Supplementary Methods

1. Whole-exome sequencing for tumor tissue

For the generation of standard exome capture libraries, we used the Agilent SureSelect Target Enrichment protocol (Santa Clara, CA) for Illumina paired-end sequencing library (ver. C2, December 2018, San Diego, CA) together with 1 μg input gDNA. In all cases, the SureSelect Human All Exon V6 probe set was used. The quantification of DNA and the DNA quality is measured by PicoGreen and agarose gel electrophoresis. We used 200ng of DNA diluted in EB Buffer and sheared to a target peak size of 150-200 bp using the Covaris LE220 focused-ultrasonicator (Covaris, Woburn, MA) according to the manufacturer's recommendations. Load the 8 microTUBE Strip into the tube holder of the ultrasonicator and shear the DNA using the following settings: mode, frequency sweeping; duty cycle, 10%; intensity, 5; cycles per burst, 200; duration, 60 seconds×6 cycles; temperature, 4℃-7℃. The fragmented DNA is repaired, an 'A' is ligated to the 3′ end, agilent adapters are then ligated to the fragments. Once ligation had been assessed, the adapter ligated product is polymerase chain reaction (PCR) amplified. For exome capture, 250 ng of DNA library was mixed with hybridization buffers, blocking mixes, RNase block and 5 μL 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 heated thermal cycler lid option at 105℃ for 24 hours on PCR machine. The captured DNA was then washing and amplified. The final purified product is then quantified using quantitative polymerase chain reaction (qPCR) according to the qPCR Quantification Protocol Guide Guide (KAPA Library Quantification kits for Illumina Sequencing platforms) and qualified using the TapeStation DNA screentape D1000 (Agilent). And then we sequenced using the HiSeq 2500 platform (Illumina).

2. Whole exome sequencing data analysis

Fastq files generated by whole exome sequencing (WES) were mapped on human genome (hg19) through Burrows-Wheeler Aligner (BWA) (ver. 0.7.12-r1039) with BWA-MEM algorithm [1]. The aligned bam files were sorted by coordinates using SAMtools (v0.1.19) [2]. Then, Genome Analysis Toolkit (GATK, v3.6 and v.4.13) performed duplicate marking, indel realignment, and base recalibration for coordinate-sorted hg19-aligned reads [3]. For somatic mutation detection, MuTect2 from GATKv4.13 was used by comparing BAM files of tumor and those of matched normal samples. Possible germline events were filtered by gnomAD [4] (population allele fraction < 2.5e-6), and variant effect predictors [5] annotated the remained events. For SMC03 sample, which lacks matched normal control, MuTect2 was performed with tumor-only mode, followed by possible germline event elimination. For further analysis, called mutations with mutant reads ≤ 4 were filtered out. For the hypermutated tumor sample, mutational signature (COSMIC v2) analysis was performed by R

package, deconstructSigs (v1.8.0) [6].

3. WES-based copy number variation analysis

Python package, ngCGH was used to detect WES-based copy number ratios of tumor samples to their corresponding normal samples with window size=1,000 sequencing reads. We defined copy number gain and amplification if the copy ratio is greater than 1.5 and 2, respectively. On the contrary, if the genes showed the copy ratio less than 0.75 and 0.5, the gene were classified as copy number loss and deletion, respectively. To confirm whether this classification of gene copy number variations (CNVs) was reliable or not, we investigated the association between gene CNVs and mRNA expression. As a result, *CDK4/FRS2* gene expression levels of *CDK4/FRS2*-amplified tumors were significantly higher than those of tumors with no *CDK4/FRS2* gain or amplification (Wilcoxon rank sum $p=1.5\times10^{-6}$ and 0.0018, respectively) (Fig. 4C, S14 Fig.). HDLBP loss tumors showed significant down-regulation of HDLBP gene expression compared to tumors with no HDLBP loss/deletion (Wilcoxon rank sum $p=1.1\times10^{-6}$) although the HDLBP-deleted tumor displayed comparable gene expression levels with no HDLBP loss/deletion tumors.

Also, accurate *CDK4* and *MDM2* co-amplification status is critical since their co-amplification is used for the molecular diagnosis of STS, and therefore, we re-investigated their amplification status with consideration of tumor purity and ploidy. we adopted ABSOLUTE [7] algorithm to calculate tumor purity, ploidy, and absolute copy number values from ngCGH results (S13 Table) except for four samples (SMC7, SMC12, YCC8, and YCC17) as ABSOLTUE failed to model their data. Based on tumor ploidy and total copy number values from ABSOLUTE, we classified CNV as well. If tumor ploidy >= 4, CNVs were defined as amplification and gain when copy number > ploidy+4 and > ploidy+2, respectively. If copy number < ploidy - 4 or ploidy*0.5, CNV was defined as deletion, and if copy number < ploidy - 2 or ploidy*0.75, CNV was defined as loss. For tumors with ploidy less than 4, CNV was defined as amplification, gain, loss, and deletion if copy number > ploidy+2, copy number > ploidy+1, copy number < 1.5, and copy number < 1, respectively (S14 Fig.). Although there are some inconsistencies between log2R-based classification and total copy number-based classification, log2R-based CDK4/MDM2 amplification and total copy number-based CDK4/MDM2 amplification were observed in exactly same tumors, and all of them showed moderate-level amplifications (11-14 copies). Since the *HDLBP*-deleted tumor SMC16 (log2R classification) showed comparable gene expression level with no *HDLBP* loss/deletion, we investigated the total copy number of HDLBP in this sample and confirmed that SMC16 was classified into deletion according to total copy number-based classification as well (HDLBP total copy number of SMC16 was 0).

4. RNA-sequencing

Total RNA concentration was estimated by Quant-IT RiboGreen (Invitrogen, Carlsbad, CA). To determine the DV200 (% of RNA fragments > 200 bp) value, samples were run on the TapeStation RNA ScreenTape (Agilent). Overall, 100 ng of total RNA was subjected to sequencing library construction using a TruSeq RNA Access library prep kit (Illumina) according to the manufacturer's protocol. Briefly, the total RNA was first fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first strand cDNA using SuperScript II reverse transcriptase (#18064014, Invitrogen) and random primers. This was 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 subsequently, ligation of the adapters. The products are then purified and enriched with PCR to create the cDNA library. All libraries were normalized and six were pooled into a single hybridization/capture reaction. Pooled libraries were incubated with a cocktail of biotinylated oligos, corresponding to coding regions of the genome. Targeted library molecules were captured via hybridized biotinylated oligo probes using streptavidin-conjugated beads. After two rounds of hybridization/capture reactions, the enriched library molecules were subjected to a second round of PCR amplification. The captured libraries were quantified using KAPA Library Quantification kits for Illumina Sequencing platforms according to the qPCR Quantification Protocol Guide (#KK4854, KAPA BIOSYSTEMS), and assessed using the TapeStation D1000 ScreenTape (# 5067-5582, Agilent Technologies). Indexed libraries were then submitted to an Illumina HiSeq2500 (Illumina), and paired-end (2×100 bp) sequencing was performed by Macrogen Incorporated (Seoul, Korea).

5. RNA-sequencing data analysis

Sequence reads from RNA-sequencing were aligned on hg19 by STAR_(v2.6.1d) [8]. For gene expression profiling, Cufflinks (v2.2.1) quantified the aligned reads in Fragments Per Kilobase Million (FPKM) [9]. For further analysis, log2-transformed FPKM values were used except for DEG analysis. To extract DEGs between responders and non-responders, normalized reads counts per gene were obtained with DEGseq [10] (R package), and then the resulting reads counts were applied to DEseq2 [11] (R package) to obtain DEGs. To calculate the Microenvironment Cell Populations-counter scores in tumor samples, MCPcounter (R package) was used [12]. Geneset enrichment analysis (GSEA) was performed by GSEA-P [13]. GSEA results were visualized on Cytoscape (v3.7.1) Enrichment Map [14]. Single sample GSEA (ssGSEA) algorithm in GSVA [15] (R package) was applied to estimate geneset activation score in tumor samples for HU_ANGIOGENESIS_UP MSigDB geneset [16] and Responder and Non-Responder DEGs. Gene fusion detection was performed by STAR-Fusion (v1.5.0) with STAR-aligned bam files [17].

References

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

2. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078-9.

3. 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.

4. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. Nature 2020;581:434-43.

5. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl variant effect predictor. Genome Biol. 2016;17:122.

6. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:31.

7. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol. 2012;30:413-21.

8. 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.

9. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7:562-78.

10. Wang L, Feng Z, Wang X, Wang X, Zhang X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics. 2010;26:136-8.

11. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

12. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016;17:218.

13. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. Bioinformatics. 2007;23:3251-3.

14. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One. 2010;5:e13984.

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

16. Hu J, Bianchi F, Ferguson M, Cesario A, Margaritora S, Granone P, et al. Gene expression

signature for angiogenic and nonangiogenic non-small-cell lung cancer. Oncogene. 2005;24:1212-9.
17. Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. Genome Biol. 2019;20:213.