 CA1437991); R.J. Cho (by an unrestricted gift grant from the Samsung Advanced Institute of Technology); T.M. Mauro (by NIH Grants AR051930 and R01AG028492, and the Medical Research Service, Department of Veterans Affairs); and A. Balmain (by NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases Program Project Grant 5-P01-AR050440-05).

Received March 7, 2011; revised May 12, 2011; accepted May 17, 2011; published OnlineFirst June 29, 2011.

REFERENCES

Temporal Dissection of Tumorigenesis in Primary Cancers

Steffen Durinck, Christine Ho, Nicholas J. Wang, et al.

Cancer Discovery 2011;1:137-143. Published OnlineFirst June 29, 2011.

Updated version
Access the most recent version of this article at:
doi:10.1158/2159-8290.CD-11-0028

Supplementary Material
Access the most recent supplemental material at:
http://cancerdiscovery.aacrjournals.org/content/suppl/2011/06/29/2159-8290.CD-11-0028.DC1

Cited articles
This article cites 25 articles, 9 of which you can access for free at:
http://cancerdiscovery.aacrjournals.org/content/1/2/137.full#ref-list-1

Citing articles
This article has been cited by 15 HighWire-hosted articles. Access the articles at:
http://cancerdiscovery.aacrjournals.org/content/1/2/137.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.