Comprehensive molecular characterization of urothelial bladder carcinoma

The Cancer Genome Atlas Research Network*

Urothelial carcinoma of the bladder is a common malignancy that causes approximately 150,000 deaths per year worldwide. So far, no molecularly targeted agents have been approved for treatment of the disease. As part of The Cancer Genome Atlas project, we report here an integrated analysis of 131 urothelial carcinomas to provide a comprehensive landscape of molecular alterations. There were statistically significant recurrent mutations in 32 genes, including multiple genes involved in cell-cycle regulation, chromatin regulation, and kinase signalling pathways, as well as 9 genes not previously reported as significantly mutated in any cancer. RNA sequencing revealed four expression subtypes, two of which (papillary-like and basal/squamous-like) were also evident in microRNA sequencing and protein data. Whole-genome and RNA sequencing identified recurrent in-frame activating FGFR3–TACC3 fusions and expression or integration of several viruses (including HPV16) that are associated with gene inactivation. Our analyses identified potential therapeutic targets in 69% of the tumours, including 42% with targets in the phosphatidylinositol-3-OH kinase/AKT/mTOR pathway and 45% with targets (including ERBB2) in the RTK/SoCK pathway. Chromatin regulatory genes were more frequently mutated in urothelial carcinoma than in any other common cancer studied so far, indicating the future possibility of targeted therapy for chromatin abnormalities.

Urothelial carcinoma of the bladder is a major cause of morbidity and mortality worldwide, causing an estimated 150,000 deaths per year. Previous studies have identified multiple regions of somatic copy number alteration, including amplification of PPARG, E2F3, EGFR, CCND1 and MDM2, as well as loss of CDKN2A and RB1 (refs 2, 3). Sequencing of candidate pathways has identified recurrent mutations in TP53, FGFR3, PIK3CA, TSC1, RB1 and HRAS (refs 2, 3). Whole-exome sequencing of nine bladder cancers, followed by a replication analysis of 88 cancers, identified mutations at >10% frequency in several chromatin remodelling genes: KDM6A, CREBBP, EP300 and ARIDIA (ref. 4). Focused molecular analyses5–8 have delineated tumour subtypes and identified kinase-activating FGFR3 gene fusions5,6.

We report here a comprehensive, integrated study of 131 high-grade muscle-invasive urothelial bladder carcinomas as part of The Cancer Genome Atlas (TCGA) project. Included are data on DNA copy number, somatic mutation, messenger RNA and microRNA (miRNA) expression, protein and phosphorylated protein expression, DNA methylation, transcript splice variation, gene fusion, viral integration, pathway perturbation, clinical correlates and histopathology to characterize the molecular landscape of urothelial carcinoma. This study identifies a number of mutations and regions of copy number variation that involve genes not previously reported as altered in a significant fraction of bladder cancers. It also identifies potential therapeutic targets in most of the samples analysed.

Demographic, clinical and pathological data

Samples (from 19 tissue source sites) consisted of 131 chemotherapy-naive, muscle-invasive, high-grade urothelial tumours (T2–T4a, Nx, Mx), as well as peripheral blood (n = 118) and/or tumour-adjacent, histologically normal-appearing bladder tissue (n = 23). Cases were retained only if they met the following criteria: tumour nuclei constituted ≥60% of all nuclei; tumour necrosis was ≤20% of the specimen; and variant histologies (squamous or small cell) were ≤50% (Supplementary Information, section ‘Biospecimen collection and clinical data’). Clinical and demographic characteristics are described in Supplementary Data 1.1. Five expert genitourinary pathologists re-reviewed all of the cases for multiple parameters, including the extent of variant histology (Supplementary Fig. 1.1a and Supplementary Information, section ‘Biospecimen collection and clinical data’).

Somatic DNA alterations

The tumours displayed a large number of DNA alterations, slightly fewer than in lung cancer and melanoma, but more than in other adult malignancies studied by TCGA (Fig. 1). On average, there were 302 exonic mutations, 204 segmental alterations in genomic copy number and 22 genomic rearrangements per sample. We analysed somatic copy number alterations (CNAs) using both SNP 6.0 arrays and low-pass whole-genome sequencing; the two were strongly concordant (Supplementary Methods 6.1 and Supplementary Fig. 6.1). There were 22 significant arm-level copy number changes (Supplementary Data 6.1.1), and GISTIC (genomic identification of significant targets in cancer) (Supplementary Methods 6.2) identified 27 amplified and 30 deleted recurrent focal somatic CNAs (Supplementary Data 6.2.1 and 6.3.1). Focal amplifications involved genes previously reported to be altered in bladder cancer (Fig. 1c and Supplementary Fig. 6.2.1) and some not previously implicated. The latter included PVR1A, BCL2L1 and ZNF703. The most common recurrent focal deletion, seen in 47% of samples, contained CDKN2A (9p21.3) and correlated with reduced expression (Fig. 1 and Supplementary Fig. 2.7). Other focal deletions containing <10 genes appeared to target PDE4D, RB1, FHIT, CREBBP, IKZF2, FOXQ1, FAM190A (also called CCSE1), LRPIB and WWOX.

Whole-exome sequencing of 130 tumours and matched normal samples targeted 186,260 exons in 18,091 genes (mean coverage 100-fold, with 82% of target bases covered >30×). MuTect identified 39,312 somatic mutations (including 38,012 point mutations and 1,138 indels (insertions or deletions)), yielding mean and median somatic mutation rates of 7.7 and 5.5 per megabase (Mb), respectively (Fig. 1a and Supplementary Table 2.1.1). Thirty-two genes showed statistically significant levels of recurrent somatic mutation (Fig. 1b and Supplementary Table 2.1.2) by analysis using MutSig 1.5 (refs 9, 11) (Supplementary

*A list of authors and affiliations appears at the end of the paper.
Methods 2.2). Three other genes identified by MutSig were not considered further because of low or undetectable expression (Supplementary Fig. 2.1.1). A similar analysis considering only mutations in the COSMIC database identified three more significantly mutated genes: ERBB2, ATM and CTNNB1 (Supplementary Table 2.1.3). We validated the mutation findings in three ways: targeted re-sequencing of all genes in any other TCGA cancer type or reported in another study as opposed to the single gene listed here, RNA expression level for selected genes, expressed as fold change from the median value for all samples. Tumour samples were grouped into three clusters (red, blue and green) using consensus NMF clustering (see the main text and Supplementary Fig. 2.1.2). Three samples with no copy number data and two samples with no mutations in the genes were not used in the clustering and are shown in grey.

Nearly half (49%) of the samples had TP53 mutations (Fig. 1b), which were mutually exclusive in their relationship with amplification (9%) and overexpression (29%) of MDM2; hence, TP53 function was inactivated in 76% of samples. Most RB1 mutations were inactivating, were associated with significantly reduced mRNA level (Supplementary Fig. 2.7) and were mutually exclusive with CDKN2A deletions (Supplementary Fig. 2.8 and Supplementary Table 2.8.1). FGFR3 mutations (12%) typically affected known kinase-activating sites. PIK3CA mutations were relatively common (20%), clustering in the helical domain near E545 (Supplementary Fig. 2.4). Most TSC1 mutations (8%) were truncating, and six were homozygous (allele fraction > 0.5).

Many of the 32 genes identified in Fig. 1b have not previously been reported as statistically significantly mutated in bladder cancer: MLL2 (also called KMT2D; 27%), CDKN1A* (14%), ERCC2* (12%), STAG2 (11%), RXRA* (9%), ELF3* (8%), NFE2L2 (8%), KLF5* (8%), TXNIP (7%), FOXO1* (5%), RHOB* (5%), FOXA1 (5%), PAIP1* (5%), BTG2* (5%), ZFP36L1 (5%), RHOA (4%) and CCND3 (4%). The nine genes marked with asterisks have not been reported as significantly mutated genes in any other TCGA cancer type or reported in another study as opposed to the single gene listed here.

Figure 1 | The genomic landscape of bladder cancer. a. Mutation rate and type, histological subtype, smoking status, gender, tumour stage and cluster type. b. Genes with statistically significant levels of mutation (MutSig, false discovery rate < 0.1) and mutation types. c. Deletions and amplifications for genomic regions with statistically significant fold copy number changes (GISTIC2.0). 'Copy number' refers to absolute copy number. d. RNA expression level for selected genes, expressed as fold change from the median value for all samples. Tumour samples were grouped into three clusters (red, blue and green) using consensus NMF clustering (see the main text and Supplementary Fig. 2.1.2). Three samples with no copy number data and two samples with no mutations in the genes were not used in the clustering and are shown in grey.
showed markedly increased expression of genes involved in genotoxic metabolism and the reactive oxygen species (ROS) response (Supplementary Figs 2.5.1–2.5.3 and Supplementary Data 2.5.2). Furthermore, nine samples had mutations in redox regulator TXNIP (ref. 16) (five of them inactivating) and were mutually exclusive of samples with NFE2L2 mutations, providing another mechanism for dysregulation of redox metabolism. Predominant inactivating mutations were seen in STAG2, an X-linked cohesin complex component required for separation of sister chromatids during cell division13 (Supplementary Fig. 2.4).

Unsupervised clustering by non-negative matrix factorization of mutations and focal somatic CNAs in 125 samples identified three distinct groups (Fig. 1a and Supplementary Fig. 2.1.2). Group A (red), classified as 'locally amplified', is highly enriched in focal somatic CNAs in several genes, as well as mutations in MLL2 (Fig. 1 and Supplementary Tables 2.1.4 and 2.1.5). Group B (blue), classified as 'papillary CDKN2A-deficient FGFR3 mutant', is enriched in papillary histology. Nearly all group B samples show loss of CDKN2A, and most have one or more alterations in FGFR3. Group C (green), classified as 'TP53/cell-cycle-mutant', shows TP53 mutations in nearly all samples, as well as enrichment with RB1 mutations and amplifications of E2F3 and CCNE1 (Fig. 1 and Supplementary Table 2.1.4). These differences in pattern of mutation suggest the possibility of different oncogenic mechanisms.

Seventy-two per cent of the cancers in this study were from current or past smokers, consistent with extensive epidemiological studies indicating an association between smoking and urothelial cancer risk. In contrast with lung cancer, however, there was no statistically significant association between smoking status and the mutational spectrum, frequency of mutation in any significantly mutated gene, occurrence of focal somatic CNAs or expression subtype (Supplementary Tables 2.9.1 and 2.9.2). Never-smokers did have a slightly higher fraction of C>G mutations than did current/former smokers (28.5% versus 23.8%, P = 0.032; Supplementary Figs 2.3.2 and 2.3.3). Unsupervised clustering of promoter CpG island DNA methylation data revealed a major subgroup (34%) of tumours (CIMP) characterized by cancer-specific DNA hypermethylation (Supplementary Fig. 7.1). Multivariate regression analysis with age, sex and tumour stage as covariates identified smoking pack-years as the only significant predictor of CIMP phenotype, as has also been reported for colorectal cancer16.

Fifty-one per cent of mutations overall were Tp*T/C- (T/G) (Supplementary Table 2.1.1), a class of mutation recently reported to be mediated by one of the DNA cytosine deaminases, APOBEC (refs 19, 20). APOBEC3B was expressed at high levels in all of the tumours, suggesting a major role for APOBEC-mediated mutagenesis in bladder carcinogenesis (Supplementary Figs 12.1 and 12.2).

Four genes involved in epigenetic regulation were significantly mutated genes: MLL2, ARID1A, KDM6A and EP300 (Fig. 1). Truncating mutations were significantly enriched in each of those genes (Supplementary Fig. 2.2 and Supplementary Data 2.2.1–2.2.2). Three of the genes had previously been identified as mutated in uterine carcinomas8, but mutation of MLL2, which encodes a histone H3 lysine 4 (H3K4) methyltransferase, is a novel finding. Several other chromatin-regulating genes had mutation rates ≥10% but were not statistically significant by MutSig analysis: ML3, MLL, CREBBP, CHD7 and SRCAP. Many other epigenetic regulators were mutated at lower frequency but were also enriched with truncating mutations, indicating functional significance (Supplementary Fig. 2.2 and Supplementary Data 2.2.1 and 2.2.2). Non-silent mutations in chromatin regulatory genes overall were significantly enriched in bladder cancer in comparison with the entire exome, in contrast with all other epithelial cancers studied so far in the TCGA project (Supplementary Table 2.10). Mutations in MLL2 and KDM6A (the latter encoding a histone H3 lysine 27 (H3K27) demethylase) were mutually exclusive (Supplementary Fig. 2.8 and Supplementary Table 2.8.1), suggesting that mutations in the two genes have redundant downstream effects on carcinogenesis or that the combined loss is synthetically lethal.

Chromosomal rearrangements and viral integration

To identify structural rearrangements and pattern sequences, we used low-pass, paired-end, whole-genome sequencing (WGS; 6–8× coverage) of 114 tumours and RNA sequencing of all tumours. We detected 2,529 structural aberrations, including 1,153 that involve gene–gene fusions. Among the translocations, 379 were inter-chromosomal, 237 were intra-chromosomal, 274 were the result of inversions and 263 resulted from deletions (Supplementary Table 3.1). We found several recurrent translocations of probable pathogenic significance, including an intra-chromosomal translocation on chromosome 4 involving FGFR3 and TACC3 (n = 3). The breakpoints were in intron 16 (two cases) or exon 17 (one case) of FGFR3 and intron 10 of TACC3 (confirmed by DNA sequencing and RNA-seq). All three lead to fusion mRNA products for which the predicted proteins include the amino-terminal 758 amino acids of FGFR3 fused with the carboxy-terminal 191 amino acids of TACC3 (Fig. 2a). On the basis of the structure of the FGFR3–TACC3 fusion protein, we predict that it can auto-dimerize, leading to constitutive activation of the kinase domain of FGFR3. FGFR3–TACC3 fusion, which was recently described in both glioblastoma17a and bladder cancer17, represents a promising therapeutic target. The ERBB2 gene was also involved in translocations in four tumours, all with different fusion partners and all confirmed by DNA sequencing, RNA-seq or both. In one case, exons 4 to 29 of ERBB2 were fused to the promoter plus exon 1 of DIP2B, and the fusion product was amplified (Fig. 2b). Two other fusion products resulted in novel mRNA products, the biological significance of which is not known.

We identified viral DNAs in 7 of 122 tumours (6%), and viral transcripts in 5 of 122 (4%). Three tumours expressed cytomegalovirus (CMV) transcripts (encoding RL5A, RNA2.7, RL9A, RNA1.2, UL5 and UL22A), one expressed BK polyoma virus and one expressed human papilloma virus 16 (HPV16). HPV16 and human herpesvirus 6B DNA were each identified in one other sample but without expression. None of the tumours expressing CMV showed evidence of CMV integration into the host genome, suggesting the presence of a stable episome. In the BK-positive tumour, two BK genes were integrated into GRB14, a signalling adaptor protein for receptor tyrosine kinases. In the HPV-16-expressing case, the virus integrated into BCL2L1, an apoptosis-regulating gene (Fig. 2c). In that tumour, BCL2L1 was amplified (~6×) and...

Figure 2 | Structural rearrangements and viral integration. a, FGFR3–TACC3 fusion in sample TCGA-CA-FASMH showing the breakpoints in the two genes, the breakpoint junction sequences and the predicted fusion protein. b, Rearrangement involving DIP2B and ERBB2 in TCGA-DK-A216. The ERBB2 gene has swapped its promoter with that of DIP2B, resulting in overexpression of ERBB2. c, Insertion of human papilloma virus 16 (HPV16) into the BCL2L1 gene on chromosome 20 in TCGA-GC-A316. The region of BCL2L1 into which the virus has integrated and the integration junction sequence are shown.
overexpressed (≈10× median; >2× any of the other samples). Overall, these findings indicate that viral infection may have a role in the development of a small percentage of urothelial carcinomas.

**mRNA, miRNA and protein expression**

Analysis of RNA-seq data from 129 tumours identified four clusters (clusters I-IV) (Fig. 3 and Supplementary Fig. 4.1). Cluster I (‘papillary-like’) is enriched in tumours with papillary morphology (P = 0.0002), FGFR3 mutations (P = 0.0007, q = 0.02), FGFR3 copy number gain (P = 0.04, q = 0.1) and elevated FGFR3 expression (P < 0.0001) (Fig. 3a). It includes all three samples with FGFR3–TACC3 fusions. Cluster I samples also show significantly lower expression of miR-99a and miR-100, miRNAs that downregulate FGFR3 expression (P = 0.0002, Figs 3a and Supplementary Fig. 5.3). Cluster I samples also show lower expression of miR-145 and miR-125b, which have been reported as frequently downregulated in bladder cancer. Tumours with FGFR3 alterations, and perhaps other tumours that share the cluster I expression profile, may respond to inhibitors of FGFR or its downstream targets.

Reverse-phase protein array (RPRA) data indicate that clusters I and II express high HER2 (ERBB2) levels and an elevated oestrogen receptor beta (ESR2) signalling signature, indicating potential targets for hormone therapies such as tamoxifen or raloxifene (Fig. 3d). In fact, HER2 protein levels in a subset of the tumours are comparable to breast cancers (Fig. 3d). Therefore, urothelial bladder cancers (and other cancer types) may prove sensitive to inhibition of glycolysis or related metabolic pathways.

**Pathway analysis and therapeutic targeting**

Integrated analysis of the mutation and copy-number data revealed three main pathways as frequently dysregulated in bladder cancer: cell cycle regulation (altered in 93% of cases); kinase and phosphatidylinositol-3-OH kinase (PI(3)K) signalling (72%); and chromatin remodelling, including mutations/somatic CNAs in histone-modifying genes (89%) and components of the SWI/SNF nucleosome remodelling complex (64%) (Fig. 4a). To complement these results for well-defined pathways, we applied network analysis methods to examine other possible interactions between genes and pathways (Fig. 4b). In particular, we used the TieDIE algorithm to search for causal regulatory interactions within the PARADIGM network, which connects mutated genes to active transcriptional hubs.

The analysis identified a sub-network linking mutated histone-modifying genes to a large array of activated transcription factors, indicating potential far-reaching effects of histone modification on other pathways (Supplementary Fig. 8.2.1) converging on MYC/MAX regulation. Both MYC and MAX showed differential activity of transcription factors FOXA2 and SP1, implicating de-differentiation processes as a result of the mutations. Our network analysis also identified HSP90AA1 as a critical signalling hub, indicating that inhibitors of HSP90 may have therapeutic value in urothelial carcinoma. Although the linkages between mutations and transcriptional changes were statistically significant in terms of their proximity in the network (as determined by permutation tests; see Supplementary Fig. 8.2), further studies will be needed to assess the biological relevance of the findings.
Integrated analysis also identified mutations, copy number alterations or RNA expression changes affecting the PI(3)K AKT/mTOR pathway in 42% of the tumours (Fig. 5a). Included were activating mutations in PIK3CA (17%; potentially responsive to PI(3)K inhibitors), mutation or deletion of TSC1 or TSC2 (9%; potentially responsive to mTOR inhibitors) and overexpression of AKT3 (10%; potentially responsive to AKT inhibitors). We also observed mutations, genomic amplifications or gene fusions that affect the RTK/RAS pathway in 44% of the tumours (Fig. 5b, c). Included were events that can activate FGFR3 (17%; potentially responsive to FGFR inhibitors or antibodies), amplification of EGFR (9%; potentially responsive to EGFR antibodies or inhibitors), mutations of ERBB3 (6%; potentially responsive to HER2 inhibitors) and ERBB2 (16%; potentially responsive to HER2 inhibitors).

Figure 4 | Altered pathways and networks in bladder cancer. a, Somatic mutations and copy number alterations (CNA) in components of the p53/Rb pathway, RTK/RAS/PI(3)K pathway, histone modification system and SWI/SNF complex. Red, activating genetic alterations; blue, inactivating genetic alterations. Percentages shown denote activation or inactivation of at least one allele. The network connecting mutated histone-modifying genes to transcription factors with differential activity (methodology and larger implicated network in Supplementary Fig. 8.2.1). Each gene is depicted as a multi-ring circle with various levels of data, plotted such that each ‘spoke’ in the ring represents a single patient sample (same sample ordering for all genes), ‘PARADIGM’ ring, bioinformatically inferred levels of gene activity (red, higher activity); ‘Transcriptional activity’, mean mRNA levels of all of the targets of each transcription factor; ‘expression’, mRNA levels relative to normal (red, high); ‘Mutation in gene’, somatic mutation; ‘Mutation in histone modifier genes’, somatic mutation in at least one such gene; ‘IPL anti-correlation’, genes with PARADIGM integrated pathway levels (IPLs) inversely correlated with histone-gene mutation status. Gene–gene relationships are inferred using public resources.

Figure 5 | Potential targets in bladder cancer. a, Alterations in the PI(3)K AKT/mTOR pathway are mutually exclusive. Tumour samples are shown in columns; genes in rows. Only samples with at least one alteration are shown. AKT3 shows elevated expression in 10% of samples, independent of copy number (right panel). Hetloss, heterozygous loss. b, Receptor tyrosine kinases are altered, by any of several different mechanisms (amplification, mutation or fusion), in 45% of samples. Only mutations that are recurrent in this data set or previously reported in COSMIC are shown. c, Recurrent mutations in ERBB2 and ERBB3. The mutations shown in black are either recurrent in the TCGA data set or reported in COSMIC. Green, receptor L domain; red, furin-like cysteine-rich region; blue, growth factor receptor domain IV; yellow, tyrosine kinase domain. d, ERBB2 amplifications and recurrent mutations in other cancers profiled by TCGA. Missense mutations were counted in the following positions: G309, S310, L313, R678, T733, L755, V777, D769, V842, T862, R896 and M916F. In-frame insertions were counted between amino acids 774 and 776. Only tumour types with an alteration frequency ≥2% are shown.
sensitive to ERBB kinase inhibitors) and mutation or amplification of ERBB2 (9%; potentially sensitive to ERBB2 kinase inhibitors or antibodies). ERBB3 mutations in bladder cancer have been noted previously, but statistically significant mutation of ERBB2 in bladder cancer has not been reported. Both genes are potential therapeutic targets in other diseases. Notably, ERBB2 alterations were approximately as frequent in this study as in TCGA breast cancers, but with fewer amplifications and more mutations (Fig. 5d).

Discussion

This integrated study of 131 invasive urothelial bladder carcinomas provides numerous novel insights into disease biology and delineates multiple potential opportunities for therapeutic intervention. Treatment for muscle-invasive bladder cancer has not advanced beyond cisplatin-based combination chemotherapy and surgery in the past 30 years, and no new drugs for the disease have been approved in that time. Median survival for patients with recurrent or metastatic bladder cancer remains 14–15 months with cisplatin-based chemotherapy, and there is no widely recognized second-line therapy. With the exception of a single case report, there is also no known benefit from treatment with newer, targeted agents. Several of the genomic alterations identified in this study, particularly those involving the PI3K/akt/mTOR, CDKN2A/CDK4/CCND1 and RTK/RAS pathways, including FGFR3, are in principle to therapeutic targeting. Clinical trials based on patients with relevant druggable genomic alterations are warranted.

FGFR3 mutation is a common feature of low-grade non-invasive papillary urothelial bladder cancer, but it occurs at a much lower frequency in high-grade invasive bladder cancer. The cluster analysis in Fig. 3 highlights multiple mechanisms of FGFR3 activation, and its strong association with papillary morphology. The data presented here suggest a subset of muscle-invasive cancers that can potentially be targeted through FGFR3. Similarly, ERBB2 amplification may be targetable by strategies used in breast cancer, by small-molecule tyrosine kinase inhibitors or by novel immunotherapeutic approaches (NCT01353222). The data here provide further support for several on-going ERBB2-targeted trials in bladder cancer and further define the subpopulation of cancers suited to that approach. Finally, cluster III of the integrated expression profiling analysis reveals the existence of a urothelial carcinoma subtype with cancer stem-cell expression features (including KRT14 and KRT5), perhaps providing another avenue for therapeutic targeting.

The alterations identified in epigenetic pathways also suggest new possibilities for bladder cancer treatment. Ninety-nine (76%) of the tumours analysed here had an inactivating mutation in one or more of the chromatin regulatory genes, and 53 (41%) had at least two such mutations. Overall, the bladder cancers showed a mutational spectrum highly enriched with mutations in chromatin regulatory genes (Supplementary Table 2.10). Furthermore, integrated network analyses revealed a profound impact of those mutations on the activity levels of various transcription factors and pathways implicated in cancer. Drugs that target chromatin modifications—for example, recently developed agents that bind acetyl-lysine binding motifs (bromodomains)—might prove useful for treatment of the subset of bladder tumours that exhibit abnormalities in chromatin-modifying enzymes.

Our findings overall indicate bladder cancer as a prime candidate for exploration of that approach to therapy.

METHODS SUMMARY

Tumour and normal samples were obtained with institutional-review-board-approved consent and processed using a modified AllPrep kit (Qiagen) to obtain purified DNA and RNA. Quality-control analyses revealed only modest batch effects (Supplementary Information, section ‘Batch effects’). The tumours were profiled using Affymetrix SNP 6.0 microarrays for somatic CNAs, low-pass WGS (HiSeq) for somatic CNAs and translocations, RNA-seq (HiSeq) for mRNA and miRNA expression, Illumina Infinium (HumanMethylation450) arrays for DNA methylation, HiSeq for exome sequencing and RPPA for protein expression and phosphorylation. Statistical analysis and biological interpretation of the data were spearheaded by the TCGA genome data analysis centres. Sequence files are in CGHub (https://cghub.ucsc.edu/). All other molecular, clinical and pathological data are available through the TCGA Data Portal (https://tcga-data.nci.nih.gov/tcga/). The data can be explored through a compendium of next-generation clustered heat maps (http://bioinformatics.manderson.org/TCGA/NGCHMPortal/), the cbio Cancer Genomics Portal (http://cbioportal.org), TCGA (http://sylvioiwicki.soe.ucsc.edu/tiedie), SpliceSeq (http://bioinformatics.manderson.org/main/SpliceSeq-Overview), MiTmatch batch effects assessor (http://bioinformatics.manderson.org/tcga/MITmatch) and Regulome Explorer (http://explorer.cancerregulome.org/). Also see Supplementary Information.

Received 17 June; accepted 19 December 2013. Published online 29 January: corrected online 19 March 2014 (see full-text HTML version for details).
31. Vaske, C. J. et al. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. Bioinformatics 36, i237–i245 (2010).

32. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal carcinoma. Nature 499, 43–49 (2013).

33. Bos, R. et al. Activating HER2 mutations in HER2-negative breast cancers. Science 337, 224–227 (2012).

34. Greulich, H. et al. Functional analysis of receptor tyrosine kinase mutations in lung cancer identifies oncogenic extracellular domain mutations of ERBB3.Proc. Natl. Acad. Sci. USA 109, 14476–14481 (2012).

35. Jaiswal, B. S. Oncogenic ERBB3 mutations in human cancers. Cancer Cell 23, 639–643 (2013).

36. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology for Bladder Cancer. Vol. 21, 2012. http://www.nccn.orgprofessionals/physician_gls/f_guidelines.aspx (2012).

37. von Knecht, H. et al. Long-term survival results of a randomized trial comparing gemcitabine plus cisplatin, with methotrexate, vinblastine, doxorubicin, plus cisplatin in patients with bladder cancer. J. Clin. Oncol. 23, 4602–4608 (2005).

38. Iyer, G. et al. Genome sequencing identifies a basis for everolimus sensitivity. Science 338, 221 (2012).

39. Filipkowski, P. et al. Selective inhibition of BET bromodomains. Nature 468, 1067–1073 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to all of the patients and families who contributed to this study, as well as C. Gunter and L. Chastain for scientific editing and M. Sheth, J. Zhang and C. Ron Bouchard for administrative support. This work was supported by the following grants from the United States National Institutes of Health: US4 HG003273, US4 HG003275, US4 HG003276, US4 HG003277, US4 HG003278, US4 HG003279, U24 CA134385, U24 CA134384, U24 CA134383, U24 CA134384, U24 CA134385, U24 CA134386, U24 CA134387, U24 CA134388, U24 CA134405 and P01 CA120964. Additional personnel and funding sources are acknowledged in the Supplementary Information.

Author Information The primary and processed data used to generate the analyses presented here can be downloaded from the TCGA Data Portal (https://tcga-data.nci.nih.gov/tcga). All of the primary sequence files are deposited in the GenBank and all other data are deposited at the Data Coordinating Center (DCC) for public access (http://cancergenome.nih.gov/). We have deposited the datasets associated with the Supplementary tables and figures in the dbGaP public database. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.W.N. (weinste@mdanderson.org), S.P.L. (slerner@bcm.edu) or D.J.K. (dk@rics.bwh.harvard.edu). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.W.N. (weinste@mdanderson.org), S.P.L. (slerner@bcm.edu) or D.J.K. (dk@rics.bwh.harvard.edu). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.W.N. (weinste@mdanderson.org), S.P.L. (slerner@bcm.edu) or D.J.K. (dk@rics.bwh.harvard.edu).
