

Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin

Katherine A. Hoadley,^{1,20} Christina Yau,^{2,20} Denise M. Wolf,^{3,20} Andrew D. Cherniack,^{4,20} David Tamborero,⁵ Sam Ng,⁶ Max D.M. Leiserson,⁷ Beifang Niu,⁸ Michael D. McLellan,⁸ Vladislav Uzunangelov,⁶ Jiashan Zhang,⁹ Cyriac Kandoth,⁸ Rehan Akbani,¹⁰ Hui Shen,^{11,22} Larsson Omberg,¹² Andy Chu,¹³ Adam A. Margolin,^{12,21} Laura J. van't Veer,³ Nuria Lopez-Bigas,^{5,14} Peter W. Laird,^{11,22} Benjamin J. Raphael,⁷ Li Ding,⁸ A. Gordon Robertson,¹³ Lauren A. Byers,¹⁰ Gordon B. Mills,¹⁰ John N. Weinstein,¹⁰ Carter Van Waes,¹⁸ Zhong Chen,¹⁹ Eric A. Collisson,¹⁵ The Cancer Genome Atlas Research Network, Christopher C. Benz,^{2,*} Charles M. Perou,^{1,16,17,*} and Joshua M. Stuart^{6,*}

¹Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

²Buck Institute for Research on Aging, Novato, CA 94945, USA

³Department of Laboratory Medicine, University of California San Francisco, 2340 Sutter St, San Francisco, CA, 94115, USA

⁴The Eli and Edythe Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

⁵Research Unit on Biomedical Informatics, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Dr. Aiguader 88, Barcelona 08003, Spain

⁶Department of Biomolecular Engineering, Center for Biomolecular Sciences and Engineering, University of California, Santa Cruz, 1156 High St., Santa Cruz, CA 95064, USA

⁷Department of Computer Science and Center for Computational Molecular Biology, Brown University, 115 Waterman St, Providence RI 02912, USA

⁸The Genome Institute, Washington University, St Louis, MO 63108, USA

⁹National Cancer Institute, NIH, Bethesda, MD 20892, USA

¹⁰UT MD Anderson Cancer Center, Bioinformatics and Computational Biology, 1400 Pressler Street, Unit 1410, Houston, TX 77030, USA

¹¹USC Epigenome Center, University of Southern California Keck School of Medicine, 1450 Biggy Street, Los Angeles, CA 90033, USA

¹²Sage Bionetworks 1100 Fairview Avenue North, M1-C108, Seattle, WA 98109-1024, USA

¹³Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC V5Z 4S6, Canada

¹⁴Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys, 23, Barcelona 08010, Spain

¹⁵Department of Medicine, University of California San Francisco, 450 35d St, San Francisco, CA, 94148, USA

¹⁶Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

¹⁷Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

¹⁸Building 10, Room 4-2732, NIDCD/NIH, 10 Center Drive, Bethesda, MD 20892

¹⁹Head and Neck Surgery Branch, NIDCD/NIH, 10 Center Drive, Room 5D55, Bethesda, MD 20892

²⁰Co-first author

²¹Present address: Oregon Health Sciences University, 3181 S.W. Sam Jackson Park Road, Portland, Oregon 97239-3098, USA

²²Present address: Van Andel Research Institute, 333 Bostwick Avenue, Northeast, Grand Rapids, Michigan 49503, USA

*Correspondence: cbenz@buckinstitute.org (C.C.B.), cperou@med.unc.edu (C.M.P.), jstuart@ucsc.edu (J.M.S.)

<http://dx.doi.org/10.1016/j.cell.2014.06.049>

SUMMARY

Recent genomic analyses of pathologically defined tumor types identify “within-a-tissue” disease subtypes. However, the extent to which genomic signatures are shared across tissues is still unclear. We performed an integrative analysis using five genome-wide platforms and one proteomic platform on 3,527 specimens from 12 cancer types, revealing a unified classification into 11 major subtypes. Five subtypes were nearly identical to their tissue-of-origin counterparts, but several distinct cancer types were found to converge into common subtypes. Lung squamous, head and neck, and a subset of bladder cancers coalesced into one subtype typified by TP53 alterations, TP63 amplifications, and high

expression of immune and proliferation pathway genes. Of note, bladder cancers split into three pan-cancer subtypes. The multiplatform classification, while correlated with tissue-of-origin, provides independent information for predicting clinical outcomes. All data sets are available for data-mining from a unified resource to support further biological discoveries and insights into novel therapeutic strategies.

INTRODUCTION

Cancers are typically classified using pathologic criteria that rely heavily on the tissue site of origin. However, large-scale genomics projects are now producing detailed molecular characterizations of thousands of tumors, making a systematic molecular-based taxonomy of cancer possible. Indeed, The

Cancer Genome Atlas (TCGA) Research Network has reported integrated genome-wide studies of ten distinct malignancies: glioblastoma multiforme (GBM) ([Cancer Genome Atlas Research Network, 2008](#)), serous ovarian carcinoma (OV) ([Cancer Genome Atlas Research Network, 2011](#)), colon (COAD) and rectal (READ) adenocarcinomas ([Cancer Genome Atlas Research Network, 2012b](#)), lung squamous cell carcinoma (LUSC) ([Cancer Genome Atlas Research Network, 2012a](#)), breast cancer (BRCA) ([Cancer Genome Atlas Research Network, 2012c](#)), acute myelogenous leukemia (AML) ([Cancer Genome Atlas Research Network, 2013b](#)), endometrial cancer (UCEC) ([Kandoth et al., 2013b](#)), renal cell carcinoma (KIRC) ([Cancer Genome Atlas Research Network, 2013a](#)), and bladder urothelial adenocarcinoma ([Cancer Genome Atlas Research Network, 2014](#)). Those studies have shown that each single-tissue cancer type can be further divided into three to four molecular subtypes. The subclassification is based on recurrent genetic and epigenetic alterations that converge on common pathways (e.g., p53 and/or Rb checkpoint loss; RTK/RAS/MEK or RTK/PI3K/AKT activation). Meaningful differences in clinical behavior are often correlated with the single-tissue tumor types and, in a few cases, single-tissue subtype identification has led to therapies that target the driving subtype-specific molecular alteration(s). *EGFR* mutant lung adenocarcinomas and *ERBB2*-amplified breast cancer are two well-established examples.

To move toward a molecular taxonomy, we investigated whether tissue-of-origin categories split into subtypes based upon multiplatform genomic analyses and also extend the analysis in the other direction to look for possible convergence. We looked to see what molecular alterations are shared across cancers arising from different tissues and if previously recognized disease subtypes in fact span multiple tissues of origin. With those questions in mind, we performed a multiplatform integrative analysis of thousands of cancers from 12 tumor types in The Cancer Genome Atlas (TCGA) project. Using data from multiple assay platforms, we tested the hypothesis that molecular signatures provide a distinct taxonomy relative to the currently used tissue-of-origin-based classification. At the center of our results is the identification of 11 “integrated subtypes.” Consistent with the histological classification, tissue-of-origin features provided the dominant signal(s) for identification of most subtypes, irrespective of genomic analysis platform or combination thereof. However, approximately 10% of cases were reclassified by the molecular taxonomy, with the newly defined integrated subtypes providing a significant increase in the accuracy for the prediction of clinical outcomes.

RESULTS

Samples, Data Types, and Genomic Platforms

To identify a multitissue, molecular signature-based classification of cancer objectively, we first characterized each of the individual tumor types using six different “omic” platforms. The diverse tumor set called “Pan-Cancer-12,” is composed of 12 different malignancies. It comprises 3,527 cases assayed by at least four of the six possible data types routinely generated by TCGA: whole-exome DNA sequence (Illumina HiSeq and GAI), DNA copy-number variation (Affymetrix 6.0 microarrays), DNA

methylation (Illumina 450,000-feature microarrays), genome-wide mRNA levels (Illumina mRNA-seq), microRNA levels (Illumina microRNA-seq), and protein levels for 131 proteins and/or phosphorylated proteins (Reverse Phase Protein Arrays; RPPA). The 12 tumor types include the ten TCGA Network published data sets listed above and two additional tumor types for which manuscripts have been submitted: lung adenocarcinoma (LUAD) and head and neck squamous cell carcinoma (HNSC). This is the most comprehensive and diverse collection of tumors analyzed by systematic genomic methods to date.

We performed sample-wise clustering to derive subtypes based on six different data types separately: DNA copy number, DNA methylation, mRNA expression, microRNA expression, protein expression, and somatic point mutation (see [Extended Experimental Procedures](#), Section 1). The classification results from each single-platform analysis produced sets of 8 to 20 groups of samples that each showed high correlation with tissue of origin ([Figures S1A–S1F](#) available online) and were highly comparable with each other ([Figure S2A](#)). For example, patterns of copy-number change varied across tissue types, and subtyping of the tumors based on copy-number alterations revealed a significant correlation with tissue ($p < 6 \times 10^{-6}$, Chi-square test).

Integrated Platform Analysis—Cluster-of-Cluster Assignments

To identify disease subtypes on a more comprehensive basis than could be done using any single type of data, we developed an integrated subtype classification for all of the tumor samples in the Pan-Cancer-12 collection based on five of the data types, excluding somatic mutations. To do so, the results of the single-platform analyses were provided as input to a second-level cluster analysis using a method we refer to as cluster-of-cluster assignments (COCA), which was originally developed to define subclasses in the TCGA breast cancer cohort ([Cancer Genome Atlas Research Network, 2012c](#)). The algorithm takes as input the binary vectors that represent each of the platform-specific cluster groups and reclusters the samples according to those vectors (see [Extended Experimental Procedures](#), Section 2). One advantage of the method is that data across platforms are combined without the need for normalization steps prior to clustering. In addition, each platform influences the final integrated result with weight proportional to the number of distinct subtypes reproducibly found by consensus clustering. Thus, “large” platforms (e.g., 450,000 DNA methylation probes) with orders of magnitude more features than “small” platforms (e.g., 131 RPPA antibodies) do not dominate the solution.

In addition to the COCA classification, we used two additional, independent methods to derive Pan-Cancer-12 subtypes based on integrated data: (1) an algorithm called SuperCluster ([Kandoth et al., 2013b](#)) ([Figure S2B](#)) and (2) clustering based on inferred pathway activities from PARADIGM ([Vaske et al., 2010](#)), which integrates gene expression and DNA copy-number data with a set of predefined pathways to infer the degree of activity of 17,365 pathway features such as proteins, complexes, and cellular processes ([Figure S2C](#)). Both SuperCluster and PARADIGM produced classifications that were highly concordant with the COCA subtypes ([Figure S2D](#)). Given recent promising results that use gene networks (as opposed to the sparsely

Table 1. The 12 Pathological Disease Types, Rows, and Their Relationship to the 13 Integrated Subtypes Defined by the Cluster-of-Cluster-Assignments Method

Handle	C1-LUAD-Enriched	C2-Squamous-like	C3-BRCA/Luminal	C4-BRCA/Basal	C5-KIRC	C6-UCEC	C7-COAD/READ	C8-BLCA	C9-OV	C10-GBM	C11-Small-Variou	C12-Small-Variou	C13-AML	Total
BLCA	10	31	0	0	1	0	0	74	0	1	1	2	0	120
BRCA	2	1	688	135	5	0	0	2	0	0	0	0	1	834
COAD	0	0	0	0	0	0	182	0	0	0	0	0	0	182
GBM	3	0	0	0	2	0	0	0	0	190	0	0	0	195
HNSC	1	302	0	0	0	0	0	1	0	1	0	0	0	305
KIRC	1	0	0	0	470	0	0	0	0	2	0	2	0	475
LAML	0	0	0	0	0	0	0	0	0	0	0	0	161	161
LUAD	258	6	0	1	0	1	0	1	0	1	0	2	0	270
LUSC	28	206	0	1	0	0	0	1	0	2	0	0	0	238
OV	1	0	0	0	1	0	0	0	327	0	0	0	0	329
READ	0	0	0	0	0	0	73	0	0	0	0	0	0	73
UCEC	2	0	0	0	0	340	1	0	0	0	2	0	0	345
Totals	306	546	688	137	479	341	256	79	327	197	3	6	162	3527

The name of each COCA subtype (top row) includes a cluster number (1 to 13) and a text designation for mnemonic purposes. Two of the subtypes (numbers 11 and 12) were eliminated from further analysis because they included < 10 samples (3 and 6 samples, respectively). Hence, the text focuses on 11 subtypes, not 13.

See also [Table S1](#).

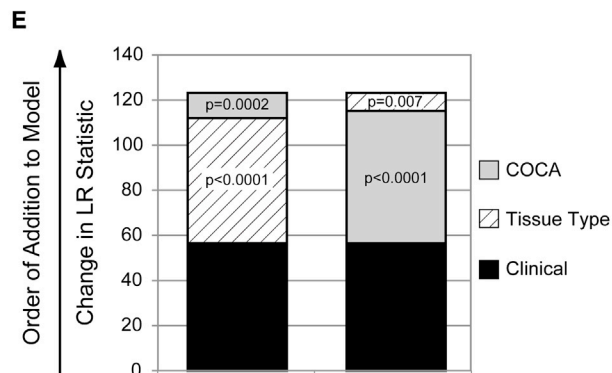
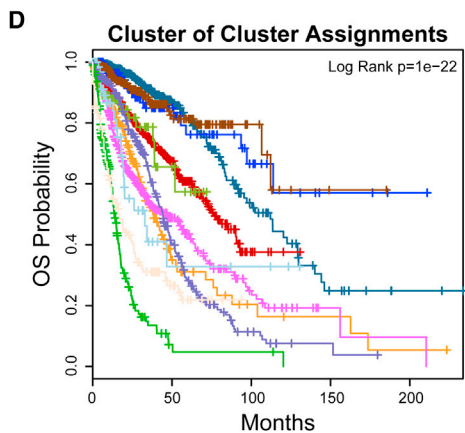
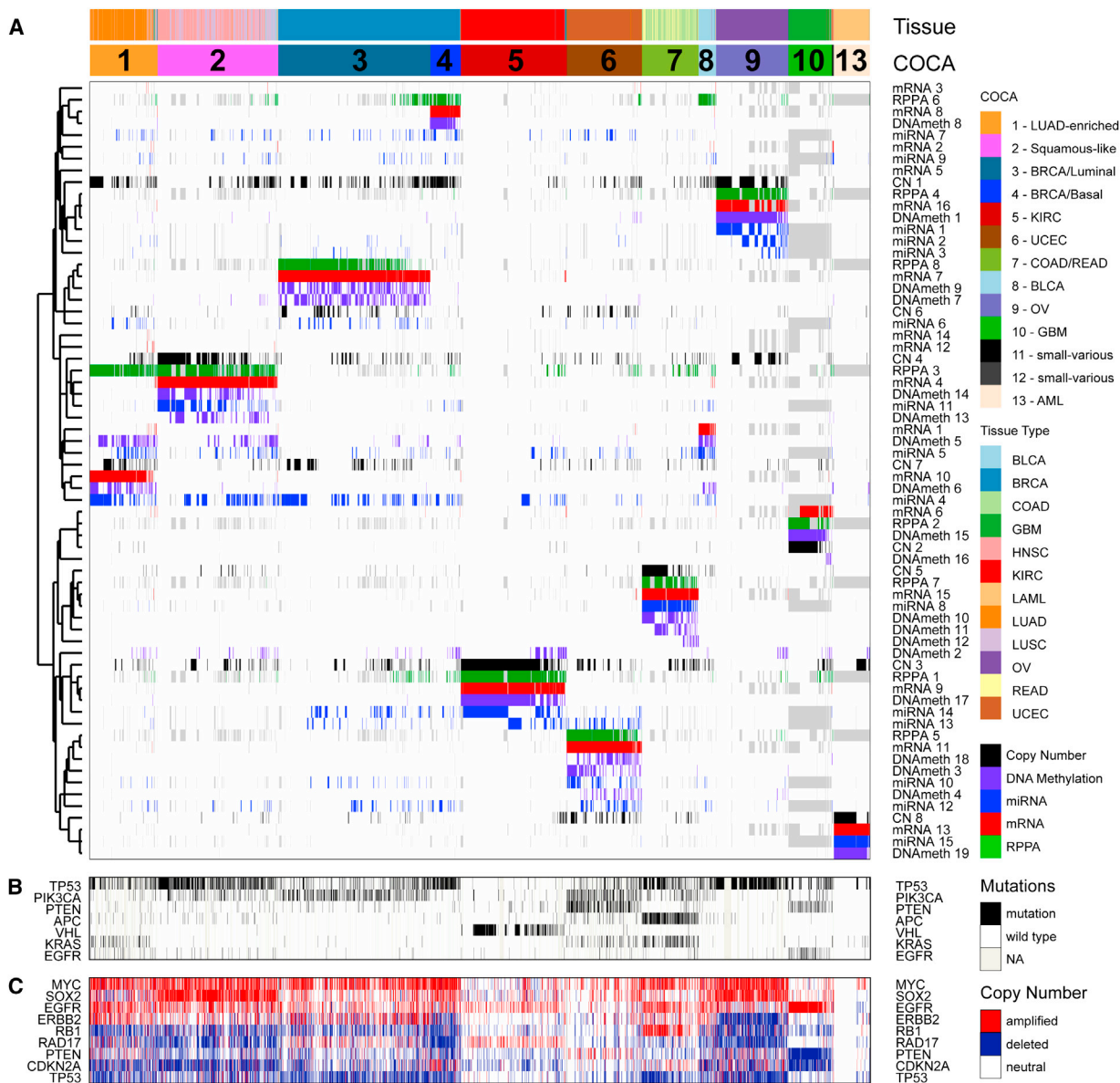
populated single-mutation space) to cluster samples based on somatic DNA variants (Hofree et al., 2013), we calculated a mutation-based clustering after first associating genes with pathways and then identifying clusters based on *mutated pathways* (Figure S1F and Data S1). Including those clusters in the identification of COCA subtypes produced highly similar results to COCA subtypes that did not use the mutation-based clusters (Figure S2D). Thus, we focus here on the COCA results obtained without the mutations, as those five other platform-based classifications required no prior biological knowledge.

The COCA algorithm identified thirteen clusters of samples, 11 of which included more than ten samples (Table S1). The two small clusters (n = 3 and 6) are noted (Table 1) but were excluded from further analyses. We refer to the remaining sample groups by cluster number and a short descriptive mnemonic (Table 1). Of the 11 COCA-integrated subtypes, five show simple, near one-to-one relationships with tissue site of origin: C5-KIRC, C6-UCEC, C9-OV, C10-GBM, and C13-LAML (Figure 1A). A sixth COCA type, C1-LUAD-enriched, is predominantly composed (258/306) of non-small cell lung (NSCLC) adenocarcinoma samples (LUAD). The second major constituent of the C1-LUAD-enriched group is a set of NSCLC squamous samples (28/306). Upon re-review of the frozen or formalin fixed sections, 11/28 lung squamous samples that cluster with the C1-LUAD-enriched group did not have squamous features and were reclassified as lung adenocarcinoma (Travis et al., 2011). NSCLCs are often difficult to classify based on histology alone (Grilley-Olson et al., 2013). That difficulty poses an important clinical challenge since histology is used to guide the selection of chemotherapy (Scagliotti et al., 2008) and to select patients for further mutational analysis (e.g., *EGFR* mutation and *ALK* fusion testing in nonsquamous NSCLC). However, the challenge

can be addressed by genomic analysis based on distinct differences in mutation spectrum (Table S2A) and distinct gene expression patterns (Figure S1A). Two clear subtypes of NSCLC (C1-LUAD-enriched and C2-squamous-like, see discussion below) are identified by COCA.

For the other five tissue types, the patterns are more complex. Either a given tissue splits into multiple COCA groups (divergence) or multiple tissue types coalesce into a single COCA group (convergence). A simple example of convergence previously described for TCGA data are the merging of colon (COAD) and rectal (READ) tumors into a single COCA group (Cancer Genome Atlas Research Network, 2012b). The expression features shared by colon and rectal samples were noted in the TCGA Network paper on the two cancer types, but we extend those findings through use of the multiplatform clustering approach (Figure 1A and Table 1).

Breast cancers (BRCA) exhibit a pattern of divergence in which two main groups of samples are distinctly identifiable. One group (C3-BRCA/Luminal) contains essentially all of the luminal (estrogen receptor-positive) (594/597) and HER2-positive tumors (66/66), whereas the other (C4-BRCA/basal) contains 131/139 of the breast basal-like tumors. Although it has previously been appreciated that basal-like breast cancers (the majority subset of triple-negative breast cancers) form a distinct subtype (Prat et al., 2013; Cancer Genome Atlas Research Network, 2012c), the findings here provide a more refined, quantitative picture of the extent of difference from luminal and basal-like breast cancers. Whereas tissue of origin is the dominant signal for combined data on almost all of the other cancer types in the Pan-Cancer-12 collection, breast basal-like cancers are as different from luminal/ER+ breast cancers as they are from cancers of the lung (Figure 1A). The data from the present study



(legend on next page)

strongly reinforce the idea that basal-like breast cancers constitute a unique disease entity.

The remaining three tissue types (HNSC, LUSC, and BLCA) provide examples of both divergence and convergence in COCA subtyping (Figure 1A and Table 1). The strongest pattern of convergence is observed for the vast majority of HNSC (301/304), LUSC (206/238), and some of the BLCA (31/120) tumors; they cluster together in a large COCA group (C2-squamous-like), perhaps reflecting similar cell type of origin or smoking as an etiologic factor. BLCA tumors also exhibit a divergence pattern, distributing predominantly into three distinct groups: 31 BLCA in the C2-squamous-like group, 10 in the C1-LUAD-enriched group, and 74 in the bladder-only group, C8-BLCA. Five other BLCA samples cluster in four different COCA groups.

Clinical Importance of the COCA Subtypes

To investigate the clinical relevance of the COCA subtypes, we performed Kaplan-Meier Survival analysis on the Pan-Cancer-12 data set. The results indicate that tissue-of-origin (Figure S3A) and COCA subtype (Figure 1D) are both prognostic, and each provides independent information (Figure 1E). Additionally, the two most commonly mutated genes in the overall data set, *TP53* (41%) and *PIK3CA* (20%), are prognostic, even across different tumor types, as are previously defined genomic signatures of cell proliferation rate (Nielsen et al., 2010) and mutated *TP53* gene-expression-based signature (Troester et al., 2006) (Figures S3B–S3F).

We next asked whether prognostic information is provided by the COCA subtypes after accounting for known clinical and tissue-of-origin features. We performed a multivariate Cox proportional hazards analysis to predict outcomes across the data set. The analysis was limited to the COCA subtypes that did not have a one-to-one relationship with tissue-of-origin tumor type (COCA1-LUAD enriched, COCA2-squamous, COCA3-Breast/luminal, COCA4-breast/basal, COCA7-COAD/READ, and COCA8-BLCA). In the model we included clinical features such as tumor size, node status, metastasis status, and age at diagnosis, as well as tissue of origin. We performed a likelihood ratio test conditioning first on the clinical variables; when either tissue-of-origin or COCA subtype was added to

the model, a large increase in the predictive fit of the model was observed, beyond what one would get with the clinical information alone (Figure 1E). That observation supports the classical model in which tissues of origin provides strong predictions of outcome. Next, we asked whether the COCA subtypes add additional independent information for predicting survival beyond the combination of tissue-of-origin and clinical features. Indeed, we observed a significant increase in statistical likelihood when COCA is added to a multivariate model that already includes the clinical and tissue-based information ($p < 0.0002$; Chi-square test; Figure 1E). Thus, while the COCA classification differs from tissue-of-origin-based classification in only ~10% of all samples, the difference does provide important molecular information that reflects tumor biology and is associated with clinical outcome.

Genomic Determinants of the Integrated COCA Subtypes

We next identified the major genomic determinants of the COCA subtypes, including somatic mutations and DNA copy-number changes (Figures 1B and 1C). For single-nucleotide variants, we analyzed a Pan-Cancer-12 list of 127 significantly mutated genes (SMGs) obtained by MuSiC analysis (Kandoth et al., 2013a). Only three of the genes are mutated at a frequency $\geq 10\%$ (*TP53*, *PIK3CA* and *PTEN*), and 11 additional are mutated at $\geq 5\%$ frequency (Table S2A). We also include a list of 291 high-confidence cancer drivers (HCDs) from Pan-Cancer-12 analysis (Tamborero et al., 2013) and identified by a combination of five complementary methods to identify signals of positive selection in the mutational pattern of genes across tumors.

A large number of correlations between COCA subtypes and somatic mutations were found (Figure 2A, Figure S4D, Data S2). Somatic mutations clearly distinguish the C1-LUAD-enriched group from the C2-squamous-like group. *KEAP1* and *STK11* are preferentially mutated in C1-LUAD-enriched tumors, whereas *CDKN2A*, *NOTCH1*, *MLL2*, and *NFE2L2*, among others, are preferentially mutated in C2-squamous-like (Figure 2A). A similarly distinct set of SMGs was seen for the C3-BRCA/luminal and C4-BRCA/basal groups; only two genes are shared (*TP53* and *PIK3CA*), and they show different

Figure 1. Integrated Cluster-Of-Cluster Assignments Analysis Reveals 11 Major Subtypes

(A) Integration of subtype classifications from 5 “omic” platforms resulted in the identification of 11 major groups/subtypes from 12 pathologically defined cancer types. The groups are identified by number and color in the second bar, with the tissue of origin specified in the top bar. The matrix of individual “omic” platform type classification/subtype schemes was clustered, and each data type is represented by a different color: copy number, black; DNA methylation, purple; miRNA, blue; mRNA, red; and RPPA, green.

(B) Mutation status for each of ten significantly mutated genes coded as: wild-type, white; mutant, red; missing data, gray.

(C) Copy-number status for each of nine important genes: amplified, red; deleted, blue; copy-number neutral, white; and missing data, gray. The color-coding schema is shown to the right.

(D) Overall survival (OS) of COCA subtypes by Kaplan-Meier plot. COCA subtypes are highly correlated with overall survival outcomes.

(E) The log-likelihood ratio (LR) statistic was estimated as we added clinical variables, COCA subtype, or tissue-type information to a cox proportional hazards model. Clinical variables included age at diagnosis, tumor size, node status, and metastasis status. The change in LR statistic as features were added to the model was assessed for significance by chi-square analysis. The set of samples was limited to the set of tumor types that did not have a one-to-one relationship with a COCA subtype: BLCA, BRCA, COAD, HNSC, LUAD, LUSC, and READ in COCA clusters COCA1-LUAD-enriched, COCA2-squamous, COCA3-BRCA/luminal, COCA4-BRCA/basal-like, COCA7-COAD/READ, and COCA8-BLCA. Left bar shows results of adding tissue-of-origin to clinical variables already part of the model, followed by a variable representing the COCA subtyping; right bar shows results when COCA is first added on to clinical variables, and then tissue-type is added. In each case the increase in the ability to predict OS was in terms of the LR.

See also Figures S1, S2, S3 and Data S1, S2, and S3.

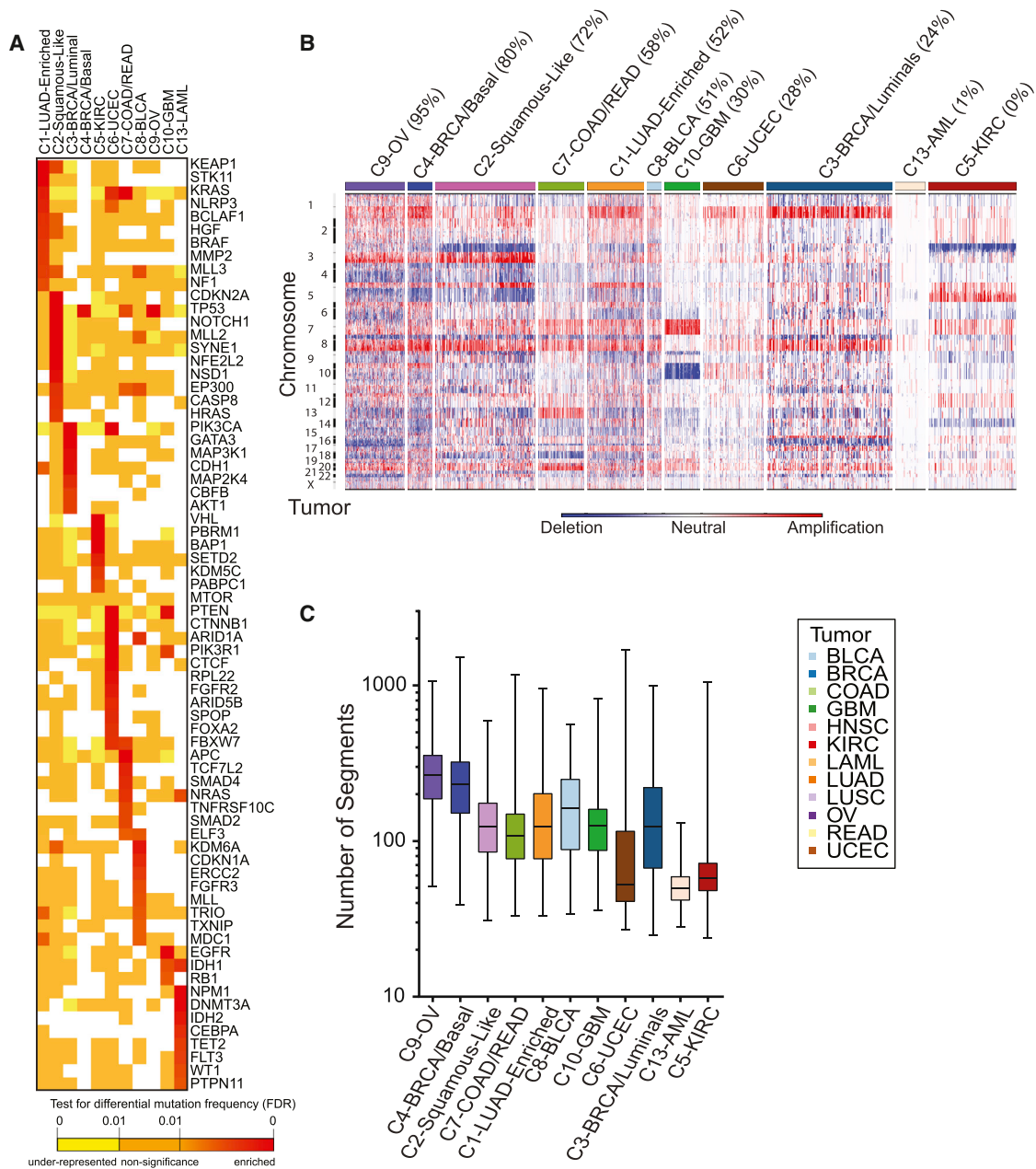


Figure 2. Genomic Determinants of the Integrative COCA Subtypes

(A) Genes from the high-confidence list of drivers (Tamborero et al., 2013) found to be mutated at a different rate within one COCA subtype compared outside it based on a two-tailed Fisher's exact test. Mutation frequency enrichment, red to orange; genes with mutations equaling the background rate, yellow; genes with no observed mutations in a subtype, white. Displayed are top-ranked genes in terms of significant mutation enrichment (FDR < 1%) in at least one COCA subtype. (B) Somatic copy-number alterations (SCNAs) in integrative clusters. SCNAs in tumors (horizontal axis) are plotted along chromosomal locations (vertical axis). The heatmap shows the presence of amplifications (red) and deletions (blue) throughout the genome. The color strip along the top indicates integrative COCA cluster membership; the number in parentheses indicates percentage of samples in a COCA subtype with TP53 mutation. COCA subtypes are ordered from highest TP53 mutant percentage to lowest.

(C) Range of copy-number segments in tumors within each integrative cluster. The box and whisker plots show the middle quartiles and the minimum and maximum number of segments in each cluster group.

See also Figure S4 and Tables S2 and S3.

mutation frequencies (Table S2A). Since the somatic mutation results were not used in any way to determine the COCA subtypes, they provide independent evidence that distinctly

different genetic events underlie the subtypes. A protein-protein interaction network analysis of mutations associated with the COCA subtypes obtained using a new version of

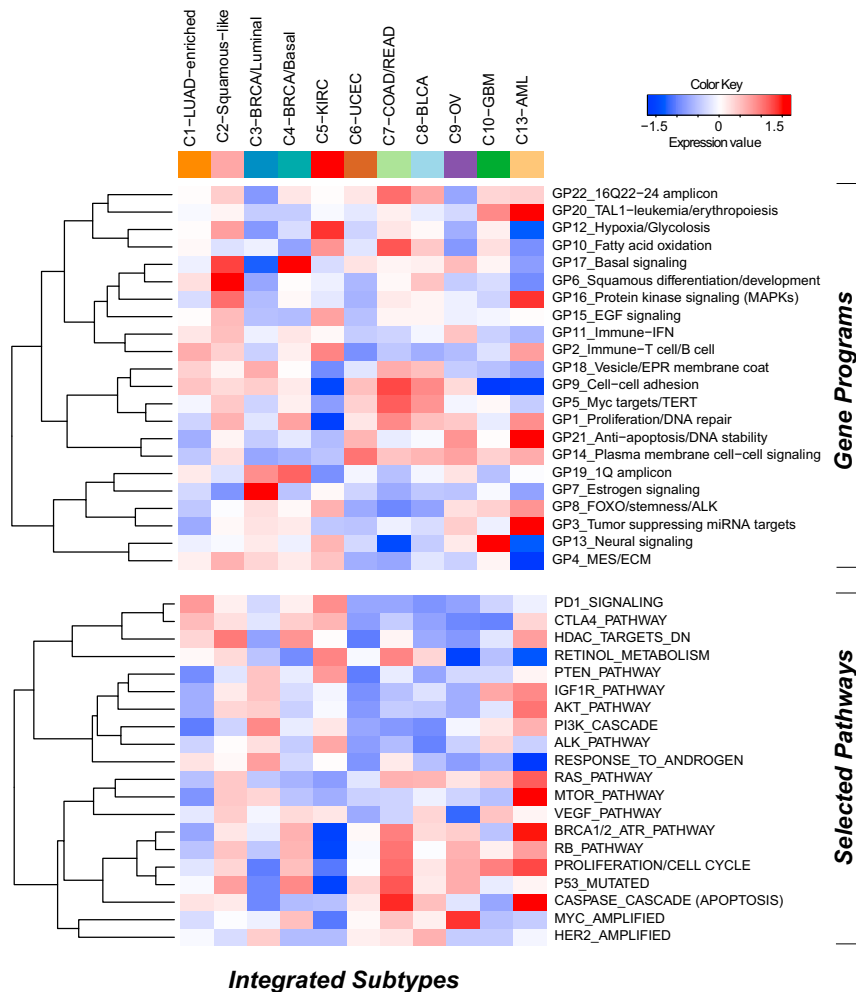


Figure 3. Subtype-Specific Patterns of Gene-Program and Selected Pathway Expression Characterizing Each Pan-Cancer-12 COCA Subtype

The heat map shows integrative subtypes in numerical order. Gene programs (top) and pathway signatures from a compendium (Table S4) containing signatures for drug targets and canonical pathways (bottom) were clustered separately from each other. Red-blue intensities reflect the means of the scores (red = high, white = average, blue = low). See also Figures S5 and S6, and Data S5.

Expression-Based Determinants of the Integrated Subtypes

We next sought to identify gene expression modules characteristic of each COCA subtype. First, we started with 6,898 sets of gene signatures documented to be coexpressed, coamplified, or to function together. From these, we identified *gene programs* as those whose genes have mRNA-seq signatures of high mutual correlation across the Pan-Cancer-12 data set. After applying a bimodality filter and weighted gene correlation network-based clustering, 22 nonredundant gene programs were identified (Table S4A, Figure S5A, Extended Experimental Procedures, Section 5, and Data S5). Linear classification with the 22 gene programs reconstituted the 11 integrated subtypes with 90% accuracy (Figure S5A and Table S4B). To view the expression-based determinants of the integrated subtypes, we plotted the

average expression level of each gene program within each COCA cluster (Figure 3). As expected, the gene programs *GP6-squamous differentiation/development*, *GP13-neural signaling* and *GP20-TAL-1-leukemia/erythropoiesis* were the most highly expressed in the C2-squamous-like, C10-GBM, and C13-LAML subtypes, respectively. As well, *GP7-Estrogen signaling* was highest in the C3-BRCA/luminal cases, whereas *GP17-basal signaling* had its highest levels in the C4-BRCA/basal cases. Activated pathway characteristics found by enrichment and subnetwork analyses based on PARADIGM inferences, many of which were consistent with the gene program analysis, are summarized in Table S4A (see Extended Experimental Procedures).

the HotNet algorithm (Vandin et al., 2012) provides an overview of the genomic determinants of the COCA subtypes (Figure S4E). The degree of genomic instability was a major determinant of subtype, as revealed in copy-number variation (CNV) data (Figure 2B). The C9-OV, C4-BRCA/basal, and C1-LUAD-enriched subtypes showed the most marked genomic instability, as assessed by average number of copy-number segments per subtype (Figure 2C), whereas AML and UCEC showed the least. Numerous COCA subtype-associated alterations implicated specific regions, arm-level copy-number changes (Figure S4A) and/or focal regions of copy-number alteration (Figure S4B). Of note were a number of previously described tissue-type-specific and subtype-specific alterations, including Chr7 gain and Chr10 loss in GBM (Cancer Genome Atlas Research Network, 2008), 3p loss and 5q gain in kidney (Cancer Genome Atlas Research Network, 2013a), 4q and 5q loss in breast basal-like cancers (Cancer Genome Atlas Research Network, 2012c), and 3p loss and 3q gain in lung squamous tumors (Cancer Genome Atlas Research Network, 2012a). Of note, the latter were seen in most C2-squamous-like tumors, regardless of tissue of origin.

Gene expression programs and PARADIGM pathways carry clinically relevant information beyond tissue of origin as evidenced by a multivariate Cox model of survival with COCA subtype as a covariate (see Table S4E). Squamous differentiation/development (GP6), proliferation/cell cycle, and estrogen signaling (GP7) were significant predictors in the model. Intriguingly, GP7, along with *fatty acid oxidation* (GP10), *tumor suppressing miRNA targets* (GP3), and the *PTEN/MTOR signaling program*, were found to be significantly associated with

patient outcome in kidney cancer using a Cox proportional hazards survival analysis (Figure S5D and Table S4F). In common with the C3-BRCA/luminal subtype cases, higher levels of *estrogen signaling* (GP7) were also associated with better prognosis for C5-KIRC cancers. Consistent with the higher frequency of elevated HER2 protein levels in bladder, colorectal, and serous endometrial cancers (Akbari et al., 2014), the *HER2-amplified* gene signature appeared elevated in the C8-BLCA, C7-COAD/READ, and C6-UCEC subtypes, as well as the C3-BRCA/luminal subtype which contains all HER2-positive breast cancers. Predictors independent of disease stage included basal signaling (GP17), associated with decreased overall survival, and the immune-related PARADIGM pathways PD1_signaling and CTLA4_pathway, both of which were associated with increased overall survival. These immune-related signatures may reflect varying amounts of lymphocyte infiltrate in the tumors as has been estimated by DNA methylation-based analysis of the Pan-Cancer-12 data set (Figure S5E). In any case, these immune cell-associated gene programs may be pertinent to emerging treatment strategies based on immune modulation. Overall, despite uneven clinical information and follow-up across the many different Pan-Cancer types, expression-based determinants of the integrated subtypes were sufficiently informative to identify pathway-based features of prognostic value that transcend tissue-of-origin cancer types.

Multiple-Platform Determinants of the Integrated Subtypes

To gain insight into the genetic and epigenetic determinants that characterize each of the COCA subtypes, we calculated differential gene scores derived from each of the separate six platforms (see Extended Experimental Procedures, Section 4) as well as PARADIGM pathway features. All differential activities were mapped to individual genes so that thematic pathways could be identified (see Data S2). Copy-number data were summarized at the gene level using GISTIC 2.0 and t tests for every gene were performed for each COCA subtype. DNA Methylation probes were associated with any gene that fell in the $\pm 1,500$ bp region surrounding gene transcriptional start sites. Genes with differential mRNA expression were identified using a SAM analysis on the RSEM values. Genes with differentially expressed protein products were determined by running a t test on the 131 protein forms represented on the RPPA data. For mutations, a Fisher's exact test on the frequency within a COCA subtype compared to outside the subtype was performed for all of a set of 291 high-confidence driver genes (Tamborero et al., 2013). Differentially expressed miRNAs for each COCA subtype were identified using a Wilcoxon rank-sum test based on the miRNA-Seq data. Genes were then identified as those predicted to be targeted by a differentially expressed miRNA that was also anticorrelated across the Pan-Cancer-12 data set.

Three approaches were used to summarize the unique features of the COCA subtypes. First, gene set enrichment analysis (GSEA) was run on the single-platform gene-based results and then clustered for visual inspection to elucidate distinctive pathways (Figure S6A). Second, a supervised Elastic Net approach was used to classify the COCA subtypes with 95% accuracy in cross-validation and the predictive features were

collected (Figure S6B; Data S3). Third, "regulatory hubs" from PARADIGM with more than 15 downstream targets and found to be differentially activated within a COCA subtype relative to other subtypes were collected (Table S5A). All three approaches revealed that each platform detects different pathways and features with respect to both COCA subtypes and data platforms. The identified discriminating features of the COCA subtypes confirm several expectations: (1) C3-BRCA/luminal was defined by protein and gene signatures for ER and GATA3 determined by the Elastic Net model as well as by PIK3CA-related signaling revealed by copy-number variation-based and mutation-based GSEA, (2) C5-KIRC was defined by multiple features of hypoxia found by mutation- and mRNA-Seq-based GSEA as well as predictive Elastic Net features, and (3) C7-COAD/READ was in part defined by APC mutations.

Convergence of the Squamous-like Subtype

A striking finding of the integrative subtype analysis was the coalescence of four distinct tumor types (LUSC, HNSC, some BLCA, and a very few LUAD) into the single C2-squamous-like subtype. We investigated the genomic- and pathway-based determinants of the subtype. The three main tumor types included shared loss of 3p and increased *TP63*, *PIK3CA*, and *SOX2* gene copies within a characteristic 3q amplicon (Figure 4A). Those regions are well known in LUSC (Cancer Genome Atlas Research Network, 2012a) and HNSC (Bhattacharya et al., 2011; Walter et al., 2013), and the results here extend that observation to include a subgroup of BLCA cases. In addition, the C2-squamous-like subtype tends to show amplification of *MYC* and loss of *CDKN2A*, *RB1*, and *TP53*. *TP53* mutation is frequent (72%), followed by a dramatic drop-off in mutation frequency to *MLL2* (20%), *PIK3CA* (19%), *CDKN2A* (18%), *NOTCH1* (16%), *NFE2L2* (10%), and *MALAT1* (6%), the only other genes mutated at $\geq 10\%$ frequency (Table S2A). Of potential interest in the C2-squamous-like group, tumors without *TP53* mutations show a higher density of *PIK3CA* mutations (Figure 1), consistent with recent evidence linking *PI3K* activation and wild-type *TP53* inactivation in HNSC (Herzog et al., 2013). Putative driver analysis identified several genes (*PIK3CA*, *MLL2*, and *KEAP1*) frequently mutated in the C2-squamous-like group but also in other COCA subtypes (Figure 4B). Of these, *FRG1B* and *CASP8* were found to be significantly more associated with HNSC by Fisher's exact test. Putative driver analysis also revealed a number of genes with higher mutation frequencies in the C2-squamous-like subtype than in any other subtype: *TP53*, *SYNE1*, *MLL2*, *CDKN2A*, *NOTCH1*, *NFE2L2*, and *EP300*, among others (Figure 4C; Figure S7A).

An extension of the HotNet algorithm (Vandin et al., 2012) was run on all genes mutated in $\geq 2\%$ of any one subtype in conjunction with the HINT physical protein-protein interaction network (Extended Experimental Procedures, Section 4; Table S3). HotNet identified four subnetworks of mutated genes characteristic of the C2-squamous-like subtype (Table S3B). The largest, most frequently mutated subnetwork (91.7% of C2-squamous-like samples) includes many well-known cancer genes and tumor suppressors, including *TP53*, *CDKN2A*, and *PTEN*. The second most mutated subnetwork (59.9%) consists of *NFE2L2*, *CUL3*, and *KEAP1*, *CCNE1*, *FBXW7*, and *NOTCH1*.

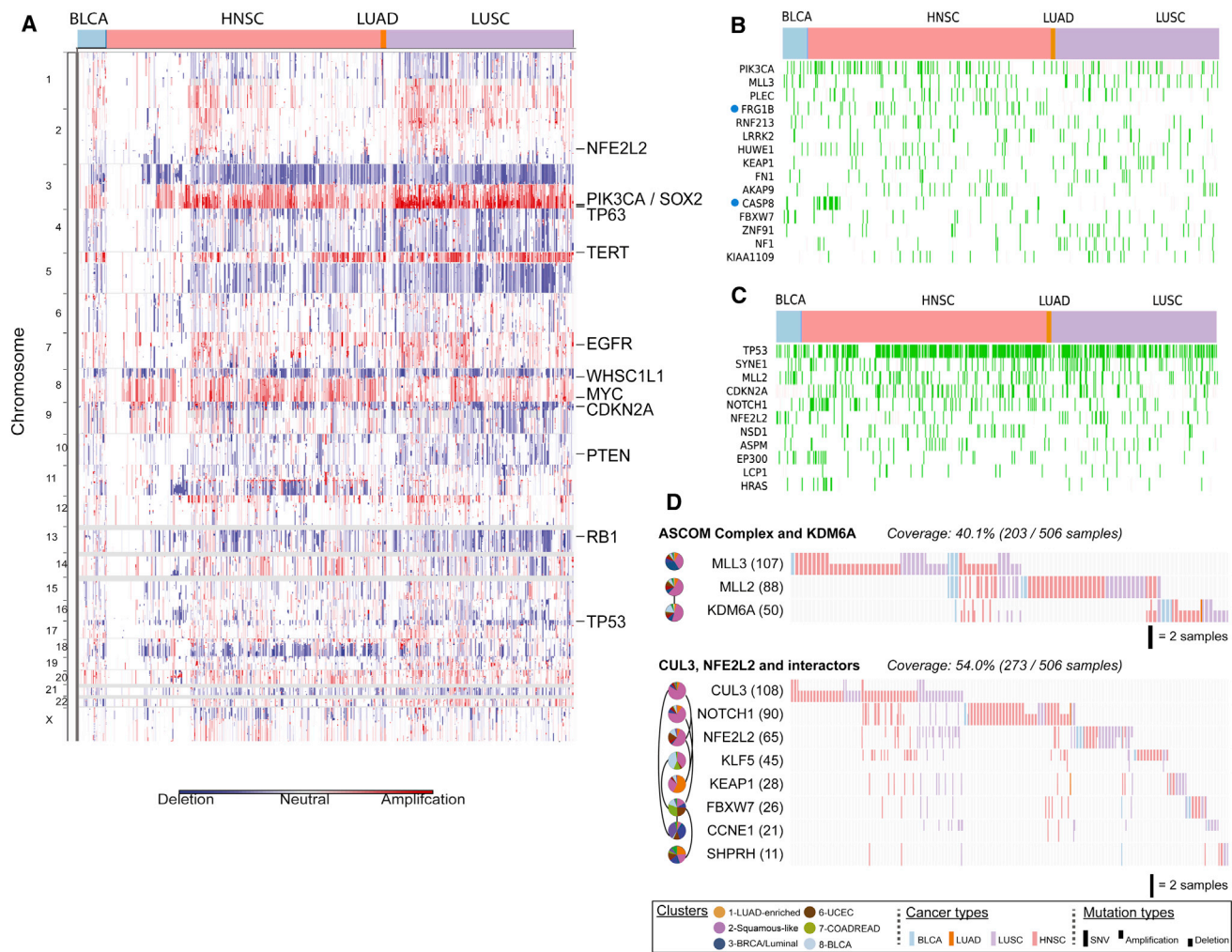


Figure 4. Genomic Determinants of the C2-Squamous-like COCA Subtype

(A) SCNAs for the C2-squamous-like subtype are shown, highlighting the importance of 3q26 gains across the different tissue-of-origin samples.

(B) Selected genes from 291 high-confidence driver (HCD) genes (Tamborero et al., 2013) mutated in >5% of C2-squamous-like samples and comparable in frequency in other subtypes. Samples with protein-affecting mutations in those genes are shown in green.

(C) HCD genes (as in B) with mutation frequency significantly higher in C2-squamous-like tumors relative to others (stated at $p < 0.01$ according to Fisher's exact test with FDR correction). The method used corrections for imbalance in the number of samples from different tissues (see [Extended Experimental Procedures](#), section 8).

(D) Two subnetworks of mutated pathways identified by an updated HotNet algorithm analysis using HINT interactions (see [Extended Experimental Procedures](#)) as mutated in at least 20% of the samples of the C2-squamous-like subtype. Links between pie charts indicate interactions among the proteins in each subnetwork. Each gene (node) is colored by wedges whose size indicates the relative proportion of the gene's mutations that are in samples from each integrated subtype. To the right of the pie charts is a gene-by-sample mutation matrix representing the mutation status of each gene across all squamous-like samples. Full ticks represent SNVs, lower ticks represent deletions and upper ticks represent amplifications. The color of each tick indicates tissue-of-origin type, with gray indicating no mutation in the corresponding sample.

See also [Figure S7](#) and [Table S5](#).

NFE2L2, *CUL3*, and *KEAP1* are well-known regulators of oxidative stress. The third most mutated subnetwork (37.1% of squamous samples) includes the ASCOM complex (*MLL2* and *MLL3*) and the putative ASCOM-interacting protein *KDM6A*. These proteins are involved in histone modifications that promote transcription. In addition, consistent with previous reports on collective motility in squamous cell carcinomas (Friedl and Gilmour, 2009), RAC and RHO signaling are also elevated in the C2-squa-

mous-like subtype based on PARADIGM analysis (Table S4 and Figure S7B).

Molecular Features Common to the Squamous, Breast Basal, and Ovarian Subtypes

Past work highlighted transcriptional similarities between the breast basal-like subtype and LUSC (Chung et al., 2002), as well as breast basal-like and serous ovarian cancers (Cancer

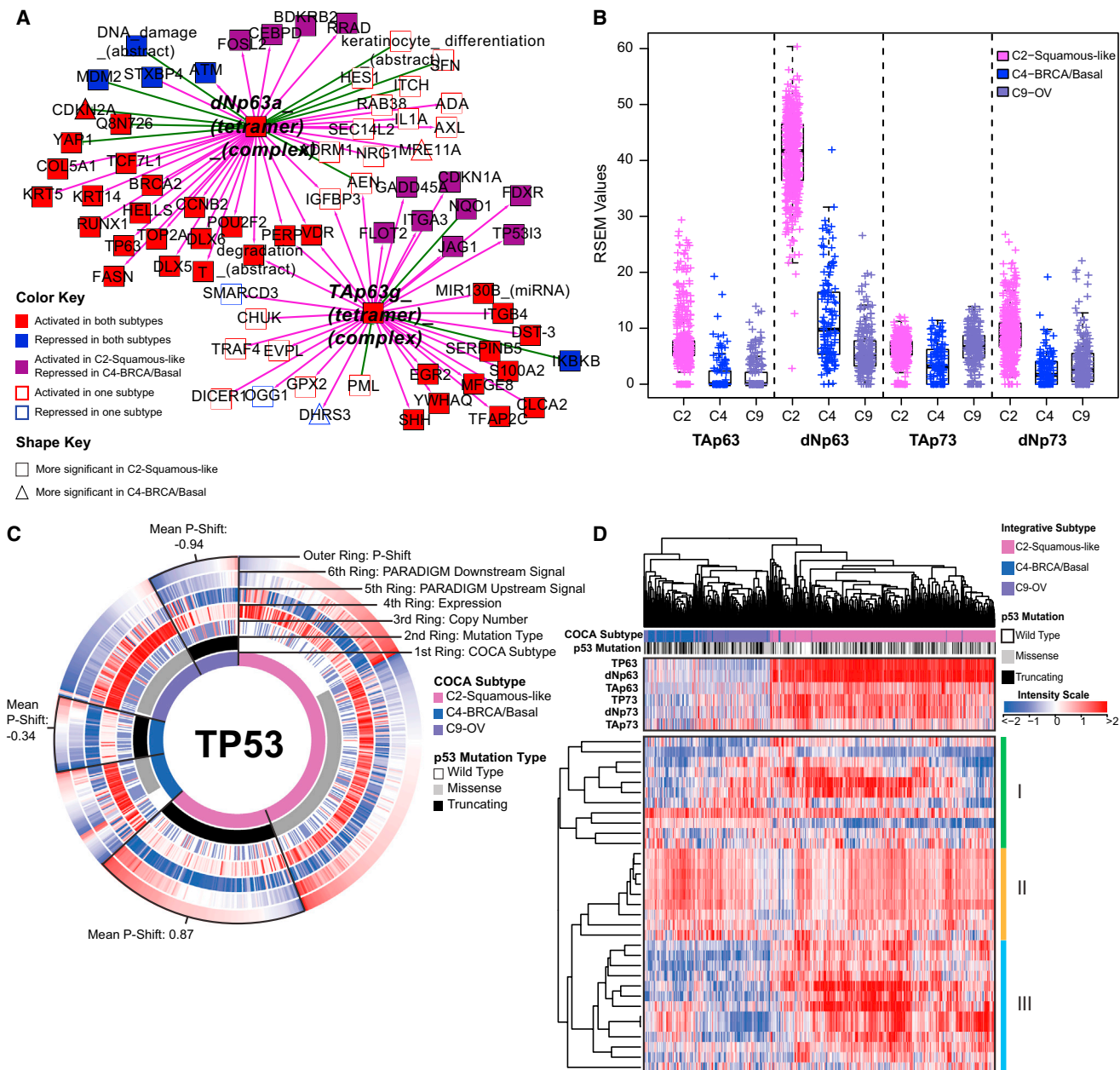


Figure 5. Comparison of Molecular Characteristics of C2-Squamous-like, C4-BRCA/Basal, and C9-OV Subtypes Reveals Differences in TP63 and TP53 Signaling

(A) Relative significance of TP63 network activation within the C2-squamous-like and C4-BRCA/basal subtypes. The network neighbors surrounding the TAp63 γ and Δ Np63 α tetramer complexes that show significant activation (or inactivation) within the C2-squamous-like and/or C4-BRCA/basal subtypes relative to all other cases were visualized using Cytoscape (Shannon et al., 2003). Node shape reflects relative significance in the one-versus-all comparison (square: more significant in C2-squamous-like, triangle: more significant in C4-BRCA/basal). Node color indicates relative activity (red: activated in C2 and C4, blue: inactivated in C2 and C4, purple: activated in C2 but inactivated in C4, white: activated or inactivated in only one subtype).

(B) Box plot of isoform-specific levels of TP63 and TP73 within three of the TP53-frequently mutated COCA subtypes (C2-squamous-like, C4-BRCA/basal, and C9-OV).

(C) CircleMap of PARADIGM-Shift differences associated with TP53 mutations within the C2, C4, and C9 COCA subtypes. Samples were ordered first by integrative subtype membership (innermost ring), then by TP53 mutation status (second ring), and finally by P-Shift (outer ring, indicating TP53 activity). The GISTIC score (indicating CNV), mRNA expression level, PARADIGM upstream and downstream activities are shown in the third, fourth, fifth and sixth rings, respectively. Red-blue color intensity reflects magnitude (red: positive, blue: negative). TP53-truncating mutants are highlighted (black outlined wedge), and the mean P-shift scores of the truncating mutants are shown. Negative P-Shift scores (outer ring blue) predict loss of function (LOF).

(legend continued on next page)

Genome Atlas Research Network, 2012c). We therefore asked if those subtypes share additional characteristics. The C9-OV (94%), C4-BRCA/basal (80%) and C2-squamous-like (72%) subtypes have the highest frequencies of *TP53* mutation. All three show a very high frequency of copy-number changes (Figure 2C), and all are significantly enriched with amplifications of 3q26 and 8q24/*cMYC* and losses of chromosomes 4q, 5q, 8p, and 18q (Figure 2B). The COCA subtypes share features common to a pan-cancer cluster identified by a parallel analysis of the transcriptional profiles of these same tumors (Martinez et al., 2014), which was found to be associated with genomic loss of *CDKN2A* (p16ARF), increased numbers of DNA double-strand breaks, high expression of cyclin B1, and upregulation of proliferation genes.

Consistent with our previous TCGA report noting the similarities between breast basal-like and serous ovarian cancers (Cancer Genome Atlas Research Network, 2012c), the copy-number profiles of the integrative subtypes place the C4-BRCA/basal subtype closest to the C9-OV subtype (Figure S4C); both are also near a cluster tree branch that contains C2-squamous-like and C8-BLCA. All six of those subtypes show *TP53* mutation and large-scale copy-number changes.

Pathway commonalities between the C4-BRCA/basal and C9-OV subtypes (Table S5B) largely recapitulate previous findings using PARADIGM analysis that both subtypes show activation of *cMYC* and *FOXM1*/proliferation signaling (Cancer Genome Atlas Research Network, 2012c). However, *HIF1A* signaling in those subtypes, despite previously being reported as high, appears less active in this pan-cancer context, probably due to the presence of other cancer types with clearly elevated *HIF1A* signaling (e.g., C5-KIRC). In terms of gene programs, C2-squamous-like tumors show high expression of the *basal signaling* gene program (*GP17*), at levels comparable with those in the C4-BRCA/basal tumors (Figure 3). In addition, both subtypes show upregulation of the *proliferation/DNA synthesis* gene program (*GP1*), as well as signatures of *TP53* mutation, *MYC* targets/*TERT*, *VEGF* signaling, and activation of the *PD1* and *CTLA4* immune costimulatory pathways (Table S4A, Figure 3). Indeed, principal components analysis showed that C2-squamous-like and C4-BRCA/basal tumors are the most similar COCA subtypes with regard to gene program/drug pathway expression (Figure S5B).

In line with those findings, a systematic search for PARADIGM pathway commonalities between the C2-squamous-like and C4-BRCA/basal tumors through the definition of a “basalness score” (Cancer Genome Atlas Research Network, 2012c) reveals shared activation of proliferation- and immune-related pathways. *TP63* network dysregulation is apparent in HNSC and LUSC (Figure S7C, Table S5), as found previously (Cancer Genome Atlas Research Network, 2012a; Walter et al., 2013). It has also been associated with normal basal stem/progenitor cell function in other organs (e.g., breast, urogenital tract

(Crum and McKeon, 2010). However, closer scrutiny of the network neighborhood surrounding the *TP63* g and *dNp63a* complexes reveals that *TP63* activation is more significant in the C2-squamous-like tumors than it is in the C4-BRCA/basals, and it involves a larger number of *TP63* network targets (Figure 5A). Indeed, *TP63* expression levels, in particular expression of the oncogenic Δ Np63 isoform, are significantly higher in the C2-squamous-like subtype than in the C4-BRCA/basal tumors (Figure 5B). Notably, we did not see *TP63* network activity or increased expression in the C9-OV subtype (Table S4A and Figure 6B).

High *TP53* mutation rates characterize several tumor types including those represented by the COCA subtypes C4-BRCA/basal, C9-OV, and C2-squamous-like (Table S2A). Surprisingly, our pathway and gene program analysis reveal a pattern of *TP53* compensation in the C2-squamous-like tumors that distinguishes them from these other subtypes with high *TP53* mutation rates. First, the C2-squamous-like tumors do not exhibit significant loss of PARADIGM-inferred *TP53* activity (Table S4A) and PARADIGM-SHIFT analysis (Ng et al., 2012) predicts loss-of-function of *TP53*-truncating mutations (observed in 43% of C4-BRCA/basal, 38% of C9-OV, and 30% of C2-squamous-like cases) at a significantly higher degree in the C4-BRCA/basal and C9-OV subtypes compared to the C2-squamous-like subtype (Figure 5C). Second, the copy-number data when aligned with *TP53* missense and truncating mutations reveals more loss of heterozygosity (LOH) in the C9-OV and C4-BRCA/basal than in the C2-squamous-like samples. The apparent higher *TP53*-pathway activity in C2-squamous-like tumors may be related to the expression of isoforms of related family members *TP63* and/or *TP73* (Figure 5B), which may compensate for *TP53* mutation in the C2-squamous-like tumors as revealed by PARADIGM-Shift analysis (Figure 5C), and as supported by functional experimental data in HNSC lines and tumors (Lu et al., 2011). In HNSC, the function of *TP63/73* in growth of HNSC is modulated in the presence of inflammatory factor *TNF- α* and *cREL*. Third, the transcriptional targets of *TP53* shared with *TP63/73* appear to be more highly expressed in the C2-squamous-like subtype than in the C9-OV or C4-BRCA/basal subtype (Figure S7D). Indeed, hierarchical clustering of 33 *TP53*-related gene signatures subsets the C2-squamous-like, C4-BRCA/basal and C9-OV tumors predominantly by subtype (left side dendrogram subtree: 99% C4-BRCA/basal/C9-OV; right-side dendrogram subtree: 98% C2-squamous-like) (Figure 5D). However, with the exception of the C4-BRCA/basal-like subtype, the levels of *TP53* activity were not predictive of overall survival when restricted to the analysis within a subtype. For the C4-BRCA/basal case, the PARADIGM-Shift scores do provide a moderate predictive degree when only the *TP53* truncating mutants are considered ($p < 0.05$). Interestingly, *TP63/73* compensatory function has been linked to cisplatin chemo-sensitivity and survival in BRCA1-related triple-negative

(D) Unsupervised clustering of C2-squamous-like, C4-BRCA/basal, and C9-OV cancers based on the expression patterns of 33 published *TP53*-related gene signatures. Sample subtype assignment (pink, C2-squamous-like, blue: C4-BRCA/basal; purple, C9-OV) and *TP53* mutation status (wild-type, white; truncating, black; missense, gray) are indicated in the column color bar. Heatmap red-blue color intensity reflects magnitude (red, positive; white, average; blue, negative). Three main clusters of correlated signatures were identified (denoted I, II, and III); see Data S4 (syn2491513) for complete list of signatures in each cluster. See also Figure S7 and Table S5.

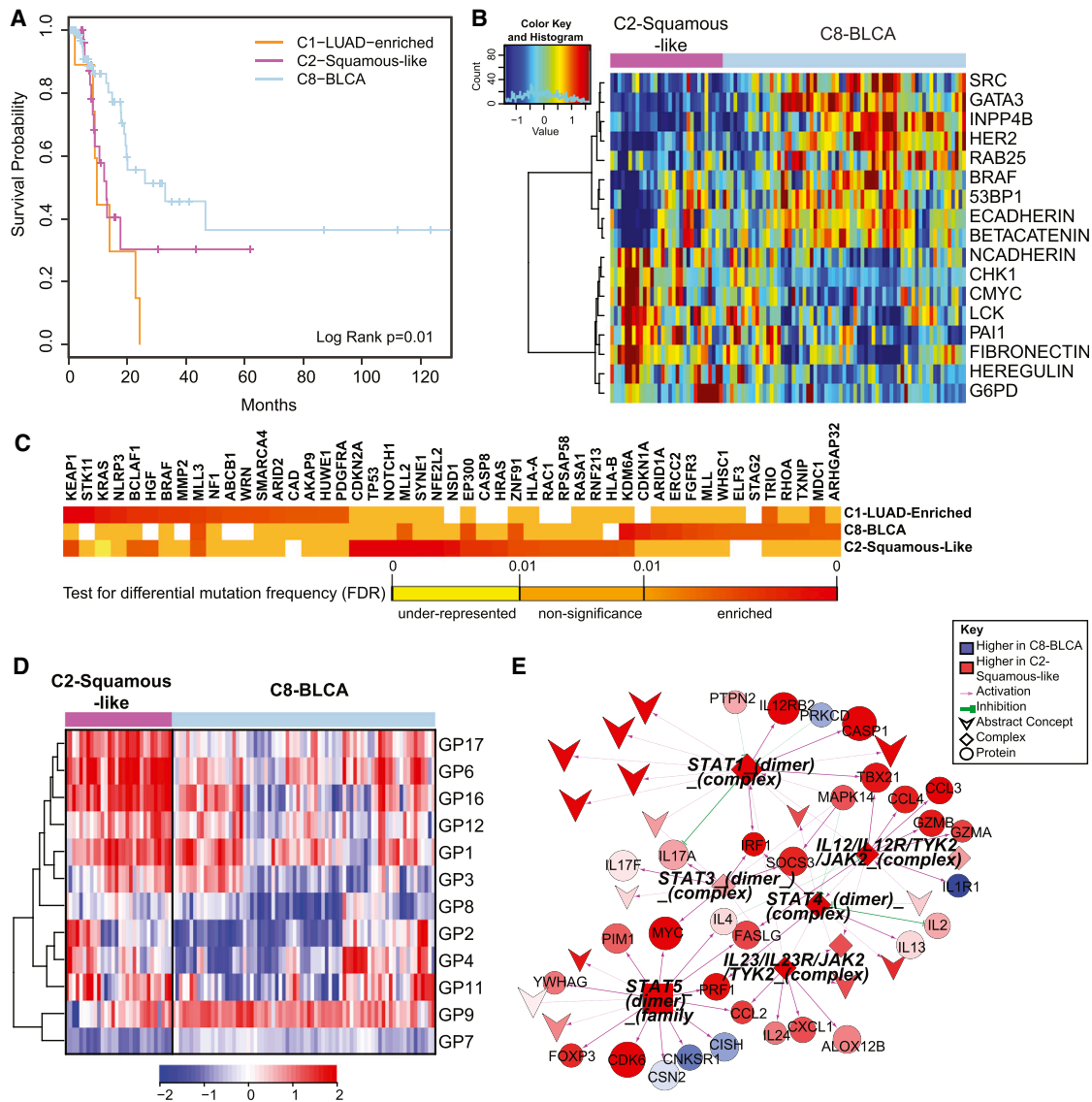


Figure 6. Divergence of the Bladder Cancer Samples across Multiple COCA Subtypes

(A) Kaplan-Meier survival analysis of bladder cancers within the C1-LUAD-enriched, C2-squamous-like, and C8-BLCA subtypes.

(B) Heatmap of 17 proteins expressed at significantly different levels within the C2-squamous-like relative to the C8-BLCA bladder cancer samples. Samples are arranged along the column by subtype (pink, C2; light blue, C8); and protein data are ordered along the rows by clustering. Rainbow color scale reflects magnitude (red, high; green, average; blue, low).

(C) HCD genes with differential mutation frequencies among the bladder samples clustered in COCA subtypes C1, C2, and C8. Differential frequencies reflect frequencies within, relative to frequencies outside of, the COCA subtype.

(D) Heatmap of 11 gene programs showing significant differential expression between the C2 and C8 bladder cancers. Samples are arranged along the column by subtype (pink, C2; light blue, C8), and gene programs are ordered along the rows by clustering. Red-blue color scale reflects magnitude (red, high; blue, low).

(E) PARADIMM subnetwork of immune-related pathway biomarkers activated in C2 bladder cancers relative to the C8 subtype. Red-blue color scale represents relative activation (red, higher in C2; blue, higher in C8). Node size reflects relative significance, and node shape denotes feature type (diamond, multiprotein complex; v-shaped, cellular process; circle, genes; square, gene family). Color of an edge reflects type of interaction within the PARADIMM SuperPathway (purple arrows, activation; green T, inhibition).

See also [Figure S8](#) and [Table S6](#).

breast cancers ([Leong et al., 2007](#)). These studies show the potential for p63/73 compensatory function for mutated or suppressed p53 in HNSCC and breast cancer, which has potential implications for targeted and standard therapy across these

malignancies. These data indicate that TP53/63/73 downstream activities are of potentially broader significance among the C2-squamous-like, C9-OV and C4-BRCA/basal subtypes, with similarly high TP53 mutation rates.

Divergence of Bladder Cancer Subtypes

Despite a relatively small sample size ($n = 120$), bladder cancer was one of the most diverse of the tumor types, with samples clustering into 7 of the 11 major COCA subtypes (Table S6). The majority of the samples fell into three main COCA groups: 10 in C1-LUAD-enriched, 31 in C2-squamous-like, and 74 in C8-BLCA. Correlation with histology showed that the bladder samples in the C2-squamous-like group did, indeed, have evidence of squamous features, although most in that subtype had less than 50% squamous differentiation upon review by a team of five urological pathologists. The genomic classifications are consistent with evidence for diverse squamous, adenocarcinoma and other variant histologies in bladder carcinoma (Willis et al., 2013). Because it is one of the most diverse tissue-of-origin tumor types in the Pan-Cancer-12 set, we looked at survival differences among the three main COCA groups of bladder cancers. Samples in the C2-squamous-like and C1-LUAD-enriched groups showed significantly worse overall survival than those in the C8-BLCA group (Figure 6A; Figure S8B). The same distinction held in proteomics-only analyses (Akbari et al., 2014), consistent with the worse overall survival of the other tumor types (LUAD, LUSC, and HNSC) that predominate in the C1-LUAD-enriched and C2-squamous-like subtypes.

We focused on the two larger subsets (C2-squamous-like and C8-BLCA) of bladder cancers, performing single-platform and integrated-platform comparisons. There are significant differences in copy number (Figure S4A), protein expression (Figure 6B), mutations (Figure 6C), gene programs (Figure 6D), and PARADIGM pathway networks (Figure 6E; Figure S8A). There is also a significant difference in 3p arm-level events; the C2-squamous-like subset shows the characteristic squamous-like pattern of 3p loss, whereas the C8-BLCA subtype does not (Figure 2B). Consistent with findings from the Pan-Cancer proteomics analysis (Akbari et al., 2014), higher HER2 and Rab25 protein levels are observed in the majority of the C8-BLCA cases relative to the C2-squamous-like bladder cases (Figure 6B). Conversely, markers of epithelial-to-mesenchymal transition (EMT) such as low E-cadherin, high fibronectin, and high N-cadherin expression are apparent in the C2-squamous-like bladder cancers (Figure 6B). Both gene program and PARADIGM analyses reveal differences in immune cell signatures; the bladder C2-squamous-like samples show higher levels of immune cell-associated signatures (Figure 6D and 6E). That difference, which has also been noted for lung squamous (Cancer Genome Atlas Research Network, 2012a) and breast basal-like cancers (Prat et al., 2010), could contribute to differences in outcome and suggest therapeutic targets.

DISCUSSION

This integrated multiplatform analysis of 12 cancer types provides independent and clinically relevant prognostic information above and beyond tumor stage and primary tissue-of-origin. Based on this study, one in ten cancer patients would be classified differently by this new molecular taxonomy versus our current tissue-of-origin tumor classification system. With respect to its therapeutic relevance, this proportion of potentially misclassified tumors is comparable to the rate of *EGFR*

mutations in unselected non-small cell lung cancers (Lynch et al., 2004; Paez et al., 2004) and *ERBB2* amplifications among all breast cancers (Cancer Genome Atlas Research Network, 2012c). If used to guide therapeutic decisions, this reclassification would affect a significant number of patients to be considered for nonstandard treatment regimens. In addition to identifying several new genomic and pathway insights between and within tissue-of-origin tumor types, this TCGA study provides a public resource compendium of individual and integrated data sets from six different “omic” platforms, comprehensively characterizing >3,500 tumors and enabling researchers to explore new questions and analytical approaches that will perpetuate this discovery process.

It is possible that each COCA subtype reflects tumors arising from distinct cell types. In this new taxonomy, cancers of non-epithelial origin (e.g., neural, muscle, connective tissue) appear most different from epithelial tumors based on virtually all molecular platforms. The next most marked difference is apparent between epithelial cancers arising from basal layer-like cells (C2-squamous-like and C4-BRCA/basal) and those with secretory functions (C1-LUAD-enriched and C3-BRCA/Luminal). Molecular commonalities within a COCA subtype suggest common oncogenic pathways. The C2-squamous-like cancers likely arise from a cellular subtype shared between environmentally exposed epithelial surfaces (e.g., oral cavity, lungs, and bladder); and malignancies from this cellular subtype possess a characteristic set of dysregulated genomic features, including *SOX2* and $\Delta Np63$ high expression (by 3q26-29 amplification) with *TP53* mutation. Although some of these pathway features have previously been reported for normal squamous tissue development and homeostasis (Crum and McKeon, 2010) and in squamous cell carcinomas of specific organ sites (Maier et al., 2011; Yang et al., 2011), they have not previously emerged collectively as a broad subtype-defining phenotype from an integrated genomic analysis of thousands of different tumors. Cancers in the C2-squamous-like subtype appear most similar to those in the C4-BRCA/basal subtype, which in turn show pathway similarities to those in the C9-ovarian. While all three COCA subtypes exhibit comparably high *TP53* mutation frequencies and expression of the *GP17_Basal signaling* gene program, the C2-squamous-like cancers are distinguished from all others by their significantly higher *TP63* and *TP73* expression, both short ($\Delta Np63$, $\Delta Np73$) and long (*TAp63*, *TAp73*) isoforms, which may partially compensate *TP53* mutation in this COCA subtype.

In this integrated analysis, bladder cancers (BLCA) emerged as the most heterogeneous of all Pan-Cancer-12 malignancies, with multiple samples falling into primarily three different integrated subtypes (C1-LUAD-enriched-like, C2-squamous-like and C8-BLCA). The clinical relevance of Pan-Cancer integrated subtyping is apparent in this divergent cohort of tumors. Survival is dependent on subtype membership, with the C2-squamous-like BLCA cases showing earlier mortality than the more common C8-BLCA cases (Figure 6A). Those BLCA cases in the C2-squamous-like subtype display immune features common to the C2-squamous-like subtype, that are pertinent to two areas of current interest among bladder cancer researchers: (1) epidemiologic and experimental evidence that

chronic cystitis and recurrent bladder infections (or other physical irritants) capable of inducing squamous metaplasia can predispose to squamous cancer of the bladder; and (2) the observation that early-stage bladder cancers are often responsive to intravesicular T cell induction by Bacillus-Calmette-Guerin (BCG) anti-TB vaccination. These findings strongly support a COCA subtype specific approach to postresection surveillance, adjuvant therapy and management of metastatic disease for bladder cancer patients.

Our results suggest that “cell-of-origin” rather than pathway-based features dominate the molecular taxonomy of diverse tumor types. There are several possible explanations for this observation. First and foremost, there are hundreds to thousands of features (mRNAs, proteins, microRNAs) with cell-type specific expression patterns, whereas pathways tend to regulate a much smaller subset of components—tens to hundreds of genes and their products. Second, pathways are often used in a cell-type-specific manner (e.g., APC-pathway in colon/rectum); therefore, pathway-based features are likely subsumed by a cell-of-origin-based classification. Further research is needed to uncover pathway dependencies within a cell-of-origin context, of which many such relationships already exist (i.e., EGFR in LUAD, BRAF in Melanoma, etc.).

In closing, the refined molecular taxonomy we describe builds on centuries of pathology and genetic research. The data sets and results have been collected into a unified resource on Synapse to support integrative bioinformatics analysis. To support navigation through these findings, the results have been made available through several portals including the UCSC Genome Browser, Gitools, and MD Anderson’s Next Generation Heatmaps (see the [Extended Experimental Procedures](#)). New methods to mine these data will enable “subtracting away” the dominant cell-of-origin signals to reveal information about pathway signaling and tumor microenvironments (e.g., stromal and immune components). This new taxonomy provides independent prognostic information above and beyond stage and tissue of origin, and further investigations may provide novel pathway-based insights with clues for personalizing therapy. Follow-up studies are needed to validate the findings reported here, and additional samples and tumor types will extend the integrated analysis. However, this initial PanCancer-12 analysis lays the groundwork for a richer classification of tumors into molecularly defined subtypes unlike all prior cancer classification systems.

EXPERIMENTAL PROCEDURES

Data for the complete set of 5,074 TCGA samples were obtained for the December 22, 2012 Pan-Cancer-12 data freeze from the Sage Bionetworks repository, Synapse. All data are made available through the Synapse website (<https://www.synapse.org>) and referenced with Synapse identifiers denoted as synN, where N provides a unique identifier within the Synapse system. All result files relevant for subtyping and downstream analyses are available from syn2468297.

Mutation Data and Predicted Driver Genes

Single-nucleotide variant calls for all samples in each of the 12 different tumor types were obtained from the official data freeze for each individual data type. Briefly, mutation calls were obtained from the separate TCGA working groups and processed to de-duplicate and re-annotate them using the ENSEMBLE

version 69 transcript database. The combined mutation annotation format (MAF) file is available from the Synapse resource.

One hundred and twenty-seven significantly mutated genes (SMGs) were identified in the entire sample set as those mutated more frequently than the background model according to MuSiC analysis as described in ([Kandoth et al., 2013a](#)). The SMG analysis was also performed by running MuSiC restricted to each COCA subtype. Lastly, genes whose mutations predominantly occur within a given COCA subtype were identified by using the list of high-confidence drivers retrieved by the combined analysis of several signals of positive selection, as described in ([Tamborero et al., 2013](#)).

Cluster-of-Cluster Assignments

Subtypes derived from each of the six platforms—mRNA-Seq, miRNA-Seq, reverse-phase protein arrays (RPPA), structural copy-number alterations (SCNA), DNA methylation, and somatic mutations—were calculated as described in the [Extended Experimental Procedures](#) section. Subtype calls for each of the six platforms were used to identify relationships among the different COCA subtypes and coded into a series of indicator variables for each subtype. The binary matrix was then used in the ConsensusClusterPlus R-package ([Wilkerson and Hayes, 2010](#)) to identify patterns of relationship among the samples. ConsensusClusterPlus was run with 80% sample resampling and 1,000 iterations of hierarchical clustering based on a Pearson correlation distance metric. More information on integrative subtyping analysis can be found in [Extended Experimental Procedures](#), Section 2. The integrated COCA subtypes are available on the Synapse resource.

Survival Analysis for Pan-Cancer-12 and Squamous Bladder Samples

Overall survival was calculated for samples using information from the enrollment and follow-up forms available at the DCC and downloaded on June 17, 2013. Kaplan-Meier survival plots were generated with the package Survival in R. A log-rank test was used to assess significance.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, eight figures, seven tables, and five data files and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.06.049>.

CONSORTIA

The members of The Cancer Genome Atlas Research Network are Rachel Abbott, Scott Abbott, B. Arman Aksoy, Kenneth Aldape, Adrian Ally, Samir Kumar Amin, Dimitris Anastassiou, J. Todd Auman, Keith A. Baggerly, Miruna Balasundaram, Saianand Balu, Stephen B. Baylin, Stephen C. Benz, Benjamin P. Berman, Brady Bernard, Ami S. Bhatt, Inanc Birol, Aaron D. Black, Tom Bodenheimer, Moiz S. Bootwalla, Jay Bowen, Ryan Bressler, Christopher A. Bristow, Angela N. Brooks, Bradley Broom, Elizabeth Buda, Robert Burton, Yaron S.N. Butterfield, Daniel Carlin, Scott L. Carter, Tod D. Casasent, Kyle Chang, Stephen Chanock, Lynda Chin, Dong-Yeon Cho, Juok Cho, Eric Chuah, Hye-Jung E. Chun, Kristian Cibulskis, Giovanni Ciriello, James Cleland, Melissa Cline, Brian Craft, Chad J. Creighton, Ludmila Danilova, Tanja Davidsen, Caleb Davis, Nathan D. Dees, Kim Delehaunty, John A. Demchok, Noreen Dhalla, Daniel DiCara, Huyen Dinh, Jason R. Dobson, Deepti Dodda, HarshaVardhan Doddapaneni, Lawrence Donehower, David J. Dooling, Gideon Dresdner, Jennifer Drummond, Andrea Eakin, Mary Edgerton, Jim M. Eldred, Greg Eley, Kyle Ellrott, Cheng Fan, Suzanne Fei, Ina Felau, Scott Frazer, Samuel S Freeman, Jessica Frick, Catrina C. Fronick, Lucinda L. Fulton, Robert Fulton, Stacey B. Gabriel, Jianjiong Gao, Julie M. Gastier-Foster, Nils Gehlenborg, Myra George, Gad Getz, Richard Gibbs, Mary Goldman, Abel Gonzalez-Perez, Benjamin Gross, Ranabir Guin, Preethi Gunaratne, Angela Hadjipanayis, Mark P. Hamilton, Stanley R. Hamilton, Leng Han, Yi Han, Hollie A. Harper, Psalm Haseley, David Haussler, D. Neil Hayes, David I. Heiman, Elena Helman, Carmen Helsel, Shelley M. Herbrich, James G. Herman, Toshinori Hinoue, Carrie Hirst, Martin Hirst, Robert A. Holt, Alan P. Hoyle, Lisa Iype, Anders Jacobsen, Stuart R. Jeffreys, Mark A. Jensen, Corbin D.

Jones, Steven J.M. Jones, Zhenlin Ju, Joonil Jung, Andre Kahles, Ari Kahn, Joelle Kalicki-Veizer, Divya Kalra, Krishna-Latha Kanchi, David W. Kane, Hoon Kim, Jaegil Kim, Theo Knijnenburg, Daniel C. Koboldt, Christie Kovar, Roger Kramer, Richard Kreisberg, Raju Kucherlapati, Marc Ladanyi, Eric S. Lander, David E. Larson, Michael S. Lawrence, Darlene Lee, Eunjung Lee, Semin Lee, William Lee, Kjong-Van Lehmann, Kalle Leinonen, Kristen M. Leraas, Seth Lerner, Douglas A. Levine, Lora Lewis, Timothy J. Ley, Haiyan I. Li, Jun Li, Wei Li, Han Liang, Tara M. Lichtenberg, Jake Lin, Ling Lin, Pei Lin, Wenbin Liu, Yingchun Liu, Yuexin Liu, Philip L. Lorenzi, Charles Lu, Yiling Lu, Lovelace J. Luquette, Singer Ma, Vincent J. Magrini, Harshad S. Mahadeshwar, Elaine R. Mardis, Adam Margolin, Marco A. Marra, Michael Mayo, Cynthia McAllister, Sean E. McGuire, Joshua F. McMichael, James Melott, Shaowu Meng, Matthew Meyerson, Piotr A. Mieczkowski, Christopher A. Miller, Martin L. Miller, Michael Miller, Richard A. Moore, Margaret Morgan, Donna Morton, Lisle E. Mose, Andrew J. Mungall, Donna Muzny, Lam Nguyen, Michael S. Noble, Houtan Noushmehr, Michelle O'Laughlin, Akinyemi I. Ojesina, Tai-Hsien Ou Yang, Brad Ozenberger, Angeliki Pantazi, Michael Parfenov, Peter J. Park, Joel S. Parker, Evan Paull, Chandra Sekhar Pedamallu, Todd Pihl, Craig Pohl, David Pot, Alexei Protopopov, Teresa Przytycka, Amie Radenbaugh, Nilsa C. Ramirez, Ricardo Ramirez, Gunnar Rättsch, Jeffrey Reid, Xiaojia Ren, Boris Reva, Sheila M. Reynolds, Suhni K. Rhie, Jeffrey Roach, Hector Rovira, Michael Ryan, Gordon Saksena, Sofie Salama, Chris Sander, Netty Santoso, Jacqueline E. Schein, Heather Schmidt, Nikolaus Schultz, Steven E. Schumacher, Jonathan Seidman, Yasin Senbabaoglu, Sahil Seth, Samantha Sharpe, Ronglai Shen, Margi Sheth, Yan Shi, Ilya Shmulevich, Grace O. Silva, Janae V. Simons, Rileen Sinha, Payal Sipahimalani, Scott M. Smith, Heidi J. Sofia, Artem Sokolov, Mathew G. Soloway, Xingzhi Song, Carrie Sougnez, Paul Spellman, Louis Staudt, Chip Stewart, Petar Stojanov, Xiaoping Su, S. Onur Sumer, Yichao Sun, Teresa Swatloski, Barbara Tabak, Angela Tam, Donghui Tan, Jiabin Tang, Roy Tarnuzzer, Barry S. Taylor, Nina Thiessen, Vestinn Thorsson, Timothy Triche Jr., David J. Van Den Berg, Fabio Vandin, Richard J. Varhol, Charles J. Vaske, Umadevi Veluvolu, Roeland Verhaak, Doug Voet, Jason Walker, John W. Wallis, Peter Waltman, Yunhu Wan, Min Wang, Wenyi Wang, Zhining Wang, Scot Waring, Nils Weinhold, Daniel J. Weisenberger, Michael C. Wendt, David Wheeler, Matthew D. Wilkerson, Richard K. Wilson, Lisa Wise, Andrew Wong, Chang-Jiun Wu, Chia-Chin Wu, Hsin-Ta Wu, Junyuan Wu, Todd Wylie, Liu Xi, Ruibin Xi, Zheng Xia, Andrew W. Xu, Da Yang, Liming Yang, Lixing Yang, Yang Yang, Jun Yao, Rong Yao, Kai Ye, Kosuke Yoshihara, Yuan Yuan, Alfred K. Yung, Travis Zack, Dong Zeng, Jean Claude Zenklusen, Hailei Zhang, Jianhua Zhang, Nianxiang Zhang, Qunyan Zhang, Wei Zhang, Wei Zhao, Siyuan Zheng, Jing Zhu, Erik Zmuda, Lihua Zou

AUTHOR CONTRIBUTIONS

K.A.H., C.Y., D.M.W., A.D.C., L.J.V.V., N.L.-B., P.W.L., L.D., G.B.M., C.C.B., C.M.P., and J.M.S. conceived and designed the experiments. K.A.H., C.Y., D.M.W., A.D.C., D.T., M.D.M.L., B.N., V.U., R.A., H.S., and B.J.R. developed key methodology. R.A., P.W.L., L.A.B., and G.B.M. acquired data sets. K.A.H., C.Y., D.M.W., A.D.C., D.T., A.C., S.N., M.D.L., B.N., M.D.M., V.U., C.K., R.A., H.S., N.L.-B., P.W.L., B.J.R., L.D., A.G.R., L.A.B., G.B.M., J.N.W., C.V.W., Z.C., E.A.C., C.C.B., C.M.P., and J.M.S. analyzed and interpreted data. J.Z., L.O., A.A.M., N.L.-B., R.A., J.N.W., and J.M.S. created data resources. K.A.H., C.Y., D.M.W., A.D.C., D.T., S.N., M.D.M.L., B.N., V.U., R.A., B.J.R., L.A.B., G.B.M., J.N.W., C.V.W., Z.C., E.A.C., C.C.B., C.M.P., J.M.S. drafted the manuscript or provided critical revisions. J.Z., C.C.B., C.M.P., and J.M.S. led and coordinated the project.

ACKNOWLEDGMENTS

We thank the thousands of patients and their loved ones who contributed materially, emotionally, and intellectually to this study. We thank D.A. Wheeler and M.L. Meyerson for scientific review of the work and M.P. Schroeder for help setting up Gitools. We thank K.R. Shaw, B.A. Ozenberger, H.J. Sofia, C.M. Hutter, and J.C. Zenklusen for administrative support. This work was supported by grants from the Chapman Foundation and Dell Foundation to J.N.W., a MD Anderson Physician Scientist Award to L.A.B., a Burroughs Welcome Career Award at the Scientific Interface to B.J.R., and support from

the following grants from the United States National Institutes of Health: K08 CA137153, K08 CA176561, P50 CA083639, R01 CA071468, R01 HG005690, R01 HG007069, R01CA180006, R21 CA155679, U01 CA168394, U24 CA143858, U24 CA143867-05, U24 CA143883, U24 CA143848, U24 CA143858, U24 CA143866, U54 CA112970, U24 CA143799, U24 CA143835, U24 CA143840, U24 CA143843, U24 CA143845, U24 CA143848, U24 CA143858, U24 CA143866, U24 CA143867, U24 CA143882, U24 CA143883, U24 CA144025, U54 HG003273, U54 HG003067, U54 HG003079, ZIA-DC-000073, ZIA-DC-000074, and P30 CA016672 for the MD Anderson CCSG Functional Proteomics Core.

Received: December 23, 2013

Revised: April 4, 2014

Accepted: June 23, 2014

Published: August 7, 2014

REFERENCES

- Akbani, R., Ng, P.K., Werner, H.M., Shahmoradgol, M., Zhang, F., Ju, Z., Liu, W., Yang, J.Y., Yoshihara, K., Li, J., et al. (2014). A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat Commun* 5, 3887.
- Bhattacharya, A., Roy, R., Snijders, A.M., Hamilton, G., Paquette, J., Tokuyasu, T., Bengtsson, H., Jordan, R.C., Olshen, A.B., Pinkel, D., et al. (2011). Two distinct routes to oral cancer differing in genome instability and risk for cervical node metastasis. *Clin. Cancer Res.* 17, 7024–7034.
- Cancer Genome Atlas Research Network (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455, 1061–1068.
- Cancer Genome Atlas Research Network (2011). Integrated genomic analyses of ovarian carcinoma. *Nature* 474, 609–615.
- Cancer Genome Atlas Research Network (2012a). Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 489, 519–525.
- Cancer Genome Atlas Research Network (2012b). Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487, 330–337.
- Cancer Genome Atlas Research Network (2012c). Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70.
- Cancer Genome Atlas Research Network (2013a). Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 499, 43–49.
- Cancer Genome Atlas Research Network (2013b). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 368, 2059–2074.
- Cancer Genome Atlas Research Network (2014). Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507, 315–322.
- Chung, C.H., Bernard, P.S., and Perou, C.M. (2002). Molecular portraits and the family tree of cancer. *Nat. Genet. Suppl.* 32, 533–540.
- Crum, C.P., and McKeon, F.D. (2010). p63 in epithelial survival, germ cell surveillance, and neoplasia. *Annu. Rev. Pathol.* 5, 349–371.
- Friedl, P., and Gilmour, D. (2009). Collective cell migration in morphogenesis, regeneration and cancer. *Nat. Rev. Mol. Cell Biol.* 10, 445–457.
- Grilley-Olson, J.E., Hayes, D.N., Moore, D.T., Leslie, K.O., Wilkerson, M.D., Qaqish, B.F., Hayward, M.C., Cabanski, C.R., Yin, X., Socinski, M.A., et al. (2013). Validation of interobserver agreement in lung cancer assessment: hematoxylin-eosin diagnostic reproducibility for non-small cell lung cancer: the 2004 World Health Organization classification and therapeutically relevant subsets. *Arch. Pathol. Lab. Med.* 137, 32–40.
- Herzog, A., Bian, Y., Vander Broek, R., Hall, B., Coupar, J., Cheng, H., Sowers, A.L., Cook, J.D., Mitchell, J.B., Chen, Z., et al. (2013). PI3K/mTOR inhibitor PF-04691502 antitumor activity is enhanced with induction of wild-type TP53 in human xenograft and murine knockout models of head and neck cancer. *Clin. Cancer Res.* 19, 3808–3819.
- Hofree, M., Shen, J.P., Carter, H., Gross, A., and Ideker, T. (2013). Network-based stratification of tumor mutations. *Nat. Methods* 10, 1108–1115.

- Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., et al. (2013a). Mutational landscape and significance across 12 major cancer types. *Nature* 502, 333–339.
- Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R., Benz, C.C., et al.; Cancer Genome Atlas Research Network (2013b). Integrated genomic characterization of endometrial carcinoma. *Nature* 497, 67–73.
- Leong, C.O., Vidnovic, N., DeYoung, M.P., Sgroi, D., and Ellisen, L.W. (2007). The p63/p73 network mediates chemosensitivity to cisplatin in a biologically defined subset of primary breast cancers. *J. Clin. Invest.* 117, 1370–1380.
- Lu, H., Yang, X., Duggal, P., Allen, C.T., Yan, B., Cohen, J., Nottingham, L., Romano, R.A., Sinha, S., King, K.E., et al. (2011). TNF- α promotes c-REL/ Δ Np63 α interaction and TAp73 dissociation from key genes that mediate growth arrest and apoptosis in head and neck cancer. *Cancer Res.* 71, 6867–6877.
- Lynch, T.J., Bell, D.W., Sordella, R., Gurubhagavatula, S., Okimoto, R.A., Brannigan, B.W., Harris, P.L., Haserlat, S.M., Supko, J.G., Haluska, F.G., et al. (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* 350, 2129–2139.
- Maier, S., Wilbertz, T., Braun, M., Scheble, V., Reischl, M., Mikut, R., Menon, R., Nikolov, P., Petersen, K., Beschoner, C., et al. (2011). SOX2 amplification is a common event in squamous cell carcinomas of different organ sites. *Hum. Pathol.* 42, 1078–1088.
- Martinez, E., Yoshihara, K., Kim, H., Mills, G.B., Trevino, V., and Verhaak, R.G. (2014). Comparison of gene expression patterns across twelve tumor types identifies a cancer supercluster characterized by TP53 mutations and cell cycle defects. *Oncogene*. Published online August 4, 2014. <http://dx.doi.org/10.1038/onc.2014.216>.
- Ng, S., Collisson, E.A., Sokolov, A., Goldstein, T., Gonzalez-Perez, A., Lopez-Bigas, N., Benz, C., Haussler, D., and Stuart, J.M. (2012). PARADIGM-SHIFT predicts the function of mutations in multiple cancers using pathway impact analysis. *Bioinformatics* 28, i640–i646.
- Nielsen, T.O., Parker, J.S., Leung, S., Voduc, D., Ebbert, M., Vickery, T., Davies, S.R., Snider, J., Stijleman, I.J., Reed, J., et al. (2010). A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin. Cancer Res.* 16, 5222–5232.
- Paez, J.G., Jänne, P.A., Lee, J.C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F.J., Lindeman, N., Boggon, T.J., et al. (2004). EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 304, 1497–1500.
- Prat, A., Parker, J.S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J.I., He, X., and Perou, C.M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* 12, R68.
- Prat, A., Adamo, B., Cheang, M.C., Anders, C.K., Carey, L.A., and Perou, C.M. (2013). Molecular characterization of basal-like and non-basal-like triple-negative breast cancer. *Oncologist* 18, 123–133.
- Scagliotti, G.V., Parikh, P., von Pawel, J., Biesma, B., Vansteenkiste, J., Manegold, C., Serwatowski, P., Gatzemeier, U., Digumarti, R., Zukin, M., et al. (2008). Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naïve patients with advanced-stage non-small-cell lung cancer. *J. Clin. Oncol.* 26, 3543–3551.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
- Tamborero, D., Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Kandoth, C., Reimand, J., Lawrence, M.S., Getz, G., Bader, G.D., Ding, L., and Lopez-Bigas, N. (2013). Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci Rep* 3, 2650.
- Travis, W.D., Brambilla, E., Noguchi, M., Nicholson, A.G., Geisinger, K.R., Yatabe, Y., Beer, D.G., Powell, C.A., Riely, G.J., Van Schil, P.E., et al. (2011). International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J. Thorac. Oncol.* 6, 244–285.
- Troester, M.A., Herschkowitz, J.I., Oh, D.S., He, X., Hoadley, K.A., Barbier, C.S., and Perou, C.M. (2006). Gene expression patterns associated with p53 status in breast cancer. *BMC Cancer* 6, 276.
- Vandin, F., Clay, P., Upfal, E., and Raphael, B.J. (2012). Discovery of mutated subnetworks associated with clinical data in cancer. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing*, 55–66.
- Vaske, C.J., Benz, S.C., Sanborn, J.Z., Earl, D., Szeto, C., Zhu, J., Haussler, D., and Stuart, J.M. (2010). Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* 26, i237–i245.
- Walter, V., Yin, X., Wilkerson, M.D., Cabanski, C.R., Zhao, N., Du, Y., Ang, M.K., Hayward, M.C., Salazar, A.H., Hoadley, K.A., et al. (2013). Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes. *PLoS ONE* 8, e56823.
- Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26, 1572–1573.
- Willis, D.L., Porten, S.P., and Kamat, A.M. (2013). Should histologic variants alter definitive treatment of bladder cancer? *Curr. Opin. Urol.* 23, 435–443.
- Yang, X., Lu, H., Yan, B., Romano, R.A., Bian, Y., Friedman, J., Duggal, P., Allen, C., Chuang, R., Ehsanian, R., et al. (2011). Δ Np63 versatily regulates a Broad NF- κ B gene program and promotes squamous epithelial proliferation, migration, and inflammation. *Cancer Res.* 71, 3688–3700.