HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2017 March 30.

Published in final edited form as:

Nat Genet. ; 43(9): 875–878. doi:10.1038/ng.907.

Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder

A full list of authors and affiliations appears at the end of the article.

Abstract

Transitional cell carcinoma (TCC) is the most common type of bladder cancer. Here we sequenced the exomes of nine individuals with TCC and screened all the somatically mutated genes in a prevalence set of 88 additional individuals with TCC with different tumor stages and grades. In our study, we discovered a variety of genes previously unknown to be mutated in TCC. Notably, we identified genetic aberrations of the chromatin remodeling genes (*UTX, MLL-MLL3*, *CREBBP-EP300*, *NCOR1*, *ARID1A* and *CHD6*) in 59% of our 97 subjects with TCC. Of these genes, we showed *UTX* to be altered substantially more frequently in tumors of low stages and grades, highlighting its potential role in the classification and diagnosis of bladder cancer. Our results provide an overview of the genetic basis of TCC and suggest that aberration of chromatin regulation might be a hallmark of bladder cancer.

Bladder cancer is the ninth most common cancer worldwide[1], with TCC being the predominant form (representing 90% of cases). Clinically, there are two distinct groups of TCCs: ~70% of the affected individuals have superficial non–muscle-invasive TCCs (NMI-TCCs, stage Ta or T1), which tend to recur but which are generally not life-threatening, and ~30% of individuals have muscle-invasive TCCs (MI-TCCs, stages T2–T4), which are associated with a high risk of death from distant metastases[2]. Previous studies based on candidate gene approaches suggested that the two subgroups of TCCs may have divergent genetic backgrounds[2]: NMI-TCCs often harbor mutations in *FGFR3* and the Ras gene family, and MI-TCCs usually have defects in *TP53* and *RB1* in high-grade tumors. These discoveries provided important insights into potential diagnoses and therapeutic applications, yet no comprehensive analysis of this cancer has been performed. In this study,

Correspondence should be addressed to Z.C. (caizhiming2000@yahoo.com.cn), J.W. (wangj@genomics.org.cn) or H.Y. (yanghm@genomics.org.cn).

[12]These authors contributed equally to this work.

Accession codes. All sequencing data from this study are deposited in NCBI Sequence Read Archive under the accession number SRA038181.

Note: Supplementary information is available on the Nature Genetics website.

AUTHOR CONTRIBUTIONS

Jun Wang, Z.C., Jian Wang, H.Y., S.L. and Y.G. managed the project. A.T., X. Li, L.Z., Z. Li, F.Z., X. Zhao, C. Liang, C. Liu, Y.W., L.S., Z.J., Jing Chen, S. Wu, Z.Z., R. Yang, J. Zhao, C.X., Z.G., J.Y., H. Zhang and W.Y. prepared the samples. X.H., R.W., P.H., H.J., J.L. and X. Zhang performed the sequencing. Y.G., G.G., Y.H., S.G., C.C., M.H., W.J., R. Ye, Z. Liu, S. Wan, H. Zheng, K.K., M.L.N. and Y.L. performed the bioinformatic analysis. Y.H., X.H., Jinnong Chen, S.Y., X. Liu, D.F. and J. Zou performed the validation of somatic mutations. G.G. and Y.H. wrote the paper. Y.G., Jun Wang, Z.C., X.H., Y.L., D.T. and X.S. revised the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

we aimed to screen TCC systematically to identify other previously unidentified bladder-cancer–associated genes.

We performed whole-exome sequencing of the genomic DNA from tissue samples from nine individuals with MI-TCC (stage ≥ T2; Supplementary Table 1) and their matched peripheral blood samples. In this stage, called here the 'discovery screen', we primarily focused on MI-TCC because of its poor prognosis and survival. We generated sequencing reads with an Illumina GAII platform (Illumina, Inc.) and aligned them to the reference human genome (hg18) using MAQ[3] and BWA[4] software (Online Methods). On average, we sequenced the 34-Mb targeted exome regions of each sample to a mean depth of 100× or greater (Supplementary Fig. 1a and Supplementary Table 2). Over 95% of the target regions were covered sufficiently for confident variant calling (defined as ≥ 10×; Supplementary Fig. 1b).

To identify somatic mutations, we compared the sequencing data generated from the tumor-normal sample pairs of each subject (Online Methods). To eliminate any previously described germline variants, we cross-referenced potential somatic mutations against the dbSNP130 and SNP datasets of Han Chinese in Beijing (CHB) and Japanese in Toyko (JPT) from the three pilot studies in the 1000 Genomes Project. Based on these criteria, we identified 465 predicted somatic mutations, including 105 synonymous mutations, 284 missense mutations, 45 nonsense mutations, 2 splice-site changes and 29 small coding insertions or deletions (indels) (Supplementary Table 3). Of the 221 predicted somatic substitutions and 12 indels that we randomly selected for validation, we confirmed 200 (90.5%) and 8 (66.7%) by genotyping and Sanger sequencing, respectively (Online Methods). The mutation spectrum in the nine TCC samples was dominated by C:G>T:A transitions (Supplementary Fig. 2), as has been reported in several other human cancers[5].

To further evaluate the mutation prevalence of 328 genes that had at least one non-silent somatic mutation in the discovery screen, we determined the coding sequences of these genes in a prevalence screen of 88 additional individuals with TCC (Supplementary Table 1). We included both individuals with MI-TCC (51 individuals) and NMI-TCC (37 individuals) to investigate whether there are genes that may be preferentially mutated in any subgroups of affected individuals. In addition, in this stage, we also screened two other well-known bladder cancer genes (*FGFR3* and *RB1*)[2] without somatic mutation in the discovery stage. In brief, we enriched the targeted exonic sequences of matched normal and tumor DNAs from the 88 untreated subjects followed by Illumina-based resequencing (Online Methods). Overall, we achieved a mean coverage depth of ~80× for all the samples sequenced, with at least ~85% of the targeted bases being sufficiently covered (≥10×; Supplementary Table 4). We determined somatic mutations using the same method as used in the discovery screen, and we selected all the non-silent mutations for verification by genotyping or Sanger sequencing (Supplementary Table 5).

We combined mutations obtained from the discovery and validation screens to prioritize the significantly mutated genes that are likely to be implicated in TCC tumorigenesis. We considered a gene to be a potential significantly mutated gene if the gene harbor confirmed non-silent mutations in at least two tumors and if its non-silent mutation rate was significantly higher ($P < 0.05$) than the background (Supplementary Table 6 and Online

Methods). In total, we identified 54 significantly mutated genes in our dataset (Supplementary Table 7). Five of the significantly mutated genes have previously well-established roles in TCC (Table 1), including *TP53* (ref. 6) (altered in 21% of TCCs), *RB1* (ref. 7) (altered in 11%), *HRAS* (ref. 8) (altered in 10%), *FGFR3* (ref. 9) (altered in 9%) and *KRAS* (ref. 8) (altered in 6% of TCCs). The other 49 significantly mutated genes were all new candidates with no previously well-defined roles in TCCs, and 33% (16 out of 49) of these genes were frequently mutated genes that showed mutations in more than 5% of TCCs (Table 1).

Most notably, we detected frequent non-silent mutations in eight genes that are involved in the chromatin remodeling process (Table 1). Of these genes, the most frequently altered were the histone demethylase gene (HDMT) *UTX* (mutated in 21% of TCCs), two histone acetyltransferase (HAT) genes, *CREBBP* and *EP300*, and the SWI/SNF-related chromatin remodeling gene *ARID1A* (with all these genes being mutated in 13% of TCCs). We also detected aberration of each of the remaining four genes in more than 5% of TCCs, including the histone methyltransferase (HMT) genes *MLL* and *MLL3* and two other chromatin remodeling genes, *NCOR1* (encoding a constitutive subunit for the N-coR-HDAC3 complex that possesses histone deacetylation (HDAT) activity[10]) and *CHD6* (encoding a component of SNF2/RAD54 helicase family that remodels chromatin to allow cell-type–specific gene expression[11]).

Genetic alterations of the genes involved in the chromatin remodeling process were also reported in various other tumors in recent studies. Inactivating mutations in *UTX* were observed in multiple types of cancers[12]; frequent mutations of *ARID1A* were identified in ovarian clear cell carcinoma[13]; and truncating mutations in the HMT gene *SETD2*, the HDMT gene *JARID1C* and another member of the SNF/SWI complex, *PBRM1*, were identified in renal cell carcinoma[14,15]. Aberrations of the chromatin remodeling genes may directly lead to the misregulation of multiple downstream effector genes, consequently promoting the tumorigenesis process[16]. Nevertheless, to our knowledge, except for *UTX*, which showed alterations in 2 of 14 TCC cell lines[12], genetic mutations in the genes involved in the chromatin remodeling process have not yet been reported in the primary tumors of TCC. In our study, we that found 57 (59%) subjects harbored non-silent mutations in chromatin remodeling genes, indicating that disruption of the chromatin remodeling machinery may be one of the main mechanisms that leads to TCC.

In addition to the frequency of mutations in a gene, the nature of the mutations can provide valuable information for classifying the mutated genes as oncogenes (dominant genes) or as tumor suppressors (recessive genes)[17]. Our observations clearly distinguished *UTX* as a tumor suppressor gene in TCC, as evidenced by its mutation pattern of significant enrichment (16 of 22 mutations; $P < 0.001$) with truncating mutations. These included 11 nonsense mutations, 4 frameshift indels and 1 splice site change (Fig. 1a). All but one of these were predicted to truncate the JmjC domain, which is essential for the demethylase activity of the protein product. We observed the same mutation pattern in *ARID1A* (Fig. 1b), in which truncating mutations accounted for 14 (78%) of 18 mutations ($P < 0.001$) and which generated defective proteins lacking the ARID DNA-binding domain or LXXLL motif. The patterns of mutations in *CREBBP* and *EP300* also suggested their potential tumor

suppressing roles in TCC. Recent studies showed that they behave as tumor-suppressor genes in various tumors[18,19]. In our screen, 73% and 54% of the mutations in these two genes are predicted to disrupt the key HAT functional region in *CREBBP* and to truncate the CH3 domain that mediates the interactions between p300 and other cellular proteins (for example, p53 and E2F1)[20] in *EP300*, respectively (Fig. 1c). Of note, we also identified recurrent truncating mutations in other frequently mutated chromatin remodeling genes (*MLL*, *MLL3* and *NCOR1*). In previous studies, *MLL* was shown to be a dominant cancer gene affected by recurrent translocations in leukemias[21]. However, our data, as well as the mutation data recently deposited in COSMIC database (see URLs), supported that *MLL* and its homolog *MLL3* are likely to act recessively in some solid tumors[22,23].

The correlations among mutations in the frequently altered genes could provide insights into their functional interactions. We also confirmed in our study the well-known examples of the negative correlations of mutations in *TP53* and *FGFR3* (ref. 24) and of mutations in *TP53* and *HRAS* (ref. 2) ($P < 0.05$; Supplementary Table 8), with no sample here having mutations in both *TP53* and *FGFR3* or *HRAS* (Fig. 2). We also found significantly negative correlations between mutations in *UTX* and the Ras genes (*KRAS* and *HRAS*). In addition, we observed positive correlations of mutations between the following pairs of genes: *ANK2* and *CHD6*, *LRP2* and *ZFHX3*, and *CHD6* and *LRP2* (all $P < 0.05$). Notably, we did detect concurrent mutations in multiple genes within the chromatin remodeling process (Fig. 2), although these correlations were not statistically significant. The eight frequently mutated chromatin remodeling genes can be classified into five subgroups according to their functional specificities: HMT (*MLL3* and *MLL*), HDMT (*UTX*), HAT (*CREBBP* and *EP300*), HDAT (*NCOR1*) and members of the nucleosomal remodeling complex (*ARID1A* and *CHD6*). We found that 16% of TCCs had mutations in at least two of the five subgroups (Fig. 2). We also observed concurrent mutations within the HAT subgroup between *CREBBP* and the highly homologous gene *EP300*. Distinct chromatin remodeling processes are often co-regulated and are functionally dependent[25]. The identification of concurrent genetic alterations in either the same or different subgroups of chromatin remodeling genes probably reflects that they can function synergistically or independently to promote tumorigenesis. We next investigated the relationship between the prevalence of each frequently mutated gene and tumor stage and/or grade. As previously shown[2], incidence of *TP53* mutations increased in TCCs of high stages or grades, whereas the frequency of *FGFR3* mutations decreased with increasing tumor stage and grade of TCC (Fig. 3). Similar to the frequency distribution of *FGFR3* mutations across different tumor stages and grades, our analysis indicated that mutations in *UTX* were likely to be associated with tumors of early developmental stage. The mutation frequency of *UTX* was significantly associated with tumor grade ($P = 0.008$), with grade 1 accounting for 35%, grade 2 accounting for 15% and grade 3 accounting for 4% of tumors with mutations in *UTX*. We also found a negative correlation between mutations in *UTX* and tumor stage. The incidences of *UTX* mutations in stage Ta and T1 and in ≥stage T2 tumors were 32% and 13%, respectively ($P = 0.04$). For the other chromatin remodeling genes (*ARID1A*, *CREBBP* and *EP300*), we found no significant correlations between their mutation frequencies and tumor stage or grade (Fig. 3). In summary, our study provides a comprehensive catalog of genetic alterations in TCC, and we discovered 49 new significantly mutated genes associated with TCC. Eight of these are

chromatin remodeling genes that harbor frequent mutations in the majority of NMI-TCCs and MI-TCCs. Most of these remodeling genes are likely to act as tumor suppressors and have been reported to play pivotal roles in various tumors other than TCC. These findings further our current understanding of bladder cancer and other human cancers and indicate that genetic alterations and epigenetic deregulations may cooperatively contribute to tumor genesis and progression. Our data can serve as a valuable basis for future studies on TCC and suggest the necessity of epigenomics research in the field of cancer studies.

URLs

COSMIC database, http://www.sanger.ac.uk/perl/genetics/CGP/cosmic (v52 release); NCBI Consensus Coding Region dataset, http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi; SAMtools, http://samtools.sourceforge.net/.

ONLINE METHODS

Sample description and preparation

The primary tumor samples and matched peripheral blood were obtained from individuals with TCC newly diagnosed at the member institutions of the Urinogenital Cancer Genomics Consortium (UCGC) in China. A signed written consent from each subject was obtained before recruitment for the study according to the regulations of the institutional ethics review boards. Detailed clinical information for the subjects is summarized in Supplementary Table 1. All the specimens were snap frozen in liquid nitrogen upon collection and immediately stored at −80 °C for further study. The hematoxylin eosin–stained sections prepared using the cancerous tissues were microscopically evaluated by two independent pathologists. In the present study, only TCCs with malignant cell purities of over 85% were selected for DNA extraction and subsequent sequencing.

Illumina based whole-exome sequencing and targeted exon sequencing

Genomic DNA for all tumors and blood samples from both the discovery and prevalence cohort were fragmented and hybridized to commercially available capture arrays for enrichment. In the discovery stage, the exome capture procedure was performed with NimbleGen 2.1M Human Exome Arrays, which are capable of enriching exonic sequences of more than 18,000 protein-coding genes deposited in the NCBI Consensus Coding Sequence Region database (see URLs). In both the discovery and prevalence screens, the enriched DNA fragments were sheared to 200 bp on average and subjected to standard Illumina Genome Analyzer library preparation according to Illumina's protocol. The shotgun libraries were sequenced with the Illumina GAII platform, and 80-bp single ended reads were generated. Image analysis and base calling were performed by the Genome Analyzer Pipeline version 1.3 using default parameters.

Read mapping and detection of somatic mutations

After removing reads containing sequencing adaptors and low quality reads with more than five unknown bases, the high quality single-end reads were aligned to the NCBI human reference genome (hg18) using MAQ[3] with the default options. To identify indels, the high-

quality reads were gap aligned to the reference sequence using BWA[4]. We then performed local realignment of the BWA-aligned reads using the Genome Analysis Toolkit (GATK)[26]. The raw lists of potential somatic substitutions were called by VarScan[27] (v2.2) based on the MAQ alignments. In this process, several heuristic rules were applied: (i) both the tumor and matched normal samples should be covered sufficiently (≥10×) at the genomic position being compared; (ii) the average base quality for a given genomic position should be no less than 15 in both the tumor and normal samples; (iii) the variants should be supported by at least 10% of the total reads in the tumors, and no high-quality variant-supporting reads were allowed in normal controls; and (iv) the variants should be supported by at least five reads in the tumors. Using the same criteria, the preliminary list of somatic indels was called out by GATK based on the local realignment results. After these two steps, germline variants could be effectively removed. To further reduce the false positive calls, variations including single nucleotide variants and indels were called with the SAMtools software package (see URLs) in the tumors. We eliminated all somatic variants that fulfilled any one of the following filtering criterion: (i) variants with Phred-like scaled consensus scores or SNP qualities <20; (ii) variants with mapping qualities <30; (iii) indels represented by only one DNA strand; or (iv) substitutions located 30 bp around the predicted indels. To deal with false positives associated with pseudo-gene issues or repeat sequences, simulated reads (80 bp in length) containing the putative mutations were generated and aligned to the reference genome. For a given variant, if more than 10% of the simulated variant-containing reads could not be uniquely mapped to the reference genome, this variant was discarded.

Validation of somatic mutations by mass spectrum or Sanger sequencing

Validation of the non-silent somatic substitutions by mass spectrum was performed with the MassArray platform of Sequenom by determining their genotypes in the tumors and matched blood. The genotyping assay and base-calling procedures were performed as previously described[28]. We considered the genotyping assay to have failed if the Sequenom software was unable to design primers for PCR amplification or base extension at the primer design stage or if the observed peak for a given assay was not significant enough for a confident call at the base-calling stage. To validate somatic non-silent substitutions and indels with Sanger sequencing, PCR primers designed for the putative somatic variants were initially used to amplify the source DNA from the tumors. If the mutations were successfully confirmed in the tumors, the same primer pairs were used to amplify the normal DNA from the same subjects to determine the somatic statuses of the observed mutations.

Statistical analysis of the significantly mutated genes

To estimate the background mutation rates, we conservatively treated all the predicted mutations identified in the discovery screen as passenger mutations. These predicted mutations were classified into 19 different categories according to their sequence context and mutation types. For each mutation category, i, we let the observed number of mutations of the category be m_i and the total number of successfully sequenced nucleotides (≥10×) of this category in the nine tumors be n_i; the background mutation rate for this category, b_i, was calculated as m_i/n_i. The estimated background mutation rates for each category are listed in Supplementary Table 6. To test whether the non-silent mutation rate of a gene was significantly higher than the background, the confirmed mutation data for the gene obtained

from the discovery and prevalence screens were combined. Then we estimated the passenger probability for each gene in turn as previously described[23]. To be specific, the probability (p_{gi}) of obtaining the observed number of mutations of each category (i) in gene g was estimated from a binominal distribution, with b_i as the success probability. The number of available nucleotides for each category was the total number of sufficiently covered (≥10×) bases for that particular category in all the 97 TCCs. The passenger probability (p_g) for gene g was calculated to be the product of the 19 category-specific probabilities, or $p_g=\Pi_{i=1}^{19} p_{gi}$. We then determined the P value for each gene by the likelihood-ratio test as previously described[29].

Statistical analysis of inactivating mutations in chromatin remodeling genes

To assess whether nonsense mutations were over-represented in the eight chromatin remodeling genes, we calculated the probability (Pn) of single base changes that would result in nonsense mutations by chance. To this end, the coding sequence of each gene was represented by its longest transcript, and every single base in the coding region was changed into the other three different bases. We obtained P_n for each gene by dividing the number of nonsense mutations observed in each gene by the length of the coding region. The significance of nonsense mutation enrichment was determined by a binomial test with the parameter P_n as the hypothesized probability of success.

Concurrence and mutual exclusion analysis

We performed the concurrence and mutual exclusion analysis on the significantly mutated genes that showed non-silent mutations in at least five tumors by permutation test as previously described[30], with minor modifications. Briefly, we kept the number of samples mutated in a given gene the same as that observed in our data and permutated the mutations across samples. For a given sample, s, we let the number of the significantly mutated genes showing mutations in that particular sample be g_s, and the probability (p_s) that the sample s was labeled as mutated in the permutation test was defined as $g_s/\sum_{s=1}^{97} gs$. For each permutation, we recorded the number of samples with concurrent and exclusive mutations (P_c and P_e, respectively) for each pair of genes and compared them with numbers observed in original data (N_c and N_e). We repeated this process 10,000,000 times and summarized the frequencies of $P_c \geq N_c$ and $P_e \geq N_e$, respectively. These frequencies were used as the empirical P values under the null hypothesis (that is, assuming no correlation between genes).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Yaoting Gui[1,12], Guangwu Guo[2,12], Yi Huang[1,12], Xueda Hu[2,12], Aifa Tang[1,3,12], Shengjie Gao[2], Renhua Wu[2], Chao Chen[2], Xianxin Li[1], Liang Zhou[1], Minghui He[2], Zesong Li[1,3], Xiaojuan Sun[3], Wenlong Jia[2], Jinnong Chen[2], Shangming Yang[2], Fangjian Zhou[4], Xiaokun Zhao[5], Shengqing Wan[2], Rui Ye[2], Chaozhao Liang[6],

Zhisheng Liu[2], Peide Huang[2], Chunxiao Liu[7], Hui Jiang[2], Yong Wang[1], Hancheng Zheng[2], Liang Sun[1], Xingwang Liu[2], Zhimao Jiang[1], Dafei Feng[2], Jing Chen[1], Song Wu[1], Jing Zou[2], Zhongfu Zhang[1], Ruilin Yang[1], Jun Zhao[1], Congjie Xu[1], Weihua Yin[1], Zhichen Guan[1], Jiongxian Ye[1], Hong Zhang[1], Jingxiang Li[2], Karsten Kristiansen[2,8], Michael L Nickerson[9], Dan Theodorescu[10,11], Yingrui Li[2], Xiuqing Zhang[2], Songgang Li[2], Jian Wang[2], Huanming Yang[2], Jun Wang[2,8], and Zhiming Cai[1,3]

Affiliations

[1]Guangdong and Shenzhen Key Laboratory of Male Reproductive Medicine and Genetics, Institute of Urology, Peking University Shenzhen Hospital, Shenzhen PKU-HKUST Medical Center, Shenzhen, China

[2]BGI-Shenzhen, Shenzhen, China

[3]Shenzhen Second People's Hospital, The First Affiliated Hospital of Shenzhen University, Shenzhen, China

[4]Department of Urology, Sun Yat-Sen University Cancer Center, Guangzhou, China

[5]Department of Urology, the Second Xiangya Hospital of Central-Southern University, Changsha, China

[6]Department of Urology, The First Affiliated Hospital of Anhui Medical University, Hefei, China

[7]Department of Urology, Zhujiang Hospital, Southern Medical University, Guangzhou, China

[8]Department of Biology, University of Copenhagen, Copenhagen, Denmark

[9]Cancer and Inflammation Program, National Cancer Institute, US National Institutes of Health (NIH), Frederick, Maryland, USA

[10]Department of Surgery, University of Colorado Comprehensive Cancer Center, Aurora, Colorado, USA

[11]Department of Pharmacology, University of Colorado Comprehensive Cancer Center, Aurora, Colorado, USA

Acknowledgments

This work was supported by grants from the National High Technology Research and Development Program of China (863 Program, 2006AA02A302 to H.Y. 2009AA022707 to X.Z.), the Promotion Program for Shenzhen Key Laboratory, Shenzhen, China (CXB200903090055A and CXB201005250016A to Z.C.), and Research Fund for the Doctoral Program of Higher Education of China (20100001110100).

References

1. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin. 2005; 55:74–108. [PubMed: 15761078]
2. Wu XR. Urothelial tumorigenesis: a tale of divergent pathways. Nat Rev Cancer. 2005; 5:713–725. [PubMed: 16110317]

3. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 2008; 18:1851–1858. [PubMed: 18714091]
4. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]
5. Greenman C, et al. Patterns of somatic mutation in human cancer genomes. Nature. 2007; 446:153–158. [PubMed: 17344846]
6. Cordon-Cardo C, et al. p53 mutations in human bladder cancer: genotypic versus phenotypic patterns. Int J Cancer. 1994; 56:347–353. [PubMed: 7906253]
7. Cairns P, Proctor AJ, Knowles MA. Loss of heterozygosity at the *RB* locus is frequent and correlates with muscle invasion in bladder carcinoma. Oncogene. 1991; 6:2305–2309. [PubMed: 1766677]
8. Jebar AH, et al. *FGFR3* and Ras gene mutations are mutually exclusive genetic events in urothelial cell carcinoma. Oncogene. 2005; 24:5218–5225. [PubMed: 15897885]
9. Cappellen D, et al. Frequent activating mutations of *FGFR3* in human bladder and cervix carcinomas. Nat Genet. 1999; 23:18–20. [PubMed: 10471491]
10. Guenther MG, Barak O, Lazar MA. The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. Mol Cell Biol. 2001; 21:6091–6101. [PubMed: 11509652]
11. Lutz T, Stoger R, Nieto A. CHD6 is a DNA-dependent ATPase and localizes at nuclear sites of mRNA synthesis. FEBS Lett. 2006; 580:5851–5857. [PubMed: 17027977]
12. van Haaften G, et al. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nat Genet. 2009; 41:521–523. [PubMed: 19330029]
13. Wiegand KC, et al. *ARID1A* mutations in endometriosis-associated ovarian carcinomas. N Engl J Med. 2010; 363:1532–1543. [PubMed: 20942669]
14. Dalgliesh GL, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature. 2010; 463:360–363. [PubMed: 20054297]
15. Varela I, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. Nature. 2011; 469:539–542. [PubMed: 21248752]
16. Chi P, Allis CD, Wang GG. Covalent histone modifications–miswritten, misinterpreted and mis-erased in human cancers. Nat Rev Cancer. 2010; 10:457–469. [PubMed: 20574448]
17. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med. 2004; 10:789–799. [PubMed: 15286780]
18. Mullighan CG, et al. *CREBBP* mutations in relapsed acute lymphoblastic leukaemia. Nature. 2011; 471:235–239. [PubMed: 21390130]
19. Pasqualucci L, et al. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. Nature. 2011; 471:189–195. [PubMed: 21390126]
20. Karamouzis MV, Konstantinopoulos PA, Papavassiliou AG. Roles of CREB-binding protein (CBP)/p300 in respiratory epithelium tumorigenesis. Cell Res. 2007; 17:324–332. [PubMed: 17372613]
21. Daser A, Rabbitts TH. The versatile mixed lineage leukaemia gene *MLL* and its many associations in leukaemogenesis. Semin Cancer Biol. 2005; 15:175–188. [PubMed: 15826832]
22. Parsons DW, et al. The genetic landscape of the childhood cancer medulloblastoma. Science. 2011; 331:435–439. [PubMed: 21163964]
23. Sjöblom T, et al. The consensus coding sequences of human breast and colorectal cancers. Science. 2006; 314:268–274. [PubMed: 16959974]
24. Bakkar AA, et al. *FGFR3* and *TP53* gene mutations define two distinct pathways in urothelial cell carcinoma of the bladder. Cancer Res. 2003; 63:8108–8112. [PubMed: 14678961]
25. Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007; 128:683–692. [PubMed: 17320506]
26. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–1303. [PubMed: 20644199]
27. Koboldt DC, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 2009; 25:2283–2285. [PubMed: 19542151]
28. Yi X, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010; 329:75–78. [PubMed: 20595611]

29. Getz G, et al. Comment on “The consensus coding sequences of human breast and colorectal cancers”. Science. 2007; 317:1500.

30. Ding L, et al. Somatic mutations affect key pathways in lung adenocarcinoma. Nature. 2008; 455:1069–1075. [PubMed: 18948947]

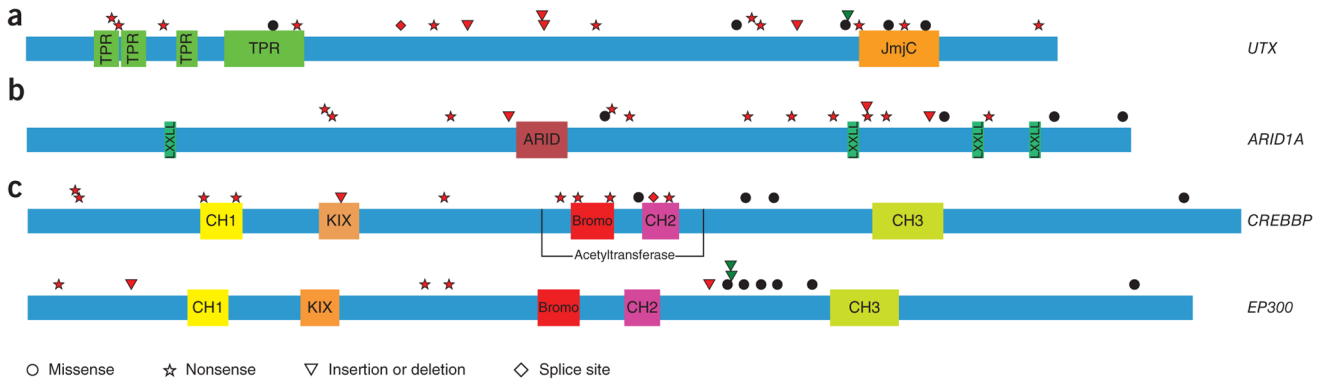

Figure 1.

Somatic mutations in *UTX*, *ARID1A* and *CREBBP-EP300*. The types and relative positions of confirmed somatic mutations are shown in the transcripts of *UTX* (**a**), *ARID1A* (**b**) and *CREBBP-EP300* (**c**) using the following symbols: red stars, nonsense mutations; bullets, missense mutations; red triangles, frame-shift indels; green triangles, in-frame indels; and diamond, mutations at splice sites. Domains and motifs in each encoded protein product, as well as the key region responsible for the histone acetyltransferase activity of CREBBP, are also indicated. TPR, tetratricopeptide repeat; JmjC, transcription factor jumonji/aspartyl beta-hydroxylase; ARID, AT-rich interactive domain; LXXLL, C-terminal leucine-rich LXXLL motif; CH1, CH2 and CH3, cysteine–histidine-rich domains; KIX, CREB-binding domain; Bromo, bromodomain.

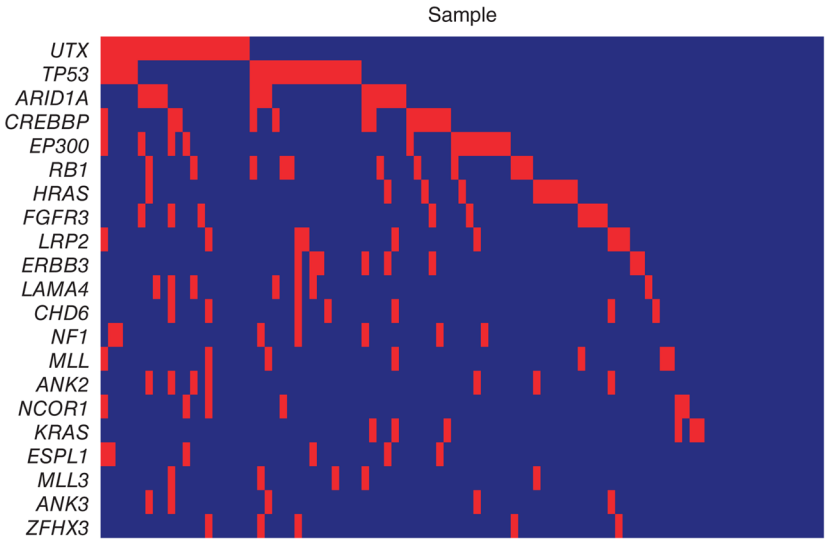

Figure 2.

Concurrent and mutually exclusive mutations observed in the frequently mutated genes. For each gene (row) indicated, tumors (columns) with or without mutations are labeled in red or blue, respectively. *P* values for the occurrence of concurrent and mutual exclusive mutations in two genes across tumors are provided in Supplementary Table 8. We selected only genes harboring non-silent mutations in at least five subjects for this analysis.

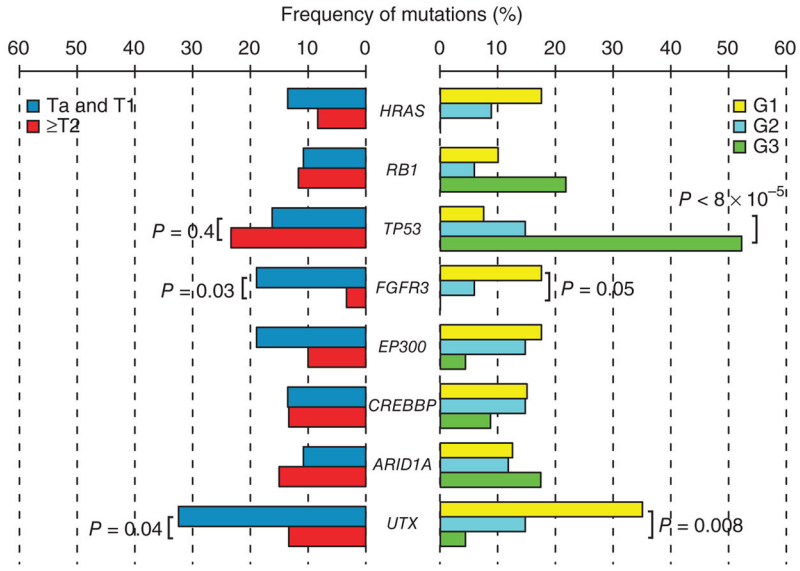

Figure 3.

Frequencies of mutations in highlighted genes across different tumor stages and grades. T, the stage of tumors under the TNM (tumor, lymph node and distant metastasis) classification system; G, grade of tumors. For each tumor stage or grade, we calculated the frequency of mutations in a given gene as the proportion of tumors harboring no silent mutations in the gene among all tumors of the indicated stage or grade. We determined the significance of the correlation between mutations in each gene (altered in at least 10% of TCCs) and tumor grade or stage using χ^2 tests.

Table 1

Frequently mutated genes identified in 97 TCCs

	Non-silent somatic changes						
Gene	**Missense**	**Nonsense, splice site or indel**	**Synonymous mutations**	**Total mutations**	**N:S ratio**	**Number (%)of subjects harboring non-silent mutations**	***P***
UTX[a]	5	17	1	23	22:1	20 (21)	2.22×10^{-37}
TP53	13	8	1	22	21:1	20 (21)	1.01×10^{-49}
CREBBP[a]	4	11	2	17	15:2	13 (13)	3.42×10^{-26}
EP300[a]	6	7	0	13	-	13 (13)	6.58×10^{-15}
ARID1A[a]	4	14	0	18	-	13 (13)	1.99×10^{-32}
RB1	3	8	1	12	11:1	11 (11)	1.00×10^{-11}
HRAS	10	0	0	10	-	10 (10)	7.89×10^{-7}
ERBB3	9	0	2	11	9:2	8 (8)	3.28×10^{-18}
CHD6[a]	6	1	2	9	7:2	7 (7)	6.00×10^{-9}
LRP2	8	1	0	9	-	9 (9)	4.47×10^{-10}
FGFR3	9	0	0	9	-	9 (9)	1.69×10^{-18}
MLL[a]	5	3	1	9	8:1	7 (7)	2.00×10^{-11}
LAMA4	8	0	1	9	8:1	7 (7)	4.47×10^{-9}
ANK2	7	0	1	8	7:1	7 (7)	3.07×10^{-8}
NF1	5	2	1	8	7:1	7 (7)	8.76×10^{-10}
ESPL1	5	1	1	7	6:1	6 (6)	2.76×10^{-11}
NCOR1[a]	5	2	0	7	–	6 (6)	3.21×10^{-6}
KRAS	6	0	1	7	6:1	6 (6)	3.29×10^{-26}
MLL3[a]	2	4	1	7	6:1	5 (5)	1.58×10^{-6}
ANK3	5	0	1	6	5:1	5 (5)	9.10×10^{-4}
ZFHX3	5	0	0	5	–	5 (5)	5.52×10^{-3}

[a]Chromatin remodeling genes.–, no synonymous mutation was identified in this gene. We confirmed all non-silent somatic changes shown in the table by Sanger sequencing or Sequenom MassARRAY. We calculated *P* values based on the number of validated non-silent mutations in each gene. Detailed information on each validated mutation is listed in Supplementary Table 5.