Supplementary Methods

1. Extraction of genomic DNA from formalin-fixed, paraffin-embedded tissues

Ten tissue slides (5 μm) of resected specimens and 20 slides of small biopsy samples were used for genomic DNA (gDNA) extraction. Tissue gDNA was extracted from formalin-fixed, paraffin-embedded (FFPE) tissue using the QIAamp sDNA FFPE Tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and eluted in a 50 μL volume. DNA yield was evaluated using a Nanodrop 8000 UV-Vis spectrometer (NanoDrop Technologies Inc., Wilmington, DE) and Qubit 4.0 fluorometer (Thermo Fisher Scientific, Waltham, MA). The DNA quantity and quality were examined using a 4200 TapeStation instrument (Agilent Technologies, Santa Clara, CA). Samples with a DNA yield > 200 ng and a median DNA fragment size of at least 350 bp were selected for targeted sequencing.

2. Targeted sequencing and bioinformatics

Targeted sequencing was performed using the K-MASTER cancer panel v1.1, which includes whole exomes of 409 cancer-related genes and intronic regions of 23 genes. gDNA was sheared using the Covaris S220 instrument (Covaris, Woburn, MA, USA). Target capture was performed using the Sure-Select XT Reagent Kit HSQ (Agilent Technologies, Santa Clara, CA), and a barcoded paired-end sequencing library was constructed. Sequencing was performed on a HiSeq 2500 instrument and 100-bp reads were generated (Illumina, San Diego, CA). One hundred samples were sequenced in a single experiment using the "high-throughput" mode, and 17 samples were sequenced using the "fast" mode.

Paired-end reads were aligned to the human reference genome (hg19) using BWA-MEM v0.7.5. Samtools v0.1.18, GATK v3.1-1, and Picard v1.93 were used for bam file handling, local realignment, and the removal of duplicate reads, respectively. Samples with a mean target coverage of < 200× were excluded from further analysis. Single-nucleotide variants with a variant allele fraction > 1% were detected using MuTect v1.1.4 and Lowfreq v0.6.1. Sequencing errors were filtered with an in-house algorithm using the data extracted from each BAM file. Insertions and deletions (indels) < 30 bp were detected using Pindel v0.2.5a4. Possible germline polymorphisms were also filtered if the allele frequency was >0.1% in any of the normal population databases, including the 1000 Genomes Project database; the Exome Aggregation Consortium (ExAC) database; the National Heart, Lung, and Blood Exome Sequencing Project database; the Korean Reference Genome Database; and the Korean Variant Archive database. Structural variants (SVs) and large indels (> 30 bp) were detected using an in-house SV caller. Copy number alterations of each gene were detected using an in-house copy number caller with copy numbers > 6 marked as amplifications and copy numbers < 0.7 designated as deletions.

3. Cell-free DNA isolation

Cell-free DNA (cfDNA) was extracted from plasma using the QIAamp Circulating Nucleic Acid kit (Qiagen, Hilden, Germany) with the QIAvac 24 Plus vacuum manifold according to the manufacturer's instructions. cfDNA was quantified using the Qubit 4.0 (Life Technologies, Burlington, Canada). cfDNA purity was analyzed using an Agilent High-Sensitivity DNA kit and a 4200 TapeStation instrument (Agilent Technologies, Santa Clara, CA). When required, additional purification was performed using the Agencourt AMPure XP kit (Beckman Coulter, Brea, CA) to remove contaminating nucleic acids. cfDNA concentration was quantified with a Qubit 4.0 fluorometer (Thermo Fisher Scientific) using an Agilent High Sensitivity DNA kit (Agilent Technologies, Santa Clara, CA).

4. Axen Cancer Panel 1

Targeted sequencing was performed using ACP1, which includes the exomes of 88 cancer-related genes and the intronic regions of three genes. *Hybridization-based* enrichment libraries were prepared manually following the manufacturer's protocol (SureSelectXT HS and XT Low input Target Enrichment System for the Illumina paired-end sequencing library). In summary, 20 ng of cfDNA was amplified with individual index and molecular barcodes and hybridized with capture oligos during library preparation. The captured sequences were then enriched with streptavidin-conjugated paramagnetic beads and further amplified before being subjected to Illumina sequencing. The fragment sizes of all libraries were measured using a 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA), and qPCR was performed using the LightCycler 480 System (Roche, Carlsbad, CA) with the Kapa Library Quantification Kit (KK4854, KAPA Biosystems, Wilmington, DE). Sequencing was performed using the Illumina NextSeq500 platform with an average read length of 2×150 bp and a total depth of 5,000×, per the manufacturer's instructions.

Adaptor sequences and low-quality bases were removed from the raw sequencing reads using the Agilent Genomics Toolkit (AGeNT, Agilent Software). Trimmed reads were aligned to the human reference genome (hg19) using Burrows–Wheeler Aligner-MEM (BWA-MEM) [23]. Duplicate reads were marked according to molecular barcode information using LocatIt (Agilent Software). Poorly mapped reads were removed using Sambamba [24], and base quality score recalibration was performed using the Genome Analysis Tool Kit (GATK) [25]. Somatic mutations were detected using the MuTect2 algorithm [2] and variants were annotated using SnpEff and SnpSift [26]. To reduce the effect of false-positive variants, we applied additional filtration criteria: (1) variants with a minor allele frequency > 5% in the genome aggregation database (gnomAD)* and ExAC database were excluded [7]; (2) variants with mutated read counts < 5 were excluded; and (3) variants with total read depth < 200 were excluded.