HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2020 September 24.

Published in final edited form as:

Nat Genet. 2013 December ; 45(12): 1459–1463. doi:10.1038/ng.2798.

Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation

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

Abstract

Bladder cancer is one of the most common cancers worldwide, with transitional cell carcinoma (TCC) being the predominant form. Here we report a genomic analysis of TCC by both whole-genome and whole-exome sequencing of 99 individuals with TCC. Beyond confirming recurrent mutations in genes previously identified as being mutated in TCC, we identified additional altered genes and pathways that were implicated in TCC. Notably, we discovered frequent alterations in *STAG2* and *ESPL1*, two genes involved in the sister chromatid cohesion and segregation (SCCS) process. Furthermore, we also detected a recurrent fusion involving *FGFR3* and *TACC3,* another component of SCCS, by transcriptome sequencing of 42 DNA-sequenced tumors. Overall, 32 of the 99 tumors (32%) harbored genetic alterations in the SCCS process. Our analysis provides evidence that genetic alterations affecting the SCCS process may be involved in bladder tumorigenesis and identifies a new therapeutic possibility for bladder cancer.

Bladder cancer is one of the most common genitourinary malignancies in the world, with an estimated 386,300 new cases and 150,200 deaths in 2008 alone[1]. Previous studies showed that bladder cancer represents a heterogeneous disease, with two distinct subtypes—superficial and invasive—showing variable clinical presentation and genetic background[2]. Recently, we reported frequent mutations in eight chromatin-remodeling genes (*UTX, MLL-MLL3, CREBBP-EP300, NCOR1, ARID1A* and *CHD6*) in TCC of the urinary bladder[3].

Correspondence should be addressed to Z.C. (caizhiming2000@163.com), Jun Wang (wangj@genomics.org.cn) or Y.G. (guiyaoting@hotmail.com).

AUTHOR CONTRIBUTIONS

Jun Wang, Z.C., Jian Wang, H.Y., Y. Li, X. Zhang and Y.G. managed the project. X.S., S. Wu, Z. Li, A.T., X.L., Xiaokun Zhao, S.Z., M.Q., L. Xie, X. Zou, L. Xing, Z. Lv, H.M., C. Liang, J.L., C. Liu, C. Li, J.C., Y. Lai, Zheguang Lin and F. Zhou prepared the samples. P.H., P.S., F.F., Y.Y. and Xin Zhao performed the sequencing. G.G., C.C., Y.H., W.J., Q.Z., Z.Y., R.Y., Zhao Lin, S. Wan, M.L.N., M.D., S.G., Z.G., L.L., X.F., J.Y., F. Zhang, S.T. and D.T. performed the bioinformatic analysis. X.S., P.H., Z. Li, P.S., F.F., X.H., Z.J., H.C. and H.C.C. performed the validation of somatic mutations, CNAs and gene fusion events. G.G. wrote the manuscript. G.G., X.S., C.C., S. Wu, P.H. and Z. Li revised the manuscript.

URLs. 1000 Genomes Project, http://www.1000genomes.org/; KEGG (Kyoto Encyclopedia of Genes and Genomes) database, http://www.genome.jp/kegg/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. All sequencing data from this study have been deposited in the Sequence Read Archive (SRA) under accession SRA063495.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Nevertheless, a comprehensive catalog of genetic alterations is far from complete, and the key drivers of TCC tumorigenesis remain poorly understood.

To improve understanding of the genetic basis of TCC, we performed whole-exome sequencing of tumor and matched peripheral blood samples from 99 individuals with TCC (Supplementary Fig. 1 and Supplementary Tables 1 and 2). After several rigorous bioinformatic analysis steps (Online Methods), we identified 11,240 candidate somatic mutations, and validation experiments on 1,119 predicted somatic substitutions and 91 indels confirmed 1,023 (91%) and 67 (74%), respectively, by Sanger sequencing (Supplementary Tables 3 and 4). We also analyzed DNA from the same cohort by whole-genome sequencing to detect copy number alterations (CNAs) and obtained fourfold mean haploid coverage for each sample (Supplementary Fig. 2). In addition, we performed transcriptome sequencing (RNA-seq) on a subset of 42 DNA-sequenced tumors, including 16 for which matched, morphologically normal bladder tissue was available (Supplementary Table 5).

To identify genes whose mutations were associated with TCC, we assessed the statistical significance of the observed mutation prevalence for each gene as described previously[3]. In total, we identified 37 significantly mutated genes (Fig. 1 and Supplementary Table 6), which included 7 well-known bladder cancer genes (*TP53* (ref. 4), *HRAS*[5], *FGFR3* (ref. 6), *PIK3CA*[7], *RB1* (ref. 8), *KRAS*[5] and *TSC1* (ref. 9)). Consistent with our previous study of TCC[3], eight chromatin-remodeling genes (*UTX*, *ARID1A*, *MLL-MLL3*, *CREBBP-EP300*, *NCOR1* and *CHD6*) were significantly mutated in the cohort (Supplementary Fig. 3). In addition, we examined mutation frequencies for genes and gene families (Supplementary Table 4) and observed frequent mutations in multiple tumors in other chromatin-remodeling genes, including the histone demethylase genes *KDM6A* (*UTX*) and *UTY*(30%), the chromatin-remodeler genes *ARID1A* and *ARID4A* (17%), the histone lysine methyltransferase genes *KMT2A* (*MLL*), *KMT2C* (*MLL3*) and *KMT2E* (*MLL5*) (16%), the histone acetyltransferase genes *EP300* and *EP400* (15%), the SWI/SNF complex–related genes *SMARCA4* and *SMARAC1* (7%) and the histone demethylase genes *KDM5A* (*JARID1A*) and *KDM5B* (*JARID1B*) (6%). In total, at least 57 of the 99 cases (58%) harbored somatic mutations in chromatin-remodeling genes, further indicating that altered epigenetic regulation of chromatin and chromatin post-translational modifications may be a major driver mechanism in TCC[3].

We also found 13 new significantly mutated genes that have not previously been reported in TCC (Fig. 1). *STAG2* (encoding stromal antigen 2) was particularly notable, ranking first in mutational significance among the 13 newly identified mutated genes and harboring a greater number of nonsynonymous mutations ($P = 8 \times 10^{-11}$) and a higher ratio of nonsynonymous to synonymous mutations ($P = 0.02$) than expected by chance. *STAG2*, a gene on the X chromosome, encodes a component of the cohesin complex associated with the SCCS process of the cell cycle, which regulates the separation of sister chromatids during cell division. We found that 11 cases (11%) harbored mutations in *STAG2*, and 9 of the 11 mutations were truncating mutations (3 small frameshifting indels and 4 nonsense and 2 splice-site mutations) (Fig. 2a and Supplementary Fig. 4). In addition, *STAG2* genomic deletions were observed in 5 of the 99 tumors (Fig. 2b and Supplementary Fig. 5). By

screening all exons of *STAG2* in an additional 50 tumor-normal pairs form individuals with TCC by Sanger sequencing, we identified 5 somatic mutations in 4 tumors (Supplementary Table 7). We also examined the methylation status of the *STAG2* promoter in 30 TCCs by bisulfite sequencing and found *STAG2* promoter hypermethylation in 7 tumors (23%) relative to matched normal samples (Supplementary Table 8).

To explore the association between *STAG2* alterations and individual survival, we performed Kaplan-Meier survival analysis on our data and found that individuals with *STAG2* somatic alterations had a much worse prognosis compared to individuals with wild-type *STAG2* (Fig. 2c). Similarly, we also observed significant differences ($P < 0.001$) in survival between cases with and without *STAG2* somatic alterations in the superficial and invasive subtypes of TCC (Supplementary Fig. 6), suggesting that *STAG2* may be an independent predictor of worse outcome in TCC. A previous report showed that inactivation of *STAG2* causes chromatid cohesion defects and aneuploidy in Ewing sarcoma, glioblastoma and melanoma[10]. In our study, we evaluated aneuploidy by quantifying the extent of copy number changes on chromosome arms and found that tumors with genetic alterations in *STAG2* had more aneuploidy ($P = 0.01$) than tumors with wild-type SCCS genes (Supplementary Fig. 7), which is consistent with a published study[10]. In contrast with the observation in solid tumors, many *STAG2*-mutated primary leukemias have completely normal karyotypes[11,12]. These contradictory observations in different types of cancer suggest that further comprehensive investigation of *STAG2* is warranted to fully explore its exact role in causing aneuploidy and in TCC tumorigenesis. In addition to aberrations in well-known and emerging cancer genes, our study also identified frequent mutations in several genes not yet associated with TCC, such as *ERCC2* (Supplementary Fig. 8), *TRRAP* and *FAT4* (Supplementary Note), and further experiments are highly recommended to investigate the functions of these genes in TCC.

We also profiled the 99 tumors for CNAs using whole-genome sequencing (Supplementary Fig. 9) and found abnormalities of chromosomal arms or entire chromosomes that predominantly involved gain of 5p, 8q, 13p, 20p and 20q and loss of 8p, 9p, 9q, 11p, 14p, 15p, 17p and 21p (Supplementary Fig. 10). The patterns of broad cytogenetic gain and loss were consistent with previous studies[13], and no significantly disparate pattern was observed across the different subgroups of TCC. Profiling of CNAs identified many putative cancer driver genes that may be implicated in TCC tumorigenesis (Supplementary Fig. 11). We applied GSITIC[14] analysis to whole-genome sequencing data to identify recurrent focal CNAs, and this analysis yielded 84 regions of focal amplification (Supplementary Fig. 12 and Supplementary Table 9), several of which included CNAs previously detected in bladder cancer[13], which encompassed genes such as *TRIO*, *MDM2*, *MYC*, *E2F3*, *CCND1* and *ERBB2* (Supplementary Figs. 13-15). Other amplifications defined regions of CNAs reported for the first time, to our knowledge, in bladder cancer, including *CCNE1*, *CEBPA*, *E2F1* and *MUC1* (Supplementary Fig. 12). An interesting finding was frequent amplification of *DHFR*, encoding dihydrofolate reductase, which is a target of many anticancer agents[15], at 5q (frequently loss in our cohort) in 14 tumors (14%) (Supplementary Figs. 16 and 17 and Supplementary Note). We also identified 80 recurrent focal deletion regions, including *RB1* and *CREBBP*, which were also frequently truncated in TCCs. One of the most common focal deletions, detected in 50 tumors (50%), was a deleted region at 9p21 containing

CDKN2A and *CDKN2B* (Supplementary Fig. 18) and is a widely reported genomic alteration in bladder cancer[16].

We next searched for genome rearrangements using the RNA-seq data set instead of the whole-genome sequencing data set owing to the low sequence coverage achieved in whole-genome sequencing. Overall, we detected 32 candidate rearrangements that resulted in gene fusions (Supplementary Table 10). The only recurrent fusion event involved *FGFR3* fusion in frame with *TACC3* in cases B59-3 and B100 (2/42, 5%; Fig. 3a and Supplementary Figs. 19 and 20). Analysis of junction-spanning and mate-pair reads derived from the whole-genome sequencing data confirmed the presence of the *FGFR3-TACC3* fusion in both tumors. *TACC3*, which is located 70 kb away from *FGFR3*, encodes a microtubule-associated protein that is important for spindle stability and organization in the SCCS process and has been mapped to a chromosomal region that is disrupted in some cancers[17]. In expression analysis from RNA-seq coverage, the B59-3 and B100 tumors had markedly higher, outlier expression of *TACC3* compared to normal bladder tissues and other tumors without the *FGFR3-TACC3* fusion (Fig. 3b), and *TACC3* mRNA was predominantly present in fused transcripts in both tumors (Fig. 3c and Supplementary Fig. 20c). Although *FGFR3* and *TACC3* are located in a focal region that was frequently amplified in our cohort (Supplementary Fig. 12), the low level of amplification of this region in the B100 tumor and its apparent deletion in the B59-3 tumor (Supplementary Fig. 21) indicate that the high expression of *TACC3* was unlikely to be caused by *TACC3* amplification but is likely mediated by transcriptional regulatory elements in the promoter of its fusion partner, *FGFR3*. A previous study reported a similar *FGFR3-TACC3* fusion in glioblastoma multiforme and found that the fusion protein could induce mitotic and chromosomal segregation defects and trigger aneuploidy[18]. A low frequency of *FGFR3-TACC3* fusions (9%) was also reported in a recent study using bladder cancer cell lines[19]. In addition to mutations in *FGFR3*, frequent amplification of the locus with *TACC3* and *FGFR3* in our cohort and recurrent *FGFR3-TACC3* fusion events suggest a close relationship between this locus and bladder tumorigenesis. This idea is further supported by the discovery of an association between a SNP in this locus and increased risk of bladder cancer[20].

To construct a comprehensive view of the common genetic alterations underlying the TCC genome, we performed integrative pathway analysis of the somatic mutations and CNAs. In addition to chromatin remodeling, we also discovered a number of pathways that may be implicated in TCC, including the Ras–mitogen-activated protein kinase (MAPK) and phosphoinositol 3-kinase (PI3K)-AKT signaling pathways that were commonly altered in bladder cancer[2,13,21] (Supplementary Fig. 22 and Supplementary Table 11). The cell cycle was the second most frequently altered pathway, with alterations in 86 of the tumors (86%), including that alterations in genes with known roles in the G1/S checkpoint (loss-of-function mutations in *RB1*, amplifications of *MYC*, *CCND1*, *CCNE1*, *E2F1* and *E2F3*, and deletions of *CDKN2A* and *CDKN2B*) and the G2/M checkpoint (inactivating mutations in *ATM*, *ATR*, *CREBBP*, *EP300*, *BRCA1*, *BRCA2* and *TP53* and amplifications of *MDM2*). Specifically, genes involved in the SCCS process were frequently altered in TCCs (Fig. 4). *STAG2*, *NIPBL*, *SMC1A* and *SMC3*, four genes with important roles in sister chromatid cohesion, were altered in 16%, 4%, 3% and 2% of the tumors, respectively. We also observed frequent mutations in *ESPL1* (6% of tumors), a gene encoding a protein with a

central role in chromosome segregation through its cleavage of the cohesin complexes that hold sister chromatids together. Moreover, *TACC3*, also critical for the SCCS process, was alerted by genome rearrangement in 5% of the tumors. Rare mutations in the other SCCS genes *BUB1*, *BUBR1B*, *BUB3*, *MAD1L1* (*MAD1*), *MAD2* (*MAD2L1*) and *CENPE* have been reported in several types of cancer, including bladder cancer22, but no genomic alterations in these genes were observed in our cohort. In total, we identified genetic aberrations in the SCCS process in 32 of our 99 subjects with TCC (32%). The SCCS process, a major cell cycle control mechanism during mitosis, prevents chromosome missegregation if spindle assembly is perturbed, and misfunction of the SCCS process can generate chromosomal instability and aneuploidy, the most common characteristic of human solid tumors. As expected, we found that tumors with alterations in SCCS process genes had significantly higher aneuploidy ($P = 0.01$) than tumors without alterations in these genes (Supplementary Fig. 23). Up to now, only a few SCCS process genes have been shown to be altered at an appreciable frequency[10,23]. Discovery of high-frequency alterations in the SCCS process shows that genetic disruption of the genes directly controlling chromosome cohesion and segregation might have an important role in bladder tumorigenesis.

In the present study, we performed genomic analysis of 99 TCCs by whole-genome sequencing, whole-exome sequencing and RNA-seq. Although neither whole-genome sequencing nor whole-exome sequencing in our study achieved high coverage and more sequencing data are needed to identify additional rare mutations, our analysis provides a more comprehensive catalog of genomic alterations in TCC. Notably, we found highly frequent genomic alterations in genes involved in the SCCS process, including alterations in *STAG2*, *ESPL1* and *NIPBL*, and a recurrent *FGFR3-TACC3* fusion in TCC. In contrast to other cancers in which only rare or low-frequency alterations in SCCS genes have been observed[10,22,23], bladder cancer is the first type of cancer, to our knowledge, with predominant genetic lesions in genes involved in the SCCS process. Although we do not yet understand the detailed mechanisms by which *STAG2* may contribute to TCC pathogenesis, the discovery of highly recurrent alterations in the genes involved in the SCCS process in TCC identifies yet another pathway that is undoubtedly relevant for TCC. Our analysis suggests the necessity of comprehensive investigation of SCCS genes to elucidate their exact roles in causing aneuploidy or bladder tumorigenesis.

ONLINE METHODS

Sample description and preparation.

Tumor samples with matched peripheral blood or normal controls (morphologically adjacent normal bladder tissues) were obtained from individuals newly diagnosed with TCC at the member institutions of the Urinogenital Cancer Genomics Consortium (UCGC) in China. Each subject was properly informed 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 specimens were snap frozen in liquid nitrogen upon collection and were immediately stored at −80 °C for further study. Sections stained with hematoxylin and eosin, prepared using cancerous tissues, were microscopically

evaluated by two independent pathologists. In the present study, only TCCs with malignant cell purities over 85% were selected for DNA extraction and subsequent sequencing.

Genomic DNA extraction and Illumina-based whole-genome and whole-exome sequencing.

Genomic DNA from tumor and matched peripheral blood samples for the 99 individuals with TCC was isolated using QIAamp DNA Mini kits (Qiagen), and DNA libraries were constructed according to the protocol provided by the manufacturer (Illumina). Whole-genome sequencing data were generated using the Illumina HiSeq 2000 platform in 2 × 100-bp paired-end reads.

For whole-exome sequencing, genomic DNA from the same 99 tumor-blood pairs was fragmented and subjected to whole-exome capture with the SureSelect Human All Exon 50Mb kit (Agilent Technologies) following the manufacturer's protocols. Exome capture libraries were then sequenced on the HiSeq 2000 platform according to the manufacturer's instructions, and 2 × 100-bp paired-end reads were generated.

Total RNA extraction and Illumina-based RNA-seq.

RNA from 42 bladder tumors and 16 matched normal bladder tissues was used to generate mRNA-seq libraries using the TruSeq RNA Sample Preparation kit (Illumina). mRNA-seq libraries were then sequenced on the HiSeq 2000 platform according to the manufacturer's recommendations, and 2 × 90-bp paired-end reads were generated. After removing reads containing sequencing adaptors and reads of low quality, we aligned reads to the human genome (hg18) and Ensembl annotated genes (release 64) using SOAP2 (ref. 24), allowing up to two mismatches. Gene expression levels based on RNA-seq data were measured in RPKM[25].

Whole-exome sequencing read mapping and detection of somatic mutations.

After removing whole-exome sequencing reads containing sequencing adaptors and low-quality reads with more than five unknown bases, high-quality paired-end reads were gap aligned to the NCBI human reference genome (hg18) using Burrows-Wheeler Aligner (BWA)[26]. We then performed local realignment of the BWA-aligned reads using the Genome Analysis Toolkit (GATK)[27]. Raw lists of potential somatic substitutions were called by VarScan (v2.2)[28] on the basis of BWA alignments. In this process, several heuristic rules were applied: (i) both tumor and matched blood samples should be covered sufficiently (≥8~) at the genomic positions being compared; (ii) average base quality for a given genomic position should be no less than 15 in both the tumor and blood samples; (iii) variants should be supported by at least 20% of the total reads from tumors, and no high-quality variant-supporting reads were allowed in blood samples; and (iv) variants should be supported by at least 3 reads in tumors. Using these same criteria, the preliminary list of somatic indels was called out by GATK on the basis of local realignment results. After these two steps, germline variants could be effectively removed. To further reduce the number of false positive calls, variations in tumors, including single-nucleotide variants (SNVs) and indels, were called with the SAMtools software package. We eliminated all somatic variants that fulfilled any one of the following filtering criteria: (i) variants with Phred-like scaled consensus scores or SNP qualities of <20; (ii) variants with mapping qualities of <30; (iii)

indels represented by only one DNA strand; and (iv) substitutions located within 30 bp of a dubious indel. To eliminate previously described germline variants, somatic mutations were cross-referenced against the dbSNP (version 132) database and SNP data sets from the 1000 Genomes Project. Any mutations present in these data sets were filtered out, and the remaining mutations were subjected to subsequent analyses.

Validation of somatic substitutions and indels by Sanger sequencing.

Non-silent somatic substitutions and indels were validated by Sanger sequencing based on PCR amplification. PCR primers for putative somatic variants were designed by Primer3 *in silico* and initially used to amplify source DNA from tumors. PCR was performed on the Dual 96-Well GeneAmp PCR System 9700 (Applied Biosystems), and 20 ng of template DNA from each sample was used per reaction. Products were sequenced with the 3730xl DNA Analyzer (Applied Biosystems). All sequences were analyzed with Sequencing Analysis Software Version 5.2 (Applied Biosystems). If mutations were successfully confirmed in tumors, the same primer pairs were used to amplify DNA from the blood of the same subjects to determine the somatic status of the observed mutations.

Analysis of significantly mutated genes.

The background mutation rate was estimated on the basis of the number of synonymous mutations identified by whole-exome sequencing in the 99 tumors and was defined as the product of the synonymous mutation rate and the ratio of nonsynonymous to synonymous mutations (1.4) observed in the HapMap database. Briefly, synonymous somatic mutations were classified into seven categories according to their sequence contexts and mutation types, as described in our previous study[29]. For each mutation category, i, we let the observed number of mutations for this category be m_i and the total number of successfully sequenced nucleotides (≥8×) for this category in the 99 tumors be n_i. The background mutation rate for this category, b_i, was calculated as $1.4 \times m_i/n_i$. To test whether the rate of non-silent mutations for a given gene was significantly higher than the background rate, confirmed mutation data for the gene were obtained from whole-exome sequencing. We then estimated the passenger probability for each gene in turn as described by Sjöblom *et al.*[30]. Specifically, the probability (p_{gi}) of obtaining the observed number of mutations for each category (i) in gene g was estimated from a binominal distribution, with b_i being the success probability. The number of available nucleotides for each category was the total number of sufficiently covered (≥8×) bases for that particular category in all 99 tumors. The passenger probability (p_g) for gene g was calculated as the product of the seven category-specific probabilities, where

$$p_g = \prod_{i=1}^{7} p_{gi}$$

We then determined the P value for each gene with the likelihood-ratio test as described by Getz *et al.*[31]. We considered the genes showing significantly ($P < 0.01$) higher mutation rates than the background rate and harboring non-silent mutations in at least 5 of the 99 tumors to be significantly mutated.

Screen of *STAG2* mutations in additional individuals with TCC.

We designed 33 pairs of primers that uniquely amplify exons 3 to 35 of the *STAG2* gene (Supplementary Table 12) using Primer5.0 (Premier Biosoft). PCR was performed on the Dual 96-Well GeneAmp PCR System 9700 using DNA from tumor and matched blood samples for an additional 50 individuals with TCC. Products were sequenced with a 3730xl DNA Analyzer. All sequences were analyzed with Sequencing Analysis Software Version 5.2.

STAG2 promoter methylation analysis by bisulfite sequencing.

Genomic DNA (0.5–2 μg) from tumor and normal bladder tissues from 30 cases was submitted to bisulfite modification using the EpiTect Bisulfite kit (Qiagen) according to the manufacturer's instructions. The methylation status of the putative *STAG2* promoter, which was predicted by Proscan[32], was tested by bisulfite sequencing PCR. Bisulfite PCR primers (Supplementary Table 12) were designed using the online program MethPrimer[33]. Bisulfite-treated DNA (2 μl) was amplified in a 50-μl reaction using HotStarTaq DNA polymerase (Qiagen) with the following cycling conditions: 10 min at 95 °C, 45 cycles of 1 min at 94 °C, 45 s at 61 °C and 1 min at 72 °C, and a final extension step at 72 °C for 10 min. PCR products were purified and ligated into the pMD-18 T vector. Ten white clones for each sample were randomly selected for sequencing to determine the methylation status of each CpG site. BISMA (Bisulfite Sequencing DNA Methylation Analysis) software[34] was used to determine the methylation status for each CpG site and to present the methylation pattern. The log2 ratio of fold change (tumor versus normal) in the methylation level at the *STAG2* promoter for each of the 30 subjects was calculated. The χ^2 test was performed to calculate *P* values and identify differentially methylated samples. *STAG2* promoter with ≥10% methylation, a $\log_2$ ratio of ≥1 and a calculated *P* value of ≤0.001 was defined as hypermethylated.

CNA analysis.

We used SegSeq[35] to infer somatic CNAs in TCC genomes using whole-genome sequencing reads. Resulting copy number segments were mapped to individual genes to determine gene-level copy numbers and copy gain or loss status using thresholds of ≥2.5 copies for gain and ≤1.5 copies for loss. To infer recurrently amplified or deleted genomic regions, we reimplemented the GISTIC algorithm[14], using copy numbers in 100-kb windows instead of SNP array probes as markers. G scores were calculated for genomic regions on the basis of the frequency and amplitude of amplifications or deletions. A significant CNA region was defined as having amplification G score > 0.08 or deletion G score < 0.08, corresponding to a *P*-value threshold of 0.01 from the permutation-derived null distribution[14].

Validation in CNAs by real-time quantitative PCR.

Validation in CNAs was performed by real-time quantitative PCR using SYBR Premix Ex Taq II (TAKARA) on an ABI StepOne instrument (Applied Biosystems) as previously described[36]. CNA primers were designed using Primer3 to amplify a 100- to 200-bp fragment within each CNA using sequences from the UCSC Genome Browser. Five endogenous control primers were also designed using ultraconserved sequences

(Supplementary Table 13). First, a set of gradient dilutions (1:10; 30 ng–0.003 ng) of genomic DNA from a healthy individual was used for RT-PCR amplification to generate a standard curve for each pair of primers. Then, each pair of CNA primers was matched with a pair of endogenous control primers with similar amplification efficiency. Second, each pair of CNA and endogenous control primers was used for RT-PCR amplification for each sample in parallel. ΔC_T was calculated by subtracting the endogenous control C_T value from the CNA C_T value. $\Delta\Delta C_T$ was then determined by subtracting the reference sample ΔC_T value from the test ΔC_T value. The $\log_2$ ratio, as expressed by $\Delta\Delta C_T$, for each CNA was then compared with the $\log_2$ ratio of the copy number changes detected in whole-genome sequencing data.

Validation of *DHFR* amplifications by FISH.

DHFR gene amplifications were further validated by FISH on using formalin-fixed, paraffin-embedded slides of seven TCC cases that were also measured by real-time quantitative PCR as described above. All FISH assays were performed using *DHFR* SpectrumOrange/CEP 5 Spectrum Green oligonucleotide FISH probes (Agilent Technologies)[37] according to the manufacturer's protocol. Briefly, 4-μm paraffin-embedded tissue sections were deparaffinized and digested with proteinase, and antigen retrieval was performed. Sections were then hybridized with the *DHFR* and CEP 5 probes at 37 °C for 18–24 h using an Abbott Molecular Thermobrite. After slides were washed and counterstained with DAPI, signal analysis was performed on an IX71 fluorescence microscope (Olympus). The gene amplification ratio (the gene per chromosome per cell ratio) was evaluated according to the manufacturer's protocol (Agilent Technologies).

Detection and validation of gene fusions.

We used SOAPfuse[38] to detect gene fusion events from RNA-seq data using the default parameters. After mapping RNA-seq reads against the human reference genome sequence (hg18) and Ensembl annotated genes (release 64), SOAPfuse seeks two types of reads to support fusion detection: discordant mapping paired-end reads that connect candidate gene pairs and junction reads that confirm the exact junction sites. SOAPfuse can detect fusion events generated by genome rearrangements with breakpoints in intron and exon regions. Finally, we identified a recurrent gene fusion, *FGFR3-TACC3*, in B59-3 and B100 tumors.

To validate gene fusions on the RNA and DNA levels, PCR primers (Supplementary Table 14) were designed using flanking sequences from both sides of the predicted breakpoints by Primer5.0, and PCR products across the breakpoints ranged from 100 bp to 600 bp in length. PCR was performed on a GeneAmp PCR System 9700 thermal cycler using template cDNA or DNA from both tumors and matched normal controls. Products were gel selected when there were non-specific bands and were sequenced with a 3730xl DNA Analyzer. Sequences were analyzed with Sequencing Analysis Software Version 5.2 and compared by BLAST to the reference genome (hg18) to validate rearrangement events and exact breakpoints.

Pathway enrichment analysis.

We performed pathway enrichment analysis using WebGestalt[39] by examining the distribution of genes with somatic mutations or copy number changes within the KEGG

database. The significance of mutation enrichment was determined by a hypergeometric test and was adjusted for multiple testing with the Benjamini-Hochberg false discovery rate (FDR).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Guangwu Guo[1,2,18], Xiaojuan Sun[1,3,4,5,18], Chao Chen[2,18], Song Wu[1,3,4,5,6,7,18], Peide Huang[2,18], Zesong Li[1,3,4,5,18], Michael Dean[8,18], Yi Huang[1,3,4,5], Wenlong Jia[2], Quan Zhou[2], Aifa Tang[1,3,4,5], Zuoquan Yang[2], Xianxin Li[9], Pengfei Song[2], Xiaokun Zhao[10], Rui Ye[2], Shiqiang Zhang[1,3,4,5], Zhao Lin[2], Mingfu Qi[1,3,4,5], Shengqing Wan[2], Liangfu Xie[1,3,4,5], Fan Fan[2], Michael L Nickerson[8], Xiangjun Zou[1,3,4,5], Xueda Hu[2], Li Xing[1,3,4,5], Zhaojie Lv[1,3,4,5], Hongbin Mei[1,3,4,5], Shengjie Gao[2], Chaozhao Liang[11], Zhibo Gao[2], Jingxiao Lu[9], Yuan Yu[2], Chunxiao Liu[12], Lin Li[2], Xiaodong Fang[2], Zhimao Jiang[9], Jie Yang[2], Cailing Li[9], Xin Zhao[2], Jing Chen[9], Fang Zhang[2], Yongqi Lai[9], Zheguang Lin[9], Fangjian Zhou[7], Hao Chen[1,3,4,5], Hsiao Chang Chan[13], Shirley Tsang[14], Dan Theodorescu[15], Yingrui Li[2], Xiuqing Zhang[2], Jian Wang[2], Huanming Yang[2], Yaoting Gui[9], Jun Wang[2,16,17], Zhiming Cai[1,3,4,5]

Affiliations

[1]Department of Urological Surgery, Shenzhen Second People's Hospital, The First Affiliated Hospital of Shenzhen University, Shenzhen, China.

[2]BGI-Shenzhen, Shenzhen, China.

[3]National-Regional Key Technology Engineering Laboratory for Clinical Application of Cancer Genomics, Shenzhen Second People's Hospital, The First Affiliated Hospital of Shenzhen University, Shenzhen, China.

[4]Shenzhen Engineering Laboratory of Tumor Clinical Immune Gene Therapy, Shenzhen Second People's Hospital, The First Affiliated Hospital of Shenzhen University, Shenzhen, China.

[5]Shenzhen Key Laboratory of Genitourinary Tumor, Shenzhen Second People's Hospital, The First Affiliated Hospital of Shenzhen University, Shenzhen, China.

[6]Institute of Immunology, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China.

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

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

[9]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.

[10]Department of Urology, The Second Xiangya Hospital of Central-Southern University, Changsha, China.

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

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

[13]Epithelial Cell Biology Research Center, School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong, China.

[14]BioMatrix, Rockville, Maryland, USA.

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

[16]Department of Biology, U niversity of Copenhagen, Copenhagen, Denmark.

[17]The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark.

[18]These authors contributed equally to this work.

ACKNOWLEDGMENTS

This work was supported by grants from the Chinese High-Tech (863) Program (2012AA02A201 and 2012AA02A208), the Guangdong Innovative Research Team Program (2009010016), the State Key Development Program for Basic Research of China-973 Program (2011CB809203 and 2014CB745200) and the Shenzhen Municipal Government of China (JC201005260191A, CXB201108250096A, ZDSY20120615154448514 and BGI20100001).

References

1. Jemal A et al. Global cancer statistics. CA Cancer J. Clin. 61, 69–90 (2011). [PubMed: 21296855]
2. Wu XR Urothelial tumorigenesis: a tale of divergent pathways. Nat. Rev. Cancer 5, 713–725 (2005). [PubMed: 16110317]
3. Gui Y et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. Nat. Genet 43, 875–878 (2011). [PubMed: 21822268]
4. Cordon-Cardo C et al. p53 mutations in human bladder cancer: genotypic versus phenotypic patterns. Int. J. Cancer 56, 347–353 (1994). [PubMed: 7906253]
5. Jebar AH et al. *FGFR3* and *Ras* gene mutations are mutually exclusive genetic events in urothelial cell carcinoma. Oncogene 24, 5218–5225 (2005). [PubMed: 15897885]
6. Cappellen D et al. Frequent activating mutations of *FGFR3* in human bladder and cervix carcinomas. Nat. Genet 23, 18–20 (1999). [PubMed: 10471491]
7. Platt FM et al. Spectrum of phosphatidylinositol 3-kinase pathway gene alterations in bladder cancer. Clin. Cancer Res 15, 6008–6017 (2009). [PubMed: 19789314]
8. Cairns P, Proctor AJ & Knowles MA Loss of heterozygosity at the *RB* locus is frequent and correlates with muscle invasion in bladder carcinoma. Oncogene 6, 2305–2309 (1991). [PubMed: 1766677]
9. Hornigold N et al. Mutation of the 9q34 gene *TSC1* in sporadic bladder cancer. Oncogene 18, 2657–2661 (1999). [PubMed: 10353610]
10. Solomon DA et al. Mutational inactivation of *STAG2* causes aneuploidy in human cancer. Science 333, 1039–1043 (2011). [PubMed: 21852505]

11. Welch JS et al. The origin and evolution of mutations in acute myeloid leukemia. Cell 150, 264–278 (2012). [PubMed: 22817890]
12. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult *de novo* acute myeloid leukemia. N. Engl. J. Med 368, 2059–2074 (2013). [PubMed: 23634996]
13. Mhawech-Fauceglia P, Cheney RT & Schwaller J Genetic alterations in urothelial bladder carcinoma: an updated review. Cancer 106, 1205–1216 (2006). [PubMed: 16470587]
14. Beroukhim R et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc. Natl. Acad. Sci. USA 104, 20007–20012 (2007). [PubMed: 18077431]
15. Bertino JR, Goker E, Gorlick R, Li WW & Banerjee D Resistance mechanisms to methotrexate in tumors. Oncologist 1, 223–226 (1996). [PubMed: 10387992]
16. Williamson MP, Elder PA, Shaw ME, Devlin J & Knowles MA *p16* (*CDKN2*) is a major deletion target at 9p21 in bladder cancer. Hum. Mol. Genet 4, 1569–1577 (1995). [PubMed: 8541841]
17. Peset I & Vernos I The TACC proteins: TACC-ling microtubule dynamics and centrosome function. Trends Cell Biol. 18, 379–388 (2008). [PubMed: 18656360]
18. Singh D et al. Transforming fusions of *FGFR* and *TACC* genes in human glioblastoma. Science 337, 1231–1235 (2012). [PubMed: 22837387]
19. Williams SV, Hurst CD & Knowles MA Oncogenic *FGFR3* gene fusions in bladder cancer. Hum. Mol. Genet 22, 795–803 (2013). [PubMed: 23175443]
20. Kiemeney LA et al. A sequence variant at 4p16.3 confers susceptibility to urinary bladder cancer. Nat. Genet. 42, 415–419 (2010). [PubMed: 20348956]
21. Jacobs BL, Lee CT & Montie JE Bladder cancer in 2010: how far have we come? CA Cancer J. Clin 60, 244–272 (2010). [PubMed: 20566675]
22. Kops GJ, Weaver BA & Cleveland DW On the road to cancer: aneuploidy and the mitotic checkpoint. Nat. Rev. Cancer 5, 773–785 (2005). [PubMed: 16195750]
23. Barber TD et al. Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. Proc. Natl. Acad. Sci. USA 105, 3443–3448 (2008). [PubMed: 18299561]
24. Li R et al. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25, 1966–1967 (2009). [PubMed: 19497933]
25. Mortazavi A, Williams BA, McCue K, Schaeffer L & Wold B Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5, 621–628 (2008). [PubMed: 18516045]
26. Li H & Durbin R Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009). [PubMed: 19451168]
27. McKenna A et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010). [PubMed: 20644199]
28. Koboldt DC et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 25, 2283–2285 (2009). [PubMed: 19542151]
29. Guo G et al. Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. Nat. Genet 44, 17–19 (2012).
30. Sjöblom T et al. The consensus coding sequences of human breast and colorectal cancers. Science 314, 268–274 (2006). [PubMed: 16959974]
31. Getz G et al. Comment on “The consensus coding sequences of human breast and colorectal cancers”. Science 317, 1500 (2007).
32. Prestridge DS Predicting Pol II promoter sequences using transcription factor binding sites. J. Mol. Biol 249, 923–932 (1995). [PubMed: 7791218]
33. Li LC & Dahiya R MethPrimer: designing primers for methylation PCRs. Bioinformatics 18, 1427–1431 (2002). [PubMed: 12424112]
34. Rohde C, Zhang Y, Reinhardt R & Jeltsch A BISMA—fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. BMC Bioinformatics 11, 230 (2010). [PubMed: 20459626]
35. Chiang DY et al. High-resolution mapping of copy-number alterations with massively parallel sequencing. Nat. Methods 6, 99–103 (2009). [PubMed: 19043412]

36. Charchar FJ. et al. Whole genome survey of copy number variation in the spontaneously hypertensive rat: relationship to quantitative trait loci, gene expression, and blood pressure. Hypertension 55, 1231–1238 (2010). [PubMed: 20231529]

37. Yamada NA et al. Visualization of fine-scale genomic structure by oligonucleotide-based high-resolution FISH. Cytogenet. Genome Res. 132, 248–254 (2011). [PubMed: 21178330]

38. Jia W et al. SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. Genome Biol. 14, R12 (2013). [PubMed: 23409703]

39. Zhang B, Kirov S & Snoddy J WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res. 33, W741–W748 (2005). [PubMed: 15980575]

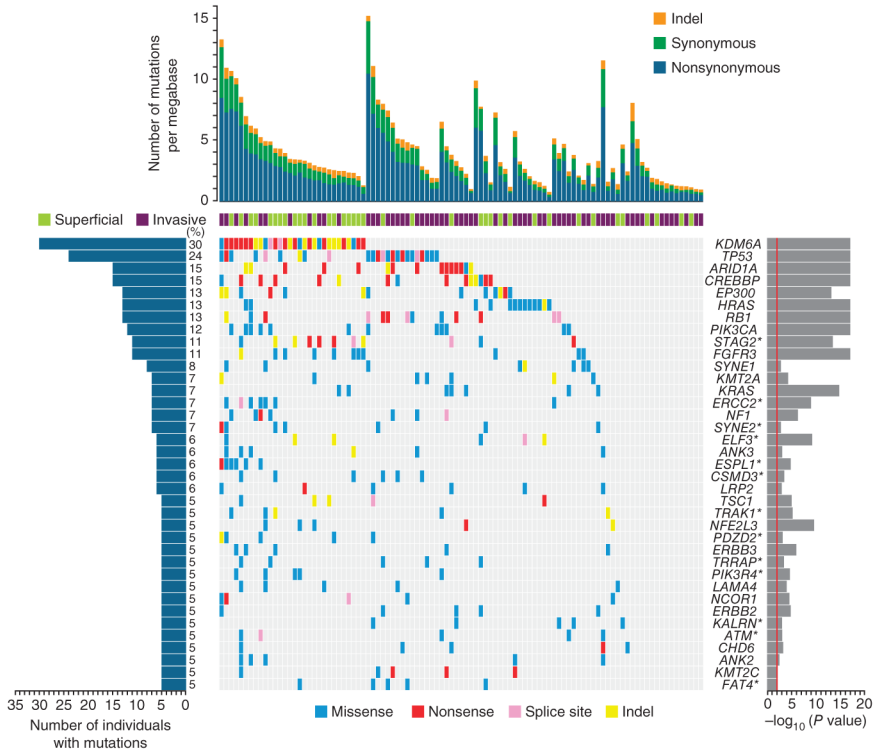

Figure 1.

Significantly mutated genes in TCC as determined by exome sequencing. Significantly mutated genes are listed on the right. The percentage of bladder tumors with mutations detected by automated calling is noted on the left. The upper histogram shows the somatic mutation rate in each of the 99 tumors. The central heatmap shows the distribution of mutations across the sequenced samples. All the mutations shown were confirmed by Sanger sequencing. Genes with asterisks had mutations newly observed in TCC in this study.

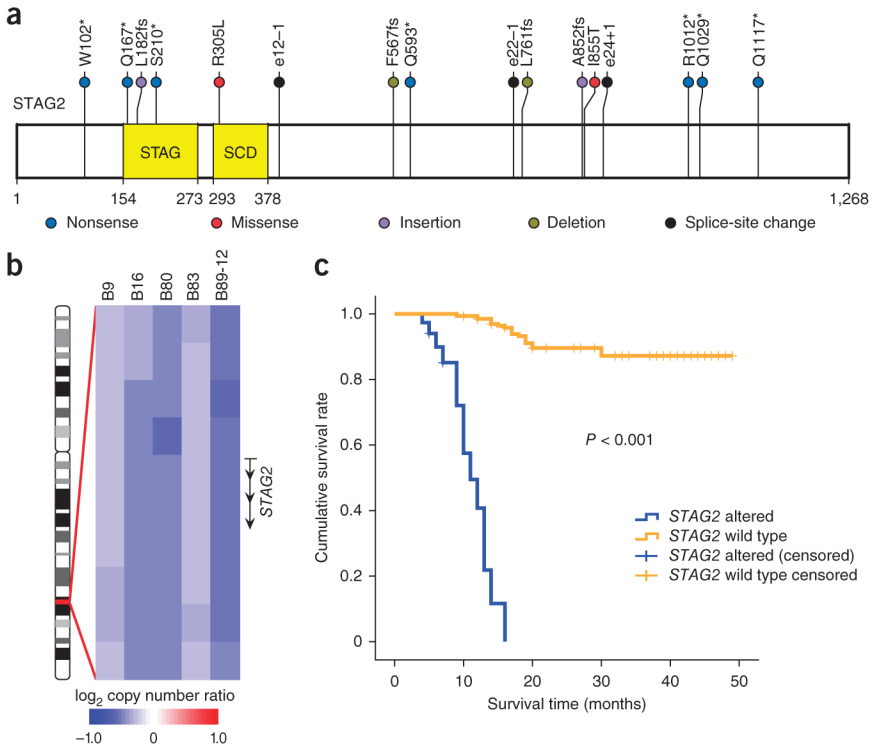

Figure 2.

STAG2 somatic mutations and copy number changes in TCC. (**a**) Somatic alterations overlaid on the STAG2 protein with the conserved protein domains highlighted. STAG, STAG domain; SCD, stromalin conservative domain. (**b**) Five tumors harboring genomic deletions of *STAG2* on the X chromosome (ideogram shown on the left). (**c**) Kaplan-Meier survival analysis of individuals with TCC ($n = 99$) shows that the survival rate for individuals with somatic *STAG2* alterations ($n = 16$) is significantly lower (log-rank test $P < 0.001$) than for individuals with wild-type *STAG2* ($n = 83$).

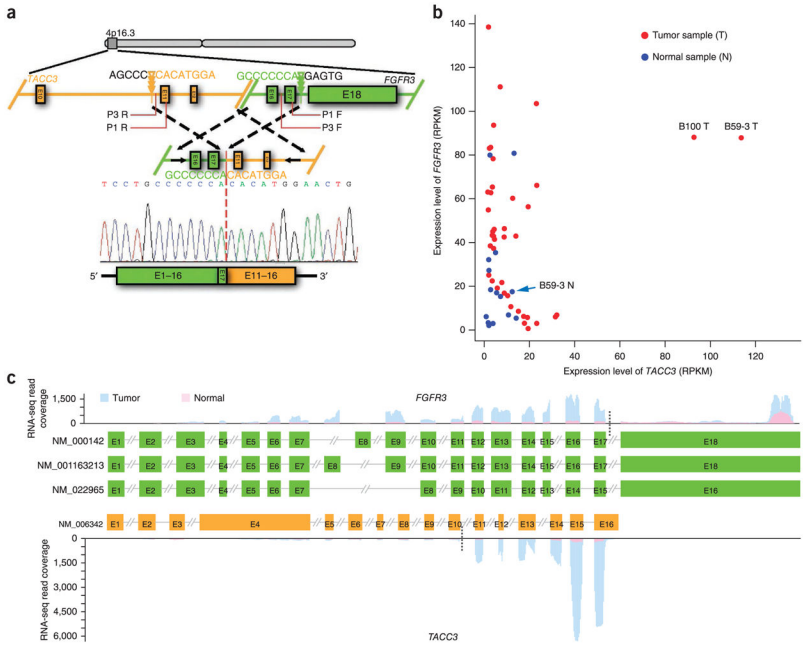

Figure 3.

FGFR3-TACC3 fusion was identified in TCC. (**a**) Genomic fusion of intron 17 of *FGFR3* with intron 10 of *TACC3* resulting in exon 17 of *FGFR3* being spliced 5' to exon 11 of *TACC3* in the fused mRNA. Triangles indicate the genomic positions of the breakpoints. Detailed information on the positions and sequences of primers P1 and P3 is provided in Supplementary Table 14. (**b**) Outlier high expression of *TACC3* in TCCs harboring *FGFR3-TACC3* gene fusions. RPKM, reads per kilobase of exon region in a gene per million mapped reads, (**c**) RNA-seq coverage analysis of *FGFR3* (top) and *TACC3* (bottom) in the tumor and matched normal bladder tissue from B59-3. Three transcripts of *FGFR3* and one transcript of *TACC3* are shown. Black dotted lines indicate breakpoints. E, exon.

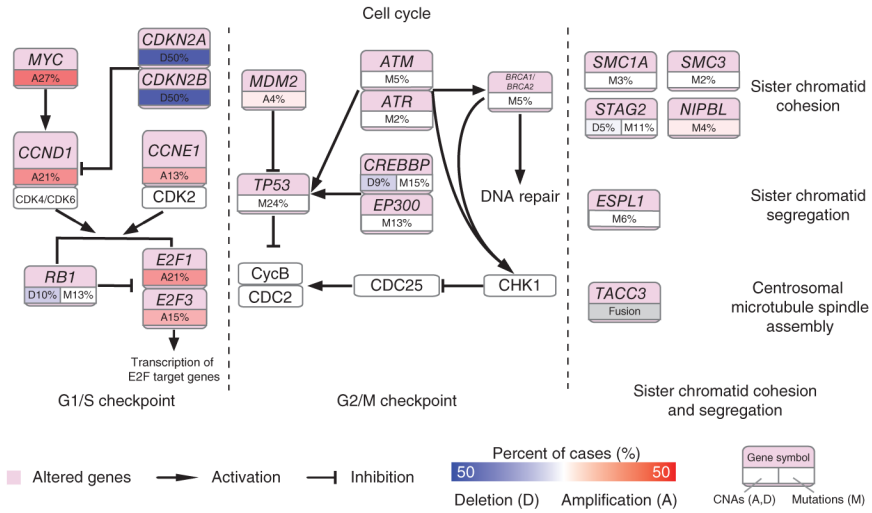

Figure 4.

Frequent genetic alterations in genes from the cell cycle pathway in TCC. Alterations are defined as somatic mutations, focal amplifications and deletions, and, in some cases, as gene fusion events. Alteration frequencies are expressed as a percentage of all cases.