Supplementary Methods

1. Tumor DNA/RNA collection

If tumor cellularity was estimated to be greater than 60% after a thorough pathological review, tumor DNA and RNA were extracted from freshly obtained tumor tissues using a QIAamp Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. For DNA preparation, we used RNaseA (cat. #19101; Qiagen). We measured concentrations and 260/280- and 260/230-nm absorption ratios with an ND1000 spectrophotometer (Nanodrop Technologies, ThermoFisher Scientific, MA) and further quantified DNA/RNA using a Qubit fluorometer (Life Technologies, Carlsbad, CA).

2. PD-L1 immunohistochemistry

PD-L1 expression status was evaluated by sampling representative tumors from each patient. Tissue sections were cut to 4 µm, mounted on Fisherbrand Superfrost Plus Microscope Slides (ThermoFisher), then dried at 60°C for 1 hour. Immunohistochemistry (IHC) staining was carried out on a Dako Autostainer Link 48 system (Agilent Technologies, Santa Clara, CA) using a Dako PD-L1 IHC 22C3 pharmDx kit (Agilent Technologies) with an EnVision FLEX visualization system and counterstained with hematoxylin according to the manufacturer’s instructions. PD-L1 protein expression was determined using a combined positive score (CPS): the number of PD-L1-stained cells (tumor cells, lymphocytes, and macrophages) divided by the total number of viable tumor cells, multiplied by 100. We defined PD-L1-positive bladder cancer (BC) as a tumor with a CPS greater than or equal to 1 (CPS $\geq$ 1).

3. Whole-exome and whole-transcriptome sequencing

Sequencing was performed using genomic DNA (gDNA) from the tumor tissues and matched blood samples using a QIAamp DNA Blood kit (Qiagen). For generation of standard exome capture libraries, we used the Agilent SureSelect Target Enrichment protocol for an Illumina-paired end sequencing library together with 1 µg of inputted gDNA. In all cases, the SureSelect Human All Exon V6 probe set was used. We assessed the quantity and quality of DNA by PicoGreen and agarose gel electrophoresis. We diluted 1 µg of gDNA in EB buffer and sheared to a target peak size of 150-200bp using the Covaris LE220 focused-ultrasonicator (Covaris Inc., Woburn, MA) according to the manufacturer’s recommendations. The fragmented DNA was repaired, and ‘A’ was ligated to the 3′-end. Then, we ligated the

fragments with Agilent adapters and amplified them using PCR. The prepared libraries were quantified using the TapeStation DNA ScreenTape D1000 (Agilent). For exome capture, 250 ng of DNA library was mixed with hybridization buffer, blocking mixes, RNase block, and 5 μg of SureSelect all exon capture library, according to the standard Agilent SureSelect Target Enrichment protocol. Hybridization to the capture baits was conducted at 65℃ using a heated thermal cycler lid option at 105℃ for 24 hours on the PCR machine. The captured DNA was washed and amplified. The final purified product was quantified by qPCR according to the qPCR Quantification Protocol Guide (KAPA Library Quantification kits for Illumina Sequencing platforms) and qualified using the TapeStation DNA ScreenTape D1000 (Agilent). Samples were multiplexed, and flow-cell clusters were created using the TruSeq Rapid Cluster kit and the TruSeq Rapid SBS kit (Illumina, San Diego, CA). Indexed libraries were sequenced using the Illumina HiSeq 2500 platform (Illumina), generating paired-end 100-bp sequence data at Macrogen Inc. (Seoul, Korea).

We estimated total RNA concentration and quality using Quant-IT RiboGreen (Invitrogen). To determine the percentage of fragments with a size greater than 200 bp (DV200), we ran samples on the TapeStation RNA ScreenTape (Agilent). A total of 100ng of total RNA was subjected to sequencing library construction using a TruSeq RNA Access Library Prep Kit (Illumina) according to the manufacturer's instructions. Briefly, total RNA was fragmented into small pieces using divalent cations at elevated temperatures. The cleaved RNA fragments were copied into first-strand cDNA using SuperScript II reverse transcriptase (Invitrogen, #18064014) and random primers, followed by second-strand cDNA synthesis using DNA polymerase I, RNase H, and dUTP. These cDNA fragments were subjected to an end-repair process, addition of a single 'A' base, and ligation of the adapters. The products were purified and enriched using PCR to create the cDNA library. All libraries were normalized, and six libraries were pooled into a single hybridization/capture reaction. Pooled libraries were incubated with a cocktail of biotinylated oligos, corresponding to the coding regions of the genome. We captured targeted library molecules via hybridized biotinylated oligo probes using streptavidin-conjugated beads. After two rounds of hybridization/capture reactions, the enriched library was subjected to a second round of PCR amplification. We quantified and assessed the captured libraries using KAPA Library Quantification Kits for Illumina Sequencing platforms according to the qPCR Quantification Protocol Guide (KAPA BIOSYSTEMS, #KK4854) and the TapeStation D1000 ScreenTape (#5067-5582, Agilent Technologies) recommendations. Indexed libraries were submitted to an Illumina HiSeq2500

(Illumina), and paired-end (2×100 bp) sequencing was performed by Macrogen Inc (South Korea).

4. Variant calling and filtering of the whole-genome sequencing data

Sequenced reads were mapped to the human reference genome (GRCh37) using the BWA-MEM algorithm [1]. The duplicated reads were removed by Picard (available at http://broadinstitute.github.io/picard), and indel realignment and base quality score recalibration were performed by GATK [2]. To establish the highly sensitive somatic variant sets, we initially considered the unions of variant calls from Mutect and Varscan2 [3,4] Variants called by both of the tools were included for future analysis. We evaluated the numbers of variant and wild-type reads in the tumor and in matched normal blood samples for every variant. To establish highly-confident somatic variant sets, we applied an additional filtering process for these variants. Briefly, we removed variants with the following features: the presence of variant reads in a matched-normal sample (≥ 2 for single-nucleotide variations [SNVs], ≥ 1 for short indels), five or more mismatched bases in the variant read, low depth of coverage (≤ 10), less than four supporting reads, average variant position in supporting reads within the first or last 10% of the read length, five or more bases in a flanking homopolymer matching one allele, a large difference in average mapping quality between reference and variant reads (≥ 35), or a large difference in average trimmed read length between reference and variant reads (≥ 25). Finally, the sequencing and mapping artifacts of the remaining variants were filtered out using a locus-specific background error matrix generated by in-house normal tissue whole-exome sequencing (WES) datasets (panel of normal).

5. Mutational signature analyses

We estimated contributions of mutational signatures to an observed mutational spectrum in each sample (i.e., the presumed amount of exposure to corresponding mutational processes). We solved the following constrained optimization problem [5]:

$$\underset{\mathbf{h}}{\operatorname{argmin}} \, ||\mathbf{v} - W\mathbf{h}||_2^2$$

where $\mathbf{v} \in \mathbb{R}_{0+}^{m\times 1}$, $W \in \mathbb{R}_{0+}^{m\times k}$, and $\mathbf{h} \in \mathbb{R}_{0+}^{k\times 1}$ (m is the number of mutation types and k is

the number of mutational signatures). For each sample, given the observed counts of each mutation type v from a sample and the pre-trained mutational signature matrix W, we calculated exposure h. We used an R package (pracma) that internally uses an active-set method to solve the above problem. The relative contributions of mutational signatures were calculated by refitting seven consensus mutational signatures previously identified in bladder cancers (COSMIC signatures 1, 2, 5, 8, 13, 29, and 40; available at https://cancer.sanger.ac.uk/cosmic/signatures) [6].

6. Copy number variations

Segmented copy number profiles were estimated for the whole-exome-sequenced samples by Sequenza algorithms using matched blood samples as controls [7]. Subclonal copy number changes, of which depth ratios were not properly fitted to the integer values of the absolute copy numbers, were excluded manually. Significant focal copy number alterations were identified from the segmented data using the Genomic Identification of Significant Targets in Cancer (GISTIC) algorithm [8].

7. Exploration of the immune microenvironment in RNA-sequencing datasets

To investigate the immune microenvironment of our BC samples, we analyzed RNA-sequencing (RNA-seq) datasets from 64 BC patients. Reads were aligned by the STAR algorithm with a two-pass protocol [9]. We quantified gene expression using RSEM in units of TPM (transcript per million) [10], according to the RNA-seq pipeline suggested by the ENCODE project (https://www.encodeproject.org/pipelines/). To estimate gene set enrichment variation against the samples of an expression data set, we performed Gene Set Variant Analysis (GSVA) with the GSVA package in R [11,12]. Enrichment of canonical pathways in comparison of two groups was estimated by the Gene Set Enrichment Analysis (GSEA) algorithm implemented in the fgsea R package.

References

1. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754-60.

2. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297-303.

3. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31:213-9.

4. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22:568-76.

5. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500:415-21.

6. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578:94-101.

7. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. Ann Oncol. 2015;26:64-70.

8. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011;12:R41.

9. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15-21.

10. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.

11. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.

12. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545-50.