Comparative oncogenomics identifies breast tumors enriched in functional tumor-initiating cells

Jason I. Herschkowitz\textsuperscript{a}, Wei Zhao\textsuperscript{b,c}, Mei Zhang\textsuperscript{d}, Jerry Usary\textsuperscript{c,d}, George Murrov\textsuperscript{d}, David Edwards\textsuperscript{b}, Jana Knezevic\textsuperscript{c}, Stephanie B. Greene\textsuperscript{e}, David Darr\textsuperscript{c,d}, Melissa A. Troester\textsuperscript{c}, Susan G. Hilsenbeck\textsuperscript{e}, Daniel Medina\textsuperscript{a}, Charles M. Perou\textsuperscript{c,d,f}, and Jeffrey M. Rosen\textsuperscript{a,1}

\textsuperscript{a}Department of Molecular and Cellular Biology and \textsuperscript{b}Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, TX 77030; and \textsuperscript{c}Curriculum in Bioinformatics and Computational Biology, \textsuperscript{d}Lineberger Comprehensive Cancer Center, \textsuperscript{e}Department of Genetics, and \textsuperscript{f}Department of Pathology and Laboratory Medicine, University of North Carolina, Chapel Hill, NC 27599

Edited by Kornelia Polyak, Dana-Farber Cancer Institute, Boston, MA, and accepted by the Editorial Board April 5, 2011 (received for review January 26, 2011)

The claudin-low subtype is a recently identified rare molecular subtype of human breast cancer that expresses low levels of tight and adherens junction genes and shows high expression of epithelial-to-mesenchymal transition (EMT) genes. These tumors are enriched in gene expression signatures derived from human tumor-initiating cells (TICs) and human mammary stem cells. Through cross-species analysis, we discovered mouse mammary tumors that have similar gene expression characteristics as human claudin-low tumors and were also enriched for the human TIC signature. Such claudin-low tumors were similarly rare but came from a number of distinct mouse models, including the p53 null transplant model. Here we present a molecular characterization of 50 p53 null mammary tumors compared with other mouse models and human breast tumor subtypes. Similar to human tumors, the murine p53 null tumors fell into multiple molecular subtypes, including two basal-like, a luminal, a claudin-low, and a subtype unique to this model. The claudin-low tumors also showed high gene expression of EMT inducers, low expression of the miR-200 family, and low to absent tight and adherens junction genes and shows high expression of epithelial-to-mesenchymal transition (EMT) genes. These tumors are enriched in gene expression signatures derived from human tumor-initiating cells (TICs) and human mammary stem cells. Through cross-species analysis, we discovered mouse mammary tumors that have similar gene expression characteristics as human claudin-low tumors and were also enriched for the human TIC signature. Such claudin-low tumors were similarly rare but came from a number of distinct mouse models, including the p53 null transplant model. Here we present a molecular characterization of 50 p53 null mammary tumors compared with other mouse models and human breast tumor subtypes. Similar to human tumors, the murine p53 null tumors fell into multiple molecular subtypes, including two basal-like, a luminal, a claudin-low, and a subtype unique to this model. The claudin-low tumors also showed high gene expression of EMT inducers, low expression of the miR-200 family, and low to absent expression of both claudin 3 and E-cadherin. These murine subtypes also contained distinct genomic DNA copy number changes, some of which are similarly altered in their cognate human subtype counterpart. Finally, limiting dilution transplantation revealed that p53 null claudin-low tumors are highly enriched for TICs compared with the more common adenocarcinomas arising in the same model, thus providing a unique preclinical mouse model to investigate the therapeutic response of TICs.

Results

p53 Null Tumors Show Variable Histology. Previously, we hypothesized that the heterogeneity observed in human BC might arise not only because of activation of different oncogenes or loss of specific tumor suppressor genes, but might also be dependent on the cell of origin in which these genetic changes occur (11). Initially transplanted p53 null mammary epithelial cells gave rise to phenotypically normal ductal outgrowths, which then stochastically developed mammary tumors. We therefore hypothesized that the deletion of this single tumor suppressor gene might give rise to a spectrum of heterogeneous tumors, depending on the cell of origin in which additional stochastic changes occurred. To test this hypothesis, we collected 50 p53 null tumors that arose in wild-type BALB/c mice after transplantation of p53 null BALB/c mammary tissue into the cleared fat pads of 3-wk-old mice (6).
Like some other genetically engineered mouse model (GEMM) mammary tumor models, the p53 null model gave rise to tumors with a diversity of histological phenotypes (Fig. S1 and Dataset S1) (9, 12). Approximately 10% of the tumors contained a majority of spindle-shaped cells, a histology originally described for carcinosarcomas, now called EMT tumors (9).

**p53 Null Tumors Cluster into Distinct Tumor Subtypes.** In a previous study we profiled 13 distinct mouse models, including the p53 null model (13). However, with only five p53 null tumor samples, we were not able to appreciate the full spectrum of molecular heterogeneity represented in this mouse model. Now, with a total of 50 tumors from the p53 null model, we see that these tumors cluster into five distinct tumor subtypes when performing hierarchical clustering analysis using our previously defined mouse intrinsic gene list (13) (Fig. 1); furthermore, we used SigClust (14) to assess the significance of this clustering and objectively determined that the p53 null model did populate multiple statistically significant groups/subtypes, as follows.

**Basal-like.** Two groups of basal-like mouse mammary tumors were observed (Fig. 1); in part, we define these groups as basal-like according to their high expression of known basal markers, including keratin 5 (K5), ID4, and TRI29 (Fig. 1D) and selective high expression of the human basal-like tumor expression cluster (Fig. S2). Basal 1 tumors (5 of 50, 10%) clustered along with a group of other mouse basal-like tumors, including BRCA1-deficient and MMTV-Wnt1 tumors. Basal 2 tumors (8 of 50, 16%) clustered next to the Basal 1 tumors but showed a higher expression of the murine luminal cluster than did Basal 1 (Fig. 1C). Basal 1 p53 null tumors showed an increased proliferation signature separating them from Basal 2 and the other p53 null subtypes (Fig. 1G and Fig. S2), and they also showed high p16 expression, which is a hallmark of impaired RB1 function (15). Basal tumors (eight of nine) tested stained positively for K5, as expected (Fig. S3 and Dataset S1); however, paradoxically, five of eight tested stained positively for the ER, of which four of five were of the Basal 2 subtype.

**Luminal.** Eight of 50 p53 null tumors (16%) clustered close together and with the mouse luminal models MMTV-Neu and MMTV-PyMT. As we have seen for other luminal models, these tumors express luminal-specific genes like XBP1 but are missing ER and estrogen responsive genes; accordingly, only one of eight of the luminal tumors stained positively for ER. Interestingly, like human luminal tumors, p53 null luminal tumors showed lower levels of p18(INK4C), and p18 null mice develop predominantly luminal-type mammary tumors (16).

**Claudin-low.** Five of 50 p53 null tumors (10%) showed the murine claudin-low expression phenotype (Fig. 1F) and significantly clustered with the previously defined murine claudin-low tumors. These tumors had an EMT tumor histology (Fig. S1) and showed expression of the human claudin-low signature (Fig. S2). In agreement with the gene expression data and immunohistochemistry on human samples (17), we observed low to absent expression of p18(INK4C) and selective expression of the murine claudin-low signature.

**Fig. 1.** Intrinsic gene set clustering analysis of 50 p53 null tumors and 117 samples from 13 GEMM previously published in Herschkowitz et al. (13). (A) Overview of the complete cluster diagram. (B) Experimental sample-associated dendrogram. Boxes indicate the p53 null tumor subtypes based on SigClust analysis. (C) Luminal epithelial expression pattern that is highly expressed in luminal p53 null tumors, MMTV-Neu, and MMTV-PyMT tumors (D). Basal epithelial expression pattern including K5 and ID4, which are highly expressed in basal-like p53 null tumors. (E) Mesenchymal genes, including snail homolog 1. (F) Genes expressed at low levels in claudin-low tumors, including CLDN3, CLDN7, and ELF5. (G) Proliferation signature. (H) Individual genes discussed within the text.
miRNAs. Because expression of a number of specific miRNAs has been associated previously with an EMT transition (18, 19), we took a candidate approach to identify miRNAs that were differentially expressed between p53 null claudin-low tumors and the other subtypes. First we evaluated the miR-200 family of miRNAs and miR-205, which are miRNAs that have been implicated in EMT and TICs (18, 20). Although a number of targets for these miRNAs have now been identified, important targets with respect to EMT are ZEB1 and ZEB2. ZEB1 and ZEB2 are expressed at high levels in claudin-low tumors (mouse and human), and as expected, these miRNAs were present at very low levels relative to the other p53 null tumors (Fig. 2B). Another cluster of miRNAs expressed at lower levels in both cancer and normal mammary stem cells contains miRNAs 182, 96, and 183 (20). Likewise, this cluster of miRNAs was expressed at low levels in murine p53 null claudin-low tumors. Additionally, miR-203, another stemness-inhibiting miRNA regulated by Zeb1 (21), was also expressed at low levels in claudin-low tumors. Marked decreases, however, were not seen for all miRNAs tested (e.g., miR-21).

It has been shown that human breast tumor subtype correlates with miRNA profiles (22, 23). We reanalyzed the Bilenkiron et al. dataset (22) to determine which tumors contained the claudin-low gene expression pattern using the Prat claudin-low predictor (17). Using a supervised analysis, 17 miRNAs were identified that were significantly differentially expressed between claudin-low tumors vs. the other BCs (Dataset S2). This included seven of the miRNAs that we had observed, including miR-200a, 200b, 200c, 149, 182, 183, and 203. These results indicate that in addition to miRNA gene expression changes, mouse and human claudin-low tumors share common miRNA expression patterns.

**p53 Null Tumor Subtypes Display Distinct Copy Number Alterations.** Presumably stochastic genetic alterations selected during neoplastic progression collaborate with the loss of p53. It is also likely that different genetic events can cause tumors to show a given phenotype, or only sensitize one particular cell type to malignant transformation; thus, specific copy number and/or mutations may be highly enriched within a specific tumor subtype, as shown for human breast tumors (24). To investigate this on the genomic level, we performed aCGH on 44 p53 null tumors using Agilent 244,000 feature DNA microarrays to determine whether there were subtype-specific copy number alterations (CNAs) (Fig. 3). In comparison with many mouse models (25, 26) the p53 null mammary tumors contained a fair amount of genomic instability. Interestingly, all five tumor subtypes contain distinct CNAs, which are listed by subtype in Dataset S3. In the p53 null basal-like tumors (both Basal 1 and Basal 2 considered together), there was loss of the distal half of chromosome 8, including INPP4B, which has now been shown to be selectively lost in human basal-like/triple-negative tumors [4q31.22–q35.2(12)] (27, 28). p53 null luminal tumors showed loss of chromosome 4 and gain of chromosome 7. The p53 null unique subtype showed very few subtype-specific events; however, when converted to human genomic coordinates, these events identify amplification of human chromosome 17q12–q21.2(2), which is a common amplicon that is distal to the HER2 amplicon. Interestingly, one of these murine tumors (2304L) that clusters in the p53 null unique subtype, but that is not contained within the SigChust defined group (Fig. 1B), showed high Her2 mRNA and protein expression and was amplified for Her2 on mouse chromosome 11 (Fig. S4); thus, the p53 null model is even able to generate HER2-amplified tumors, albeit at a low frequency.

The copy number landscape of human claudin-low tumors is not known, but the p53 null claudin-low tumors showed numerous subtype-enriched CNAs. These changes included the near-complete loss of mouse chromosome 1 and frequent but smaller losses on 7, 10, and 13.
Driving mutations/changes present in each region. Nonetheless, the possible existence of driver mutations/changes.

Fig. 3. Tumor genomic DNA copy number landscape plots for mouse p53 null tumor classes. At top is the overall pattern for all 34 tumors considered together, and then below are the landscape plots for each of the five expression-defined subtypes. Gray shading indicates the overall frequency of aberrations seen in that subtype, and the black shading indicates the group-specific CNA (P value threshold 0.05).

12, and 14. There were also specific gains on 3, 8, and 13. Many of these map to common regions of copy number changes in human BC; however, additional studies will be required to define the driving mutations/changes present in each region. Nonetheless, these claudin-low subtype-specific copy number changes do suggest the possible existence of driver mutations/changes.

The work of Bergamaschi et al. (24) identified numerous CNAs associated with some of the intrinsic subtypes. We, therefore, converted our mouse CNAs into human equivalent chromosome locations and determined that a number of significantly altered regions were in common between p53 null mouse tumors and human breast tumors (Dataset S4). Of note were the loss of two regions that occurred in both mouse and human basal-like tumors, human 4q31.22-q35.2(12) that contains INPP4B, and human 14q22.1-q23.1(4); the somatic loss of these two regions across species suggests that each contains a tumor suppressor(s) gene and that the loss of these genes may sensitize cells to become the basal-like subtype, similar to germline inactivation of BRCA1 (29).

p53 Null Claudin-Low Tumors Are Enriched for Tumor-Initiating Cells.

Similar to their utility in the isolation of mouse mammary stem cells (30), CD29 and CD24 have been used as markers that enrich for TICs in the p53 null tumor model, with the CD29+/CD24− fraction showing the TIC capabilities (12). Furthermore, an EMT program has been shown to correlate with stem-like properties, and the loss of miR-200 expression as well as a “claudin-low” signature has been suggested to characterize both normal and cancer stem cells (20). By FACS analysis, in the p53 null claudin-low tumors tested, the percentage of double-positive cells was 70–85%, as compared with a maximum of 14% in the other p53 null tumors analyzed (Fig. A and B) Interestingly, some luminal tumors exhibit very low levels of double-positive cells. This was suggestive, therefore, that there might be a high percentage of TICs in the claudin-low tumors.

To test this hypothesis, two different claudin-low p53 null tumors were FACS sorted for all four possible populations using CD29+ and CD24+, and limiting dilution transplantation was performed (Tables 1 and 2). The tumor-initiating frequency was similar between the CD29+/CD24+ and CD29+/CD24− fractions, and these two populations were highly enriched for TICs compared with the other two fractions. In addition, by transplanting FACS-sorted lineage-negative cells in limiting dilution, we determined that the tumor repopulating ability of these claudin-low phenotype tumors was >38 times greater than that of two other p53 null adenocarcinomas (T1 and T7) performed using the same methods (Fig. 4C) (31). Thus, these data indicate that an expanded population of TICs exists within these murine claudin-low tumors.

Discussion

GEMM have provided a rich resource for the study of different cancers; however, many individual models show significant molecular and histological heterogeneity (13). This heterogeneity complicates studies because multiple disease types may actually be present within a given model. One way to address this heterogeneity is to genomically characterize each tumor, then group tumors together according to important features and, most importantly, perform functional studies. The p53 null mammary transplant model is one such heterogeneous model, and we have taken advantage of this feature and identified transplantable lines that represent at least three human breast tumor subtypes. In addition, because all these tumors develop subsequent to the same initial loss of p53, the question is whether this heterogeneity is due to different collaborating oncogenes/tumor suppressors and/or different cells of origins. The cell type of origin in cancer is a highly debated topic (reviewed recently in ref. 32). Although specific genetic lesions clearly play a major role in determining the tumor phenotype, growing evidence indicates that cancer of different subtypes within an organ system may also reflect distinct cells of origin. However, it is not apparent whether a given oncogene lesion actually dictates the cell of origin or, conversely, whether the cell of origin determines which oncogenic lesions can occur. Both of these possibilities most likely exist. There is evidence that tumors generated using the same oncogene targeted to different cell lineages can be phenotypically distinct (33). Recent studies have shown that BRCA1 mutant and basal-like human tumors are enriched in gene expression profiles and surface markers of luminal progenitors (34). Similarly, inactivation of BRCA1 (and p53) in the luminal or basal cell population of the mouse mammary gland showed that only the luminal cells gave rise to tumors histologically resembling those of BRCA1 mutation carriers (35). These results, however, fall short of actually proving that these tumor types originated in these cell types. Mouse models, like the heterogeneous one presented here, can provide an invaluable tool with which to decipher the cell of origin when genetics is combined with precise lineage tracing. At present we cannot definitively answer the cell-of-origin question because the necessary reagents are not yet available to perform the lineage tracing experiments, as has been done recently using mouse models of intestinal cancer (36). However, our experiments do provide several important insights: first, tumors of the basal-like, luminal, and claudin-low phenotypes clearly arise, although at different frequencies and with a predisposition for basal-like; in particular, the Basal 1 group seems to
most faithfully recapitulate human basal-like tumors, in that it shows high expression of basal gene expression features, of the proliferation signature, and of p16 (a hallmark of RB1 loss), all of which are features of human basal-like tumors (15). Second, the luminal tumors that do arise are largely ER negative (as are the vast majority of murine tumors from other GEMMs) and therefore more similar to a luminal B than the ER-positive luminal A human subtype.

Interestingly, claudin-low p53 null tumors were also seen, although at the lowest frequency (five total). As was shown for human claudin-low tumors and cell lines, these murine tumors lack tight junction proteins, including claudin 3 and E-cadherin, and show expression features of mesenchymal cells, normal mammary stem cells, and TIC. In addition to previously defined subtypes, we also identified a phenotype unique to this model and noted that nearly 20% of the tumors were scattered throughout the cluster, indicating even greater heterogeneity within this model. For example, tumor 2304L showed clear amplification and high expression of HER2, thus even somatically HER2 amplified tumors occur within this model.

The presence of specific CNAs in different subtypes of tumors arising in the p53 null model suggests that different gains and losses are important for tumor progression subsequent to p53 loss, and these are possibly occurring within different cell types. In the p53 null basal-like tumors, there was specific loss of the distal half of chromosome 8, which is in conserved synteny with human chromosome 4. Recently, loss of this region has been seen specifically with human basal-like/triple-negative BCs, and it is thought that the target of this loss is INPP4B. This gene is selectively lost in human and murine basal-like tumors, thus suggesting that this approach of finding common regions of loss/gains across species can identify putative important tumor and/or subtype causative events. Interestingly, p53 null luminal tumors showed loss of chromosome 4. Chromosome 4 deletions and loss of heterozygosity have been reported in other luminal mouse models, including MMTV-Neu, MMTV-Myc, and MMTV-Ras (26, 37–39). Presumably there exist multiple luminal-specific tumor suppressor genes on chromosome 4. Although other subtypes showed gain of chromosome 1, p53 null claudin-low tumors showed large regions of loss on chromosome 1, which again hints at their uniqueness.

Several lines of evidence have suggested that claudin-low tumors are enriched in functional TICs, predominantly coming from expression analyses (Fig. 2A and Fig. S2). However, because of their rarity and limitations in procurement of primary human claudin-low tumors, this hypothesis has not been tested functionally using human clinical samples. We have, however, herein identified a counterpart of human claudin-low tumors in the mouse. Accordingly, we have taken advantage of this mouse model to test by limiting dilution, the gold standard functional stem cell assay, whether these tumors are enriched in TICs compared with other tumors arising in the same model. With the p53 null model we also have the advantage of being able to transplant these tumors into syngeneic mice with an appropriate microenvironment complete with normal immune function. We showed that the claudin-low murine tumors were significantly more enriched for surface markers of TICs as well as functional TICs compared with other p53 null tumors. Recent studies have shown that minority subsets of tumors from MMTV-Myc and MMTV-MET tumors cluster with our claudin-low mouse tumors (40, 41). It has not been determined whether they too are enriched in functional TICs. However, the MMTV-Myc EMT-like/claudin-low tumors were reported to show an increase in metastasis.

The murine claudin-low tumors show large percentages of CD29+/CD24− cells, MaSC-like mRNA, and miRNA expression profiles, as well as expression of other markers of MaSCs (e.g., high s-SHIP expression) (42). Therefore, it is conceivable that these tumors might have arisen from the MaSC population. Alternatively, they may have resulted from dedifferentiation of a progenitor or even a more differentiated cell. Lineage tracing experiments will be required to definitively resolve this issue.

To effectively target cancer stem cells or TICs, one pressing need is a genetically defined and renewable preclinical model to identify and test new stem cell targeted therapies. To address this need, we now have identified a mouse model that develops claudin-low tumors, in which the bulk of the tumors cells seem to be TICs. This is an example of a spontaneously occurring breast tumor with a high proportion of TICs. Thus, we now have appropriate and validated models for the investigation of important signaling pathways and

### Table 1. Limiting dilution transplantation of adenocarcinomas (T1 and T7)

<table>
<thead>
<tr>
<th>Cells injected</th>
<th>5,000</th>
<th>1,500</th>
<th>1,000</th>
<th>100</th>
<th>50</th>
<th>25</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lin−CD29+CD24+</td>
<td>4/5</td>
<td>2/8</td>
<td>2/7</td>
<td>0/8</td>
<td>0/6</td>
<td>0/2</td>
<td>0/6</td>
</tr>
<tr>
<td>Lin−CD29+CD24+</td>
<td>2/6</td>
<td>2/8</td>
<td>0/7</td>
<td>0/6</td>
<td>0/2</td>
<td>0/6</td>
<td></td>
</tr>
<tr>
<td>Lin−</td>
<td>2/3</td>
<td>6/10</td>
<td>4/9</td>
<td>2/12</td>
<td>1/10</td>
<td>0/4</td>
<td></td>
</tr>
</tbody>
</table>

Data from Zhang et al. (31).

### Table 2. Limiting dilution transplantation of claudin-low tumors (T11 and 2247R)

<table>
<thead>
<tr>
<th>Cells injected</th>
<th>500</th>
<th>250</th>
<th>100</th>
<th>50</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lin−CD29+CD24+</td>
<td>2/2</td>
<td>2/2</td>
<td>5/6</td>
<td>5/6</td>
<td>3/6</td>
</tr>
<tr>
<td>Lin−CD29+CD24L</td>
<td>1/2</td>
<td>2/2</td>
<td>5/6</td>
<td>5/6</td>
<td>3/4</td>
</tr>
<tr>
<td>Lin−CD29+CD24H</td>
<td>2/2</td>
<td>2/4</td>
<td>0/4</td>
<td>2/6</td>
<td>0/2</td>
</tr>
</tbody>
</table>

Fig. 4. p53 null claudin-low tumors express markers of stem cells and are enriched for tumor-initiating ability. (A) Claudin-low tumors have high percentages of double-positive (CD29+/CD24+) cells compared with (B) other p53 null tumors. (C) Limiting dilution transplantation of claudin-low vs. adenocarcinoma cells. Sample sizes are implied by the sizes of the circles (area is proportional to sample size).
therapeutics. Because of their transplantability into syngeneic hosts, this panel of tumors provides a valuable resource for preclinical testing of novel therapeutics. These tumors should serve as excellent models for both the general study of BC stem cells and preclinical models for testing stem cell targeted agents, enabling translation into the clinic. Finally, the finding that claudin-low tumors have an enrichment of functional TICs challenges the popular paradigm that TICs always need to be a rare subpopulation (43).

Materials and Methods

Mice. All animal protocols were reviewed and approved by the Animal Protocol Review Committee at Baylor College of Medicine and University of North Carolina, Chapel Hill.

Preparation of Single Cells. Tumors were processed and digested into single cells as previously described (12). The cells were resuspended in HBSS (Invitrogen) containing 2% FBS and 10 mM Hepes buffer (Invitrogen) before labeling with antibodies.

Flow Cytometry. Cells were labeled with antibodies (Dataset S5) at a concentration of 10 × 10^6 cells/mL under optimized conditions and were subjected to FACS analysis and sorting on an ARIA II sorter (BD Biosciences). Data analysis was performed using FlowJo (v9.1).

Transplantation. Cleavage of mammary epithelial cells and transplantation procedures were performed as previously described (44). After FACS, the designated number of cells were washed once with PBS and transplanted into the cleared fat pads of 21-d-old female BALB/c mice (Harlan).

Microarray Analysis. Total RNA was collected from 45 murine tumors and purified using the Qiagen RNeasy Mini Kit using ≥25 mg tissue. Two micrograms of total RNA was reverse transcribed, amplified, and labeled with Cy5 using a Low RNA Input Amplification kit (Agilent). The common reference RNA (45) was reverse transcribed, amplified, and labeled with Cy3. They were then cohybridized overnight to Agilent Mouse Oligo 44K Microarrays. Finally, they were washed and scanned on an Agilent scanner (G2505B) and uploaded into the database, where a Lowess normalization is automatically performed. The genes for all analyses were filtered by requiring the Lowess normalized intensity values in both channels to be >10. The log2 ratio of Cy5/Cy3 was then reported for each gene. In the final dataset, only genes that reported values in 70% or more of the samples were included.

ACKNOWLEDGMENTS. We thank the Cytometry and Cell Sorting Core Facility at Baylor College of Medicine for flow cytometry assistance. This work was supported by National Cancer Institute (NCI) Grant Multi-P1 RO1CA148761 (to J.M.R. and C.M.P.), NCI work contract N01-CN43308 (to C.M.P.), the Breast Research Foundation (C.M.P.), and Komen Postdoctoral Fellowship PDF0707744 (to J.J.H.).
Supporting Information

Herschkowitz et al. 10.1073/pnas.1018862108

SI Materials and Methods

Immunostaining. Paraffin-embedded sections (5 μm thick) were processed using standard immunostaining methods. Briefly, slides were deparaffinized and hydrated through a series of xylenes and graded ethanol steps. Heat-mediated epitope retrieval was performed in boiling citrate buffer (pH 6.0) for 15 min, then samples cooled to room temperature for 30 min. Secondary antibodies for immunofluorescence were conjugated with Alexa Fluor 488 or -594 fluorophores (1:200; Molecular Probes, Invitrogen). Immunofluorescent samples were mounted with VectaShield Hardset with DAPI mounting media (Vector Laboratories).

Real-Time PCR. Total RNA was prepared from tumors using the miRNA Kit (Qiagen). cDNA was synthesized from 10 ng of total RNA using the TaqMan MiRNA Reverse Transcription Kit with miRNA-specific RT primers (Applied Biosystems). miRNA levels were then measured using the miRNA-specific TaqMan probe provided in the MicroRNA Assays and the TaqMan Gene Expression Assay (Applied Biosystems). miRNA levels were normalized to snoRNA55 and U6 (Applied Biosystems). Student’s t test was used to compare claudin-low vs. the rest.

Microarray Platform Correction. Previously published data on 22K arrays can be found under accession no. GSE3165 in the Gene Expression Omnibus database. Platform correction (i.e., 44K vs. 22K arrays) was performed by making a systematic, gene-by-gene correction based on similar samples across platforms. For both 22K and 44K arrays, six to eight tumors from MMTV-Neu and C3(1)−/− and two pairs of BALB/c p53+/- were assayed on each platform, and for each gene on each platform a median expression ratio was determined. The assumption is then made that the median expression ratio on each platform should be the same, so then an adjustment factor is determined for each gene using these similar tumor samples across platforms. Next, all samples on the 44K platform were adjusted using this factor; note that the data shown in Fig. 1 (main text) contain the previous 122 arrays, the new 45 p53 null samples/arrays, and two p53+/- and two pairs of BALB/c p53+/- were assayed on each platform, and for each gene on each platform a median expression ratio was determined. To analyze a specific small region, we performed a one subtype vs. the rest Student t test in a segment-by-segment manner. Mouse segments were rearranged in the order of human chromosome positions by chained alignments of human genome (National Center for Biotechnology Information Build 36) against mouse (University of California, Santa Cruz Genome Browser; http://genome.ucsc.edu) (6). Last, we note that the control DNA was FVB, whereas all tumor samples were BALB/c, therefore, regions of 100% gain or 100% loss could be attributed to strain-specific germline copy number differences, and caution is needed in interpreting these data; however, the strain differences will not affect any of the subtype-specific analyses.

Statistical Methods. Limiting dilution transplantation data were analyzed using a binomial generalized linear model with a complementary log-log link (7, 8). After determining that assumptions of the Single Hit Poisson Model were not met (8), we used the more general model, fitting parameters for slope, intercept, and interaction. After verifying lack of significant interaction between dose and cell line, we tested for the main effect of cell line. Results were summarized as the “fold change in dose” (FCD) required for equal take rates. The FCD 95% confidence interval was computed from the covariance matrix of the model parameters using the delta method (p 43 in ref. 9) and back-transformation by exponentiating.

Identification of Subtype-Specific DNA Copy Number Alterations. Copy number aberration events associated with each subtype (one subtype vs. the rest of the p53 null samples) were identified using 34 arrays representing the five subtypes only. Two levels of this analysis occurred next. First, at the genome level, SWITCH dna was used to identify significantly altered regions/segments and to determine the frequency of each copy number event (by segment) within each subtype, which was then visualized in the copy number landscape plots. Briefly, the SWITCH dna method first identifies the transition point in chromosome/probe copy number assessments by calculating the F statistics recursively, then tests the significance of segments according to the segment’s average intensity and segment size. With the default setting of SWITCHdna, 15,469 segments were defined in the 34 array comparative genomic hybridization slides. The copy number of each segment was calculated by taking the average value of probes in this region. To analyze a specific small region, we performed a one subtype vs. the rest Student t test in a segment-by-segment manner. Mouse segments were rearranged in the order of human chromosome positions by chained alignments of human genome (National Center for Biotechnology Information Build 36) against mouse (University of California, Santa Cruz Genome Browser; http://genome.ucsc.edu) (6). Last, we note that the control DNA was FVB, whereas all tumor samples were BALB/c, therefore, regions of 100% gain or 100% loss could be attributed to strain-specific germline copy number differences, and caution is needed in interpreting these data; however, the strain differences will not affect any of the subtype-specific analyses.

**Fig. S1.** Morphological features of p53 null mammary tumors. p53 null tumors display variable histological features, including spindloid tumors (B and E).
Fig. S2. ANOVA of gene signatures and individual genes across five subtypes of p53 null tumors.
Fig. S3. Immunofluorescent staining. p53 null basal tumors often show staining for both keratin 5 (K5) and keratin 8 (K8) (A and B). p53 null claudin-low tumors stain for K8 (C), some with less intensity (D). Some tumors are estrogen receptor positive (E). Claudin-low tumors show little staining for CLDN3 and CDH1 (F).

Fig. S4. p53 null tumor 2304L has (A) overexpression and (B) amplification of HER2/ERBB2.

Dataset S1. Summary of p53 null tumor samples

Dataset S1 (XLS)

Dataset S2. miRNAs differentially expressed in human claudin-low tumors

Dataset S2 (XLS)

Dataset S3. Copy number alterations by subtype

Dataset S3 (XLSX)
Dataset S4.  Comparison with Bergamaschi et al. (1)

Dataset S4 (XLSX)


Dataset S5.  Antibodies

Dataset S5 (XLS)