

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/MAPK 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 year¹. 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 *ARID1A* (ref. 4). Focused molecular analyses^{5,6} have delineated tumour subtypes and identified kinase-activating *FGFR3* gene fusions^{7,8}.

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-naïve, 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 $\geq 60\%$ of all nuclei; tumour necrosis was $\leq 20\%$ of the specimen; and variant histologies (squamous or small cell) were $\leq 50\%$ (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)⁹. 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 *PVRL4*, *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 *CCSER1*), *LRP1B* 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 $>30\times$). MuTect¹⁰ 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.

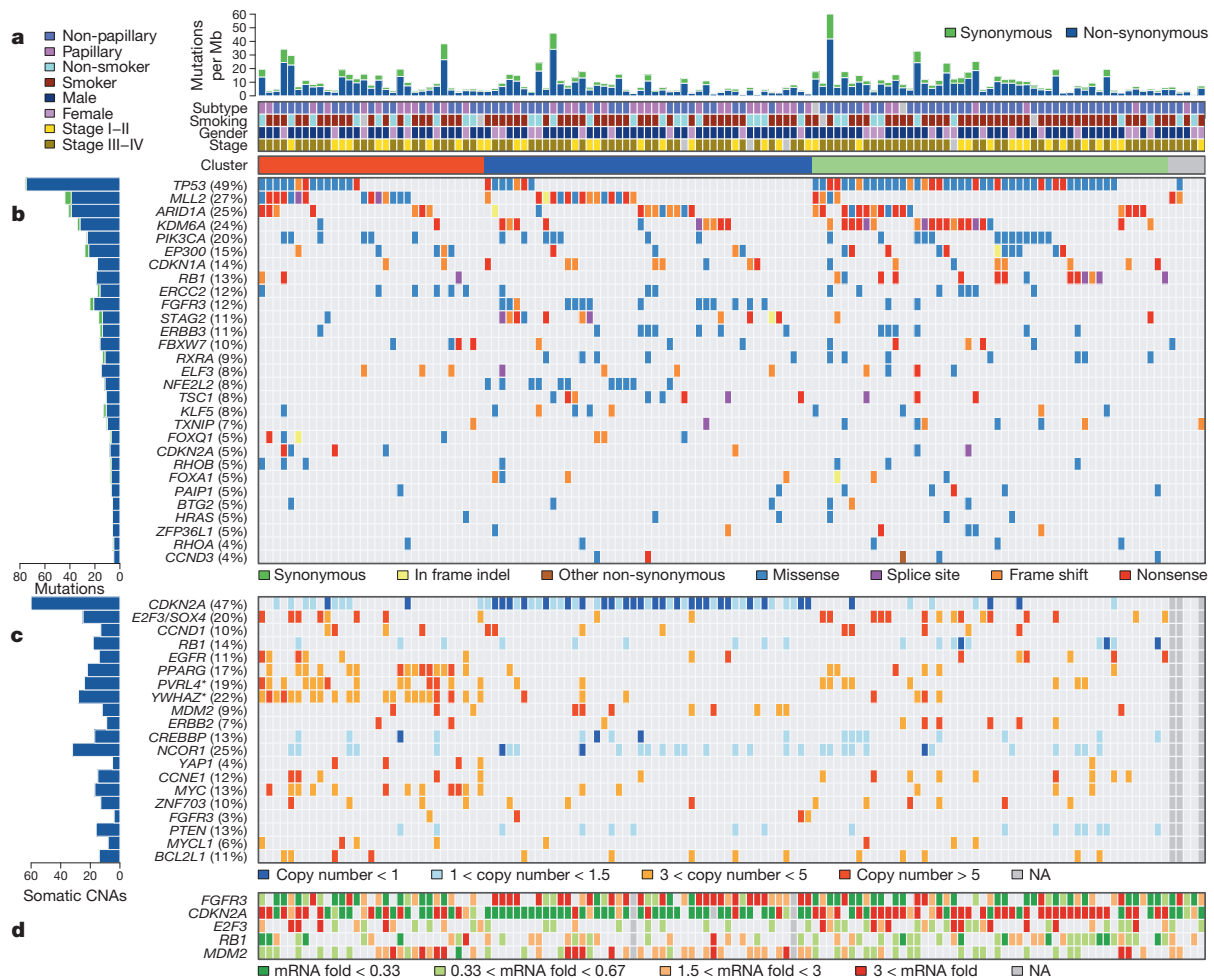


Figure 1 | The genomic landscape of bladder cancer. **a**, Mutation rate and type, histological subtype, smoking status, gender, tumour stage and cluster type. **b**, Genes with statistically significant levels of mutation (MutSig, false discovery rate < 0.1) and mutation types. **c**, Deletions and amplifications for genomic regions with statistically significant focal copy number changes (GISTIC2.0). 'Copy number' refers to absolute copy number. Note that two amplification peaks (*) contain several genes, any of which could be the target,

as opposed to the single gene listed here. **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.

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 database² 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 significantly mutated gene mutations, comparison with RNA-seq data for 123 samples and comparison with whole-genome sequence data for 18 samples. Overall, the validation rate was >99% in selected mutations by a combination of the methods (Supplementary Methods 2.4).

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%), *FOXQ1** (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 mutated at >3% frequency². *CDKN1A* (*p21*^{CIP1}), a cyclin-dependent kinase inhibitor¹², had predominantly null or truncating mutations, indicating loss of function. Fifteen of sixteen mutations in *ERCC2*, a nucleotide excision repair gene¹³, were deleterious missense mutations, suggesting dominant-negative effects. *ERCC2*-mutant tumours also had significantly fewer C>G mutations than did *ERCC2*-wild-type tumours (Supplementary Figs 2.3.1 and 2.3.2), and they trended towards higher overall mutation rate (Supplementary Fig. 2.12). Seven of twelve mutations in *RXRA* (retinoid X nuclear receptor alpha)¹⁴ occurred at the same amino acid (five S427F; two S427Y) in the ligand-binding domain. Those seven tumours showed increased expression of genes involved in adipogenesis and lipid metabolism (Supplementary Fig. 2.6 and Supplementary Data 2.6.1–2.6.3), suggesting that the mutations cause constitutive activation.

Eleven tumours (8%) had deleterious missense mutations in the Neh2 domain of *NFE2L2*, a transcription factor that regulates the anti-oxidant program in response to oxidative stress¹⁵. Those tumours

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 division¹⁷ (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 ‘focally 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 cancer¹⁸.

Fifty-one per cent of mutations overall were Tp>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). Three of the genes had previously been identified as mutated in urothelial cancers⁴, 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 $\geq 10\%$ but were not statistically significant by MutSig analysis: *MLL3*, *MLL*, *CREBBP*, *CHD7* and *SRAP*. 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 variations and pathogen 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 glioblastoma²¹ and bladder cancer^{7,8}, 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

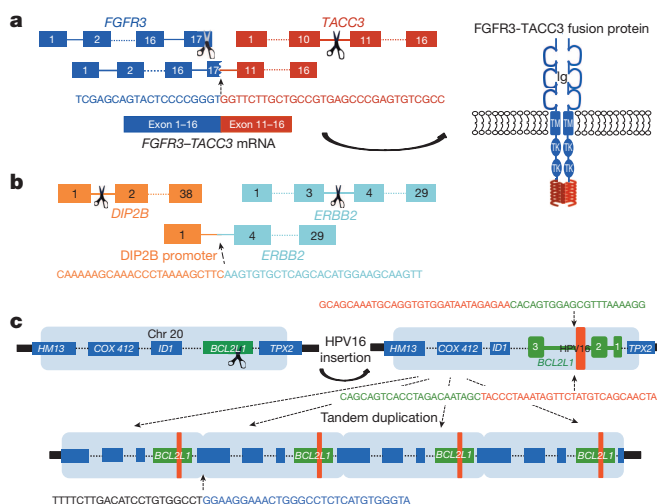


Figure 2 | Structural rearrangements and viral integration. **a**, *FGFR3*–*TACC3* fusion in sample TCGA-CF-A3MH showing the breakpoints in the two genes, the breakpoint junction sequences and the predicted fusion protein. **b**, Rearrangement involving *DIP2B* and *ERBB2* in TCGA-DK-A216. The *ERBB2* 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 ($\sim 10\times$ median; $>2\times$ 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)²². Cluster I samples also show lower expression of miR-145 and miR-125b, which have been reported as frequently downregulated in bladder cancer²³. 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 (RPPA) 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 those found in TCGA HER2-positive breast cancers²³.

For comparison, we asked whether any of the four clusters show gene signatures similar to those identified in any other tumour type(s) among the first 11 analysed by TCGA. We found that the signature of

bladder cancer cluster III ('basal/squamous-like') is similar to that of basal-like breast cancers, as well as squamous cell cancers of the head and neck and lung (Supplementary Fig. 4.2)^{24,25}. All four of those cancer types express characteristic epithelial lineage genes, including *KRT14*, *KRT5*, *KRT6A* and *EGFR*. Basal-like subtype²⁶ and squamous cell subtype²⁷ of urothelial carcinoma have been independently reported. Many of the samples in bladder cluster III express cytokeratins (that is, *KRT14* and *KRT5*) that were recently reported to mark stem/progenitor cells²⁶. Some of those samples also show a level of variant squamous histology (Fig. 3b). Bladder clusters I and II show features similar to those of luminal A breast cancer, with high mRNA and protein expression of luminal breast differentiation markers, including GATA3 and FOXA1 (Fig. 3c). Markers of urothelial differentiation such as the uroplakins (for example, UPK3A) are also highly expressed in clusters I and II, as are the epithelial marker E-cadherin and members of the miR-200 family of miRNAs (which target multiple regulators of epithelial–mesenchymal transition)²⁸ (Fig. 3c). Taken together, these observations indicate that, despite their diverse tissue origins, some bladder, breast, head and neck and lung cancers share common pathways of tumour development.

To determine whether the expression-based clusters could be seen in other data sets, we used the muscle-invasive bladder cancer samples from ref. 27, hierarchically clustering them with the genes used in our analysis. From the sample dendrogram, we identified four groups (Supplementary Fig. 4.3a). The four groups identified in the data set of ref. 27 correlated well with the four clusters identified in our TCGA data (Supplementary Fig. 4.3b).

When we analysed the RNA-seq data for transcript splice variation using SpliceSeq²⁹ (Supplementary Information, section 11), one finding of interest was an average of 3% *PKM1* and 97% *PKM2* transcripts in the tumour samples. The *PKM2* isoform of pyruvate kinase is the principal driver of a shift to aerobic glycolysis in tumours (the Warburg effect)³⁰. 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^{31,32}. 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 similar levels of pathway activity, independent of mutations in chromatin genes, suggesting that mutations in histone-modifying genes provide just one mechanism for disruption of the MYC/MAX hub. By contrast, tumours with chromatin-related mutations showed differential activity of transcription factors *FOXA2* and *SPI1*, 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.

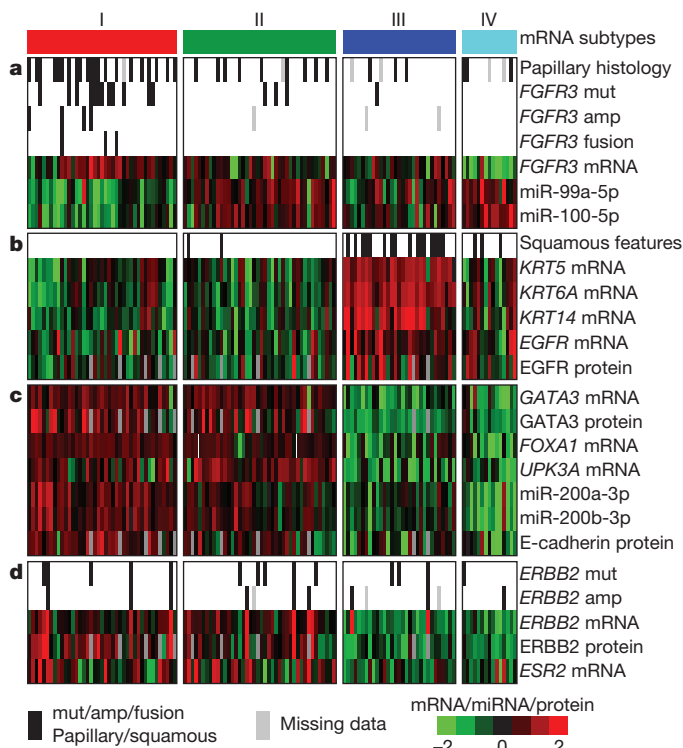


Figure 3 | Expression characteristics of bladder cancer. Integrated analysis of mRNA, miRNA and protein data led to identification of distinct subsets of urothelial carcinoma. Data for mRNA, miRNA and protein were z-normalized, and samples were organized in the horizontal direction by mRNA clustering. **a**, Papillary histology, *FGFR3* alterations, *FGFR3* expression and reduced *FGFR3*-related miRNA expression are enriched in cluster I. **b**, Expression of epithelial lineage genes and stem/progenitor cytokeratins are generally high in cluster III, some of which show variant squamous histology. **c**, Luminal breast and urothelial differentiation factors are enriched in clusters I and II. **d**, *ERBB2* mutation and oestrogen receptor beta (*ESR2*) expression are enriched in clusters I and II.

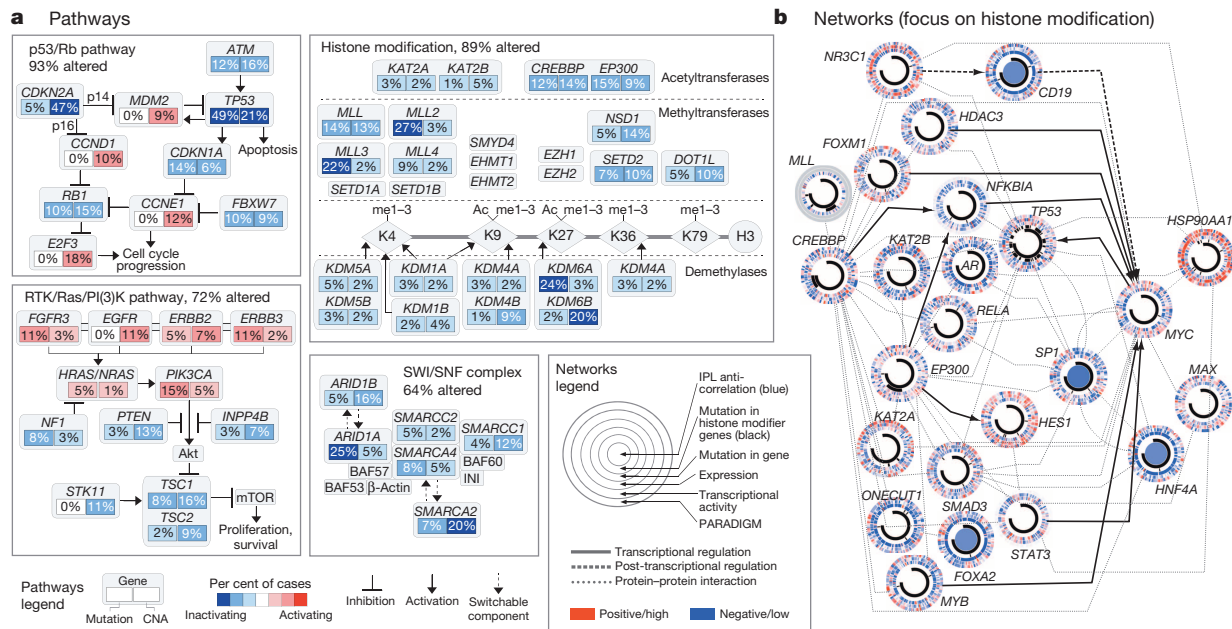


Figure 4 | Altered pathways and networks in bladder cancer. **a**, Somatic mutations and copy number alterations (CNA) in components of the p53/Rb pathway, RTK/RAS/PI(3)K pathway, histone modification system and SWI/SNF complex. Red, activating genetic alterations; blue, inactivating genetic alterations. Percentages shown denote activation or inactivation of at least one allele. **b**, 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.

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 point 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

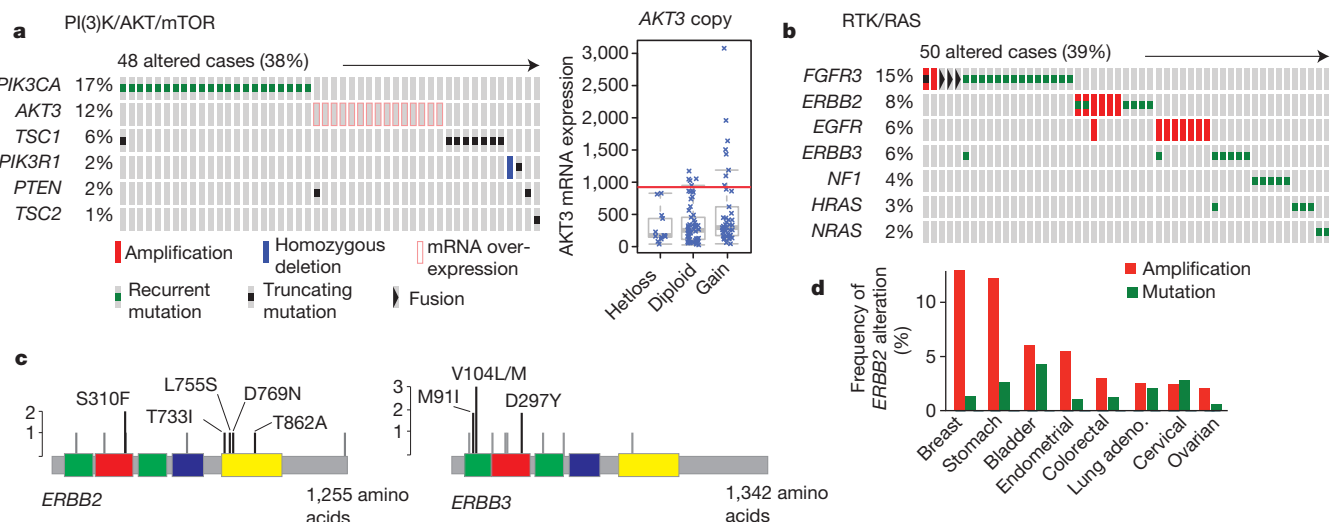


Figure 5 | Potential targets in bladder cancer. **a**, Alterations in the PI(3)K/AKT/mTOR pathway are mutually exclusive. Tumour samples are shown in columns; genes in rows. Only samples with at least one alteration are shown. AKT3 shows elevated expression in 10% of samples, independent of copy number (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 M916I. In-frame insertions were counted between amino acids 774 and 776. Only tumour types with an alteration frequency $\geq 2\%$ 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 previously⁴, but statistically significant mutation of *ERBB2* in bladder cancer has not been reported. Both genes are potential therapeutic targets in other diseases^{33–35}. Notably, *ERBB2* alterations were approximately as frequent in this study as in TCGA breast cancers, but with fewer amplifications and more mutations (Fig. 5d)²⁴.

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 years³⁶, 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 therapy³⁷. With the exception of a single case report, there is also no known benefit from treatment with newer, targeted agents³⁸. Several of the genomic alterations identified in this study, particularly those involving the PI(3)K/AKT/mTOR, CDKN2A/CDK4/CCND1 and RTK/RAS pathways, including *ERBB2* (Her-2), *ERBB3* and *FGFR3*, are amenable 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)³⁴. 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 (bromodomain)—might prove useful for treatment of the subset of bladder tumours that exhibit abnormalities in chromatin-modifying enzymes³⁹. 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.mdanderson.org/TCGA/NGCHMPortal/>), the cBio Cancer Genomics Portal (<http://cbioportal.org>), TieDIE (<http://sysbiowiki.soec.ucsc.edu/tiedie>), SpliceSeq (<http://bioinformatics.mdanderson.org/main/SpliceSeq:Overview>), MBatch batch effects assessor (<http://bioinformatics.mdanderson.org/tcgambatch/>) 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).

- Jemal, A. *et al.* Global cancer statistics. *CA Cancer J. Clin.* **61**, 69–90 (2011).
- Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
- Goebell, P. J. & Knowles, M. A. Bladder cancer or bladder cancers? Genetically distinct malignant conditions of the urothelium. *Urol. Oncol.* **28**, 409–428 (2010).
- Gui, Y. *et al.* Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nature Genet.* **43**, 875–878 (2011).
- Hurst, C. D., Platt, F. M., Taylor, C. F. & Knowles, M. A. Novel tumor subgroups of urothelial carcinoma of the bladder defined by integrated genomic analysis. *Clin. Cancer Res.* **18**, 5865–5877 (2012).
- Lindgren, D. *et al.* Integrated genomic and gene expression profiling identifies two major genomic circuits in urothelial carcinoma. *PLoS ONE* **7**, e38863 (2012).
- Williams, S. V., Hurst, C. D. & Knowles, M. A. Oncogenic *FGFR3* gene fusions in bladder cancer. *Hum. Mol. Genet.* **22**, 795–803 (2013).
- Wu, Y. M. *et al.* Identification of targetable *FGFR* gene fusions in diverse cancers. *Cancer Discov.* **3**, 636–647 (2013).
- Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
- Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnol.* **31**, 213–219 (2013).
- Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
- Warfel, N. A. & El-Deiry, W. S. p21^{WAF1} and tumorigenesis: 20 years after. *Curr. Opin. Oncol.* **25**, 52–58 (2013).
- Lehmann, A. R. The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. *Genes Dev.* **15**, 15–23 (2001).
- Tontonoz, P. *et al.* Adipocyte-specific transcription factor ARF6 is a heterodimeric complex of two nuclear hormone receptors, PPAR γ and RXR α . *Nucleic Acids Res.* **22**, 5628–5634 (1994).
- Shibata, T. *et al.* Cancer related mutations in NRF2 impair its recognition by Keap1-Cul3 E3 ligase and promote malignancy. *Proc. Natl Acad. Sci. USA* **105**, 13568–13573 (2008).
- Zhou, J., Yu, Q. & Chng, W. J. TXNIP (VDUP-1, TBP-2): a major redox regulator commonly suppressed in cancer by epigenetic mechanisms. *Int. J. Biochem. Cell Biol.* **43**, 1668–1673 (2011).
- Solomon, D. A. *et al.* Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science* **333**, 1039–1043 (2011).
- Samowitz, W. S. *et al.* Association of smoking, CpG island methylator phenotype, and V600E BRAF mutations in colon cancer. *J. Natl. Cancer Inst.* **98**, 1731–1738 (2006).
- Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
- Roberts, S. A. *et al.* Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol. Cell* **46**, 424–435 (2012).
- Singh, D. *et al.* Transforming fusions of *FGFR* and *TACC* genes in human glioblastoma. *Science* **337**, 1231–1235 (2012).
- Oneyama, C. *et al.* MicroRNA-mediated downregulation of mTOR/FGFR3 controls tumor growth induced by Src-related oncogenic pathways. *Oncogene* **30**, 3489–3501 (2011).
- Yoshino, H. *et al.* Aberrant expression of microRNAs in bladder cancer. *Nature Rev. Urol.* **10**, 396–404 (2013).
- Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
- Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
- Ho, P. L., Kurtova, A. & Chan, K. S. Normal and neoplastic urothelial stem cells: getting to the root of the problem. *Nature Rev. Urol.* **9**, 583–594 (2012).
- Sjodahl, G. *et al.* A molecular taxonomy for urothelial carcinoma. *Clin. Cancer Res.* **18**, 3377–3386 (2012).
- Korpai, M., Lee, E. S., Hu, G. & Kang, Y. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J. Biol. Chem.* **283**, 14910–14914 (2008).
- Ryan, M. C., Cleland, J., Kim, R., Wong, W. C. & Weinstein, J. N. SpliceSeq: a resource for analysis and visualization of RNA-seq data on alternative splicing and its functional impacts. *Bioinformatics* **28**, 2385–2387 (2012).
- Christofk, H. R. *et al.* The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **452**, 230–233 (2008).


31. Vaske, C. J. *et al.* Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* **26**, i237–i245 (2010).
32. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
33. Bose, R. *et al.* Activating HER2 mutations in HER2 gene amplification negative breast cancer. *Cancer Discov.* **3**, 224–237 (2013).
34. Greulich, H. *et al.* Functional analysis of receptor tyrosine kinase mutations in lung cancer identifies oncogenic extracellular domain mutations of ERBB2. *Proc. Natl Acad. Sci. USA* **109**, 14476–14481 (2012).
35. Jaiswal, B. S. *et al.* Oncogenic ERBB3 mutations in human cancers. *Cancer Cell* **23**, 603–617 (2013).
36. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology for Bladder Cancer. Vol. **1.2012**, http://www.nccn.org/professionals/physician_gls/f_guidelines.asp#site (2012).
37. von der Maase, 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. Filippakopoulos, 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: U54 HG003273, U54 HG003067, U54 HG003079, U24 CA143799, U24 CA143835, U24 CA143840, U24 CA143843, U24 CA143845, U24 CA143848, U24 CA143858, U24 CA143866, U24 CA143867, U24 CA143882, U24 CA143883, U24 CA144025 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 by registered users from The Cancer Genome Atlas at <https://tcga-data.nci.nih.gov/tcga/tcgaDownload.jsp>. All of the primary sequence files are deposited in CGHub and all other data are deposited at the Data Coordinating Center (DCC) for public access (<http://cancergenome.nih.gov/>, <https://cghub.ucsc.edu/> and https://tcga-data.nci.nih.gov/docs/publications/blca_2013/). Reprints and permissions information is available at www.nature.com/reprints. 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.N.W. (jweinste@mcdanderson.org), S.P.L. (slerner@bcm.edu) or D.J.K. (dk@rics.bwh.harvard.edu).

Author Contributions The Cancer Genome Atlas research network contributed collectively to this study. Biospecimens were provided by the tissue source sites and processed by the Biospecimen Core Resource. Data generation and analyses were performed by the genome-sequencing centres, cancer genome-characterization centres and genome data analysis centres. All data were released through the Data Coordinating Center. Project activities were coordinated by the NCI and NHGRI project teams. We also acknowledge the following TCGA investigators of the Bladder Analysis Working Group who contributed substantially to the project. Project leaders: J. N. Weinstein and S. P. Lerner. Data coordinator: C. J. Creighton. Analysis coordinators: R. Akbani and J. Kim. Manuscript coordinator: M. B. Morgan. Project coordinator: M. Sheth. Writing team: J. N. Weinstein, D. J. Kwiatkowski, S. P. Lerner, C. J. Creighton, P. W. Laird, R. Kucherlapati, R. Akbani, X. Su, K. A. Hoadley and M. C. Ryan. Clinical expertise: S. Lerner, D. J. Kwiatkowski, J. E. Rosenberg and D. Bajorin. Pathology review: H. Al-Ahmadie, B. A. Czerniak, D. Hansel, V. Reuter and B. Robinson. DNA sequence and copy number analysis: J. Kim, D. J. Kwiatkowski, A. D. Cherniack and J. E. Rosenberg. DNA methylation analysis: P. W. Laird and T. Hinoue. mRNA analysis: K. A. Hoadley, W. Y. Kim, J. S. Damrauer, W. Zhang, Y. Liu and R. Akbani. miRNA analysis: G. Robertson and A. J. Mungall. Transcript splicing analysis: M. Ryan and J. N. Weinstein. Protein analysis: R. Akbani and G. B. Mills. APOBEC: D. A. Gordenin. Pathway/integrated analysis: C. J. Creighton, N. Schultz, Evan O. Paull and J. Stuart. Chromosomal rearrangements and viral integration: X. Su, R. Kucherlapati, N. Santos, S. Lee and M. Parfenov. Batch effects: R. Akbani and J. N. Weinstein. Manuscript review: R. Gibbs, C. Gunter and M. Meyerson. Contact Pls: J. N. Weinstein, S. P. Lerner and D. J. Kwiatkowski.

 This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

The Cancer Genome Atlas Research Network

Analysis working group: The University of Texas MD Anderson Cancer Center John N. Weinstein^{1,2}, Rehan Akbani¹, Bradley M. Broom¹, Wenyi Wang¹, Roeland G. W. Verhaak¹, David McConkey³, **Baylor College of Medicine** Seth Lerner^{4,5}, Margaret Morgan^{5,6}, Chad J. Creighton⁷, Carolyn Smith⁸, **Broad Institute** David J.

Kwiatkowski^{9,10,11}, Andrew D. Cherniack⁹, Jaegil Kim⁹, Chandra Sekhar Pedamallu^{9,12}, Michael S. Noble⁹, **Memorial Sloan-Kettering Cancer Center** Hikmat A. Al-Ahmadie¹³, Victor E. Reuter¹³, Jonathan E. Rosenberg¹³, Dean F. Bajorin¹³, Bernard H. Bochner¹³, David B. Solit¹³, **Oregon Health and Science University, Department of Urology** Theresa Koppie¹⁴, **Weill Medical College of Cornell University** Brian Robinson¹⁵, **National Institute of Environmental Health Sciences** Dmitry A. Gordenin¹⁶, David Fargo¹⁶, Leszek J. Klimczak¹⁶, Steven A. Roberts¹⁶, **Optimum Therapeutics LLC** Jessie Au¹⁷, **University of Southern California Epigenome Center** Peter W. Laird¹⁸, Toshinori Hinoue¹⁸, **Computational Biology Center, Memorial Sloan-Kettering Cancer Center** Nikolaus Schultz¹⁹, Ricardo Ramirez¹⁹, **UCSD Department of Pathology** Donna Hansel²⁰, **Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill** Katherine A. Hoadley²¹, William Y. Kim^{21,22,23}, **Department of Genetics, University of North Carolina at Chapel Hill** Jeffrey S. Damrauer^{21,22}, **The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University** Stephen B. Baylin²⁴, **Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency** Andrew J. Mungall³⁰, A. Gordon Robertson³⁰, Andy Chu³⁰.

Genome Sequencing Center: Broad Institute David J. Kwiatkowski^{9,10,11}, Carrie Sougne⁹, Kristian Cibulskis⁹, Lee Lichtenstein⁹, Andrey Sivachenko⁹, Chip Stewart⁹, Michael S. Lawrence⁹, Gad Getz^{9,25}, Eric Lander⁹, Stacey B. Gabriel⁹.

Genome characterization centres: Dan L. Duncan Cancer Center, Human Genome Sequencing Center, Baylor College of Medicine Chad J. Creighton⁷, Lawrence Donehower^{7,26}, **Broad Institute** Andrew D. Cherniack⁹, Jaegil Kim⁹, Scott L. Carter⁹, Gordon Saksena⁹, Steven E. Schumacher^{9,27}, Carrie Sougne⁹, Samuel S. Freeman⁹, Joonil Jung⁹, Chandra Sekhar Pedamallu^{9,12}, Ami S. Bhatt^{9,12}, Trevor Pugh^{9,12}, Gad Getz^{9,25}, Rameen Beroukhi^{9,12,28}, Stacey B. Gabriel⁹, Matthew Meyerson^{9,12,29}, **Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency** Andrew J. Mungall³⁰, A. Gordon Robertson³⁰, Andy Chu³⁰, Adrian Ally³⁰, Miruna Balasundaram³⁰, Yaron S. N. Butterfield³⁰, Noreen Dhalla³⁰, Carrie Hirst³⁰, Robert A. Holt³⁰, Steven J. M. Jones³⁰, Darlene Lee³⁰, Haiyan L. Li³⁰, Marco A. Marra³⁰, Michael Mayo³⁰, Richard A. Moore³⁰, Jacqueline E. Schein³⁰, Payal Sipahimalani³⁰, Angela Tam³⁰, Nina Thiessen³⁰, Tina Wong³⁰, Natasja Wye³⁰, Reanne Bowlby³⁰, Eric Chuah³⁰, Ranabir Guin³⁰, Steven J. M. Jones³⁰, Marco A. Marra³⁰, **University of Southern California Epigenome Center** Toshinori Hinoue¹⁸, Hui Shen¹⁸, Moiz S. Bootwalla¹⁸, Timothy Triche Jr¹⁸, Phillip H. Lai¹⁸, David J. Van Den Berg¹⁸, Daniel J. Weisenberger¹⁸, Peter W. Laird¹⁸, **UCSD Department of Pathology** Donna Hansel²⁰, **Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill** Katherine A. Hoadley²¹, Saianand Balu²¹, Tom Bodenheimer²¹, Jeffrey S. Damrauer^{21,22} Alan P. Hoyle²¹, Stuart R. Jefferys²¹, Shaowu Meng²¹, Lisle E. Mose²¹, Janae V. Simons²¹, Mathew G. Soloway²¹, Junyuan Wu²¹, William Y. Kim^{21,22,23}, Joel S. Parker^{21,22}, D. Neil Hayes^{21,31}, **Research Computing Center, University of North Carolina at Chapel Hill** Jeffrey Roach³², **Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill** Elizabeth Buda³³, **Department of Biology, University of North Carolina at Chapel Hill** Corbin D. Jones^{33,34}, Piotr A. Mieczkowski³⁴, Donghui Tan³⁴, Umadevi Veluvolu³⁴, Scot Waring³⁴, **Eshelman School of Pharmacy, University of North Carolina at Chapel Hill** J. Todd Auman³⁵, **Department of Genetics, University of North Carolina at Chapel Hill** Charles M. Perou²², Matthew D. Wilkerson²², **Department of Genetics, Harvard Medical School** Netty Santoso³⁶, Michael Parfenov³⁶, Xiaojia Ren³⁶, Angeliki Pantazi³⁶, Angela Hadjipanyis^{36,37}, Jonathan Seidman³⁶, Raju Kucherlapati^{36,37}, **The Center for Biomedical Informatics, Harvard Medical School** Semin Lee³⁸, Lixing Yang³⁸, Peter J. Park^{37,38,39}, **Cancer Biology Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University** Stephen B. Baylin²⁴, **Division of Genetics, Brigham and Women's Hospital** Andrew Wei Xu³⁷, **Institute for Applied Cancer Science, Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center** Alexei Protopopov⁴⁰, Jianhua Zhang⁴⁰, Christopher Bristow⁴⁰, Harshad S. Mahadeshwar⁴⁰, Sahil Seth⁴⁰, Xingzhi Song⁴⁰, Jiabin Tang⁴⁰, Dong Zeng⁴⁰, Lynda Chin^{9,40}, **The University of Texas MD Anderson Cancer Center, Department of Pathology** Charles Guo⁴¹.

Genome data analysis centres: The University of Texas M.D. Anderson Cancer Center John N. Weinstein^{1,2}, Rehan Akbani¹, Bradley M. Broom¹, David McConkey³, Tod D. Casasent¹⁰, Wenbin Liu^{1,2}, Zhenlin Ju^{1,2}, Thomas Motter¹, Bo Peng¹, Michael Ryan¹, Wenyi Wang¹, Roeland G. W. Verhaak¹, Xiaoping Su¹, Ji-Yeon Yang^{1,2}, Philip L. Lorenzi¹, Hui Yao¹, Nianxiang Zhang¹, Jiexin Zhang¹, Gordon B. Mills², **Broad Institute** Jaegil Kim⁹, Michael S. Noble⁹, Juok Cho⁹, Daniel DiCara⁹, Scott Frazer⁹, Nils Gehlenborg⁹, David I. Heiman⁹, Pei Lin⁹, Yingchun Liu⁹, Petar Stojanov^{9,12}, Doug Voet⁹, Hailei Zhang⁹, Lihua Zou⁹, Lynda Chin^{9,40}, Gad Getz^{9,25}, **Institute for Systems Biology** Brady Bernard⁴², Dick Kreisberg⁴², Sheila Reynolds⁴², Hector Rovira⁴², Ilya Shmulevich⁴², **Computational Biology Center, Memorial Sloan-Kettering Cancer Center** Ricardo Ramirez¹⁹, Nikolaus Schultz¹⁹, Jianjiong Gao¹⁹, Anders Jacobsen¹⁹, B. Arman Aksoy¹⁹, Yevgeniy Antipin¹⁹, Giovanni Ciriello¹⁹, Gideon Dresdner¹⁹, Benjamin Gross¹⁹, William Lee¹⁹, Boris Reva¹⁹, Ronglai Shen¹⁹, Rileen Sinha¹⁹, S. Onur Sumer¹⁹, Nils Weinhold¹⁹, Marc Ladanyi¹⁹, Chris Sander¹⁹, **Buck Institute for Research on Aging** Christopher Benz⁴³, **University of California Santa Cruz** Daniel Carlin⁴⁴, David Haussler⁴⁴, Sam Ng⁴⁴, Evan O. Paull⁴⁴, Joshua Stuart⁴⁴, Jing Zhu⁴⁴, **Department of Pathology, MD Anderson Cancer Center** Yuxin Liu⁴⁵, Wei Zhang⁴⁵, **Helen Diller Family Comprehensive Cancer Center, University of California** Barry S. Taylor⁴⁶.

Biospecimen core resource: The Research Institute at Nationwide Children's Hospital Tara M. Lichtenberg⁴⁷, Erik Zmuda⁴⁷, Thomas Barr⁴⁷, Aaron D. Black⁴⁷, Myra George⁴⁷, Benjamin Hanf⁴⁷, Carmen Helsel⁴⁷, Cynthia McAllister⁴⁷, Nilsa C. Ramirez^{47,48}, Teresa R. Tabler⁴⁷, Stephanie Weaver⁴⁷, Lisa Wise⁴⁷, Jay Bowen⁴⁷, Julie M. Gastier-Foster^{47,48}

Tissue source sites: The University of Texas MD Anderson Cancer Center John N. Weinstein^{1,2}, **Scott Department of Urology, Baylor College of Medicine** Seth Lerner^{4,5}, Weiguo Jian^{4,5}, Sebrina Tello^{4,5}, **Texas Cancer Research Biobank (TCRB), Baylor College of Medicine** Michael Ittman^{5,49}, Patricia Castro^{5,49}, Whitney D. McClenden⁵, Margaret Morgan^{5,6}, Richard Gibbs^{5,6}, **Broad Institute** Yingchun Liu⁹, **Analytical Biological Services, Inc.** Charles Saller⁵⁰, Katherine Tarvin⁵⁰, **Cleveland Clinic Foundation** Jennifer M. DiPiero⁵¹, Jennifer Owens⁵¹, **Georgia Regents University Cancer Center** Roni Bollag⁵², Qiang Li⁵², Paul Weinberger⁵², **Helen F. Graham Cancer Center at Christiana Care** Christine Czerwinski⁵³, Lori Huelsenbeck-Dill⁵³, Mary Iacocca⁵³, Nicholas Petrelli⁵³, Brenda Rabeno⁵³, Pat Swanson⁵³, **International Genomics Consortium** Troy Shelton⁵⁴, Erin Curley⁵⁴, Johanna Gardner⁵⁴, David Mallery⁵⁴, Robert Penny⁵⁴, **ILSbio, LLC** Nguyen Van Bang^{55,56}, Phan Thi Hanh^{55,56}, Bernard Kohl⁵⁵, Xuan Van Le⁵⁵, Bui Duc Phu^{55,56}, Richard Thorp⁵⁵, Nguyen Viet Tien^{55,56}, Le Quang Vinh^{55,56}, **IU School of Medicine** George Sandusky⁵⁷, **Lahey Hospital and Medical Center** Eric Burks⁵⁸, Kimberly Christ⁵⁸, Jason Gee⁵⁸, Antonia Holway⁵⁸, Alireza Moynadeh⁵⁸, Andrea Sorcini⁵⁸, Travis Sullivan⁵⁸, **Memorial Sloan-Kettering Cancer Center** Hikmat A. Al-Ahmadie¹³, Dean F. Bajorin¹³, Bernard H. Bochner¹³, Ilana R. Garcia-Grossman¹³, Ashley M. Regazzi¹³, David B. Solit¹³, Jonathan E. Rosenberg¹³, Victor E. Reuter¹³, **Oregon Health and Science University, Department of Urology** Theresa Koppie¹⁴, **University of North Carolina, Lineberger Cancer Center** Lori Boice⁵⁹, Wendy Kimryn Rathmell⁵⁹, Leigh Thorne⁵⁹, **University of Pittsburgh** Sheldon Bastacky⁶⁰, Benjamin Davies⁶⁰, Rajiv Dhir⁶⁰, Jeffrey Gingrich⁶⁰, Ronald Hrebinko⁶⁰, Jodi Maranchie⁶⁰, Joel Nelson⁶⁰, Anil Parwani⁶⁰, **Roswell Park Cancer Institute** Wiam Bshara⁶¹, Carmelo Gaudioso⁶¹, Carl Morrison⁶¹, **Ontario Tumour Bank—Hamilton site, St Joseph's Healthcare Hamilton** Vina Alexopoulou⁶², John Bartlett⁶², Jay Engel⁶², Sugy Kodeeswaran⁶², **The University of Chicago** Tatjana Antic⁶³, Peter H. O'Donnell⁶³, Norm D. Smith⁶³, Gary D. Steinberg⁶³, **University of Miami, Sylvester Comprehensive Cancer Center** Sophie Egea⁶⁴, Carmen Gomez-Fernandez⁶⁴, Lynn Herbert⁶⁴, Merce Jorda⁶⁴, Mark Soloway⁶⁴, **UT Southwestern Medical Center** Allison Beaver⁶⁵, Suzie Carter⁶⁵, Payal Kapur⁶⁵, Cheryl Lewis⁶⁵, Yair Lotan⁶⁵, **Weill Medical College of Cornell University** Brian Robinson¹⁵, **UCSD Department of Pathology** Donna Hansel²⁰, **The University of Texas MD Anderson Cancer Center, Department of Pathology** Charles Guo⁴¹, Jolanta Bondaruk⁴¹, Bogdan Czerniak⁴¹

Disease working group: The University of Texas MD Anderson Cancer Center Rehan Akbani¹, Bradley M. Broom¹, Yuxin Liu⁴⁵, Wei Zhang⁴⁵, John N. Weinstein^{1,2}, **Scott Department of Urology, Baylor College of Medicine** Seth Lerner^{4,5}, **Baylor College of Medicine** Margaret Morgan^{5,6}, **Broad Institute** Jaegil Kim⁹, Andrew D. Cherniack⁹, Samuel S. Freeman⁹, Chandra Sekhar Pedamallu^{9,12}, Michael S. Noble⁹, David J. Kwiatkowski^{9,10,11}, **Memorial Sloan-Kettering Cancer Center** Hikmat A. Al-Ahmadie¹³, Dean F. Bajorin¹³, Bernard H. Bochner¹³, David B. Solit¹³, Jonathan E. Rosenberg¹³, Victor E. Reuter¹³, **Oregon Health and Science University, Department of Urology** Theresa Koppie¹⁴, **Weill Medical College of Cornell University** Brian Robinson¹⁵, **Stanford University, Department of Urology** Eila Skinner⁶⁶, **Computational Biology Center, Memorial Sloan-Kettering Cancer Center** Ricardo Ramirez¹⁹, Nikolaus Schultze¹⁹, **UCSD Department of Pathology** Donna Hansel²⁰, **Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill** William Y. Kim^{21,22,23}, **The University of Texas MD Anderson Cancer Center, Department of Pathology** Charles Guo⁴¹, Jolanta Bondaruk⁴¹, Kenneth Aldape⁴¹, Bogdan Czerniak⁴¹

Data coordination centre: SRA International Mark A. Jensen⁶⁷, Ari B. Kahn⁶⁷, Todd D. Pihl⁶⁷, David A. Pot⁶⁷, Deepak Srinivasan⁶⁷, Yunhu Wan⁶⁷

Project team: MLF Consulting Martin L. Ferguson⁶⁸, **National Cancer Institute** Jean Claude Zenklusen⁶⁹, Tanja Davidsen⁶⁹, John A. Demchok⁶⁹, Kenna R. Mills Shaw^{3,69}, Margi Sheth⁶⁹, Roy Tarnuzzer⁶⁹, Zhining Wang⁶⁹, Liming Yang⁶⁹, **National Human Genome Research Institute** Carolyn Hutter⁷⁰, Bradley A. Ozenberger⁷⁰, Heidi J. Sofia⁷⁰, **Scintents, LLC** Greg Eley⁷¹

¹Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ²Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ³The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ⁴Scott Department of Urology, Baylor College of Medicine, Houston, Texas 77030, USA. ⁵Texas

Cancer Research Biobank (TCRB), Baylor College of Medicine, Houston, Texas 77030, USA. ⁶Human Genome Sequencing Center at Baylor College of Medicine, Houston, Texas 77030, USA. ⁷Dan L. Duncan Cancer Center, Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA. ⁸Baylor College of Medicine, Houston, Texas 77030, USA. ⁹The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University Cambridge, Massachusetts 02142, USA. ¹⁰Brigham and Women's Hospital, 75 Francis St, Boston, Massachusetts 02115, USA. ¹¹Harvard Medical School, Boston, Massachusetts 02115, USA. ¹²Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. ¹³Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ¹⁴Oregon Health and Science University, Department of Urology, 3303 SW Bond Avenue, CHH10U, Portland, Oregon 97239, USA. ¹⁵Weill Medical College of Cornell University, New York, New York 10065, USA. ¹⁶National Institute of Environmental Health Sciences, 111 T.W. Alexander Drive, Research Triangle Park, North Carolina 27709, USA. ¹⁷Optimum Therapeutics LLC, 9363 Towne Centre Drive, San Diego, California 92121, USA. ¹⁸University of Southern California Epigenome Center, University of Southern California, Los Angeles, California 90033, USA. ¹⁹Computational Biology Center, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. ²⁰UCSD Department of Pathology 9500 Gilman Drive, La Jolla, California 92093, USA. ²¹Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ²²Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ²³Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ²⁴Cancer Biology Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, Maryland 21231, USA. ²⁵Massachusetts General Hospital, Cancer Center and Department of Pathology, 55 Fruit Street, Boston, Massachusetts 02114, USA. ²⁶Department of Molecular Virology and Microbiology, Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas 77030, USA. ²⁷Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. ²⁸Department of Medicine, Harvard Medical School, Boston, Massachusetts 02215, USA. ²⁹Department of Pathology, Harvard Medical School, Boston, Massachusetts 02215, USA. ³⁰Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia V5Z 4S6, Canada. ³¹Department of Internal Medicine, Division of Medical Oncology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ³²Research Computing Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ³³Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ³⁴Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ³⁵Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ³⁶Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ³⁷Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ³⁸The Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts 02115, USA. ³⁹Informatics Program, Children's Hospital, Boston, Massachusetts 02115, USA. ⁴⁰Institute for Applied Cancer Science, Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ⁴¹The University of Texas MD Anderson Cancer Center, Department of Pathology, Unit 085; 1515 Holcombe Boulevard, Houston, Texas 77030, USA. ⁴²Institute for Systems Biology, 401 Terry Ave N, Seattle, Washington 98109, USA. ⁴³Buck Institute for Research on Aging; 8001 Redwood Blvd, Novato, California 94945, USA. ⁴⁴University California Santa Cruz, 1156 High Street, Santa Cruz, California 95064, USA. ⁴⁵Department of Pathology, MD Anderson Cancer Center, Houston, Texas 77030, USA. ⁴⁶Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, California 94158, USA. ⁴⁷The Research Institute at Nationwide Children's Hospital, Columbus, Ohio 43205, USA. ⁴⁸The Ohio State University, Columbus, Ohio 43210, USA. ⁴⁹Department of Pathology, Baylor College of Medicine, Houston, Texas 77030, USA. ⁵⁰Analytical Biological Services, Inc., 701 Cornell Drive, Wilmington, Delaware 19801, USA. ⁵¹Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA. ⁵²Georgia Regents University Cancer Center, Augusta, Georgia 30912, USA. ⁵³Helen F. Graham Cancer Center at Christiana Care, 4701 Ogletown Stanton Road, Newark, Delaware 19713, USA. ⁵⁴International Genomics Consortium, 445 N. Fifth Street, Phoenix, Arizona 85004, USA. ⁵⁵ILSbio, LLC 100 Radcliffe Drive, Chestertown, Maryland 21620, USA. ⁵⁶Hue Central Hospital, Hue City, Vietnam. ⁵⁷IU School of Medicine, Med Science Bldg, Room 128A, 635 Barnhill Drive, Indianapolis, Indiana 46202, USA. ⁵⁸Lahey Hospital and Medical Center, Burlington, Massachusetts 01805, USA. ⁵⁹University of North Carolina, Lineberger Cancer Center, 450 West Drive, Chapel Hill, North Carolina 27599, USA. ⁶⁰University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. ⁶¹Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, New York 14063, USA. ⁶²Ontario Tumour Bank—Hamilton site, St Joseph's Healthcare Hamilton, Hamilton, Ontario L8N 3Z5, Canada. ⁶³The University of Chicago, Chicago, Illinois 60637, USA. ⁶⁴University of Miami, Sylvester Comprehensive Cancer Center, 1550 NW 10th Avenue, Miami, Florida 33136, USA. ⁶⁵UT Southwestern Medical Center 5323 Harry Hines Blvd, Dallas, Texas 75390-9110, USA. ⁶⁶Stanford University, Department of Urology, 300 Pasteur Drive, Suite S287, Stanford, California 94305, USA. ⁶⁷SRA International, Fairfax, Virginia 22033, USA. ⁶⁸MLF Consulting, Arlington, Massachusetts 02474, USA. ⁶⁹National Cancer Institute, 31 Center Drive, 3A20, Bethesda, Maryland 20892, USA. ⁷⁰National Human Genome Research Institute, 5635 Fishers Lane, Rockville, Maryland 20852, USA. ⁷¹Scintents, LLC, Atlanta, Georgia 30666, USA.