Appendix 1 More details of Materials and Methods

Bioinformatics analysis of public data

TCGA-LUAD data download: For WES data retrieval, the TCGA data portal (https://portal.gdc.cancer.gov/) was accessed to identify LUAD cases. WES data with corresponding mutation annotation format (MAF) files were downloaded and processed through the TCGAbiolinks package in Bioconductor. Quality control measures included exclusion of germline-related variants and retention of mutations with population frequencies <1% in public databases to minimize background noise. For DNA methylation data, IDAT-formatted files containing methylation profiles were acquired and analyzed using the minfi package to derive β-values representing methylation levels. Quality filtering was implemented to remove probes and samples with excessive missing values, followed by normalization procedures. For RNA-seq data, gene expression quantification files (including raw counts and transcripts per million [TPM] values) were obtained and preprocessed through TCGAbiolinks. Normalization was performed using z-score transformation, with subsequent elimination of low-expression or missing genes.

Multi-omics integration via Multi-Omics VI Clustering and Subtyping (MOVICS): During preprocessing, sample alignment across omics datasets was ensured through unique patient identifiers, with quality-controlled filtering retaining 410 samples that were present in all three omics datasets. For WES data, a binary mutation status matrix of a 276-gene DDR panel was extracted using the maftools R package. DNA methylation levels were quantified using β-values, while RNA-seq expression was normalized to TPM values. To reduce dimensionality and eliminate redundancy, hierarchical feature optimization was implemented: methylation probes were aggregated by gene annotation using dplyr, with sample-wise mean β-values calculated per gene; protein-coding genes in RNA-seq data were retained, represented by their maximum TPM values per gene.

Genomic instability quantification: TMB (mutations/Mb) was calculated from MAF files using the tmb() function in maftools. MSI scores, aneuploidy scores, and fraction of genome altered were directly extracted from TCGA datasets. Boxplots generated via ggpubr were used to visualize TMB and MSI score differences between multi-omics subtypes (CS1 *vs.* CS2), with statistical significance assessed by Wilcoxon rank-sum test (significance threshold: $P<0.05$).

Survival analysis, LASSO regression, and stepwise Cox modeling: To mitigate multicollinearity among variables identified in univariate analysis, LASSO regression was implemented via the glmnet package, where the optimal regularization parameter ($\lambda=0.11$) was determined through 10-fold cross-validation (cv.glmnet). Stepwise multivariate Cox regression using the My.stepwise algorithm (based on coxph) refined the model to identify genes that were independently associated with survival. This iterative process revealed EXO1 as the sole gene exhibiting independent prognostic significance. A Cox proportional hazards model incorporating EXO1 expression was constructed, with forest plots and survival curves generated using survival and survvminer packages.

Gene Set Enrichment Analysis (GSEA): To identify signaling pathways associated with EXO1 expression, we performed GSEA using the TCGA-LUAD transcriptomic data. Genes were ranked based on the $\log_2$ fold change between EXO1-high and EXO1-low groups, which were defined by median EXO1 expression. The preranked gene list was then analyzed with the GSEA software using the Molecular Signatures Database (MSigDB) hallmark gene set collection. Parameters were set as follows: number of permutations = 1000, enrichment statistic = weighted, and minimum/maximum gene set size = 15/500. Gene sets with a false discovery rate (FDR) <0.25 were considered significantly enriched. To resolve ties in the ranked list, genes with identical $\log_2$ fold change values were further ordered by gene symbol.

Plasmid construction

PLKO.1 vector was used to construct an shRNA expression plasmid. Double digestion of the plasmid was performed using AgeI and EcoRI restriction enzymes, followed by agarose gel electrophoresis and purification of the linearized vector. Pre-designed shRNA oligonucleotides targeting specific genes (Sigma-Aldrich) were annealed to create double-stranded DNA fragments, which were ligated into the linearized PLKO.1 backbone using T4 DNA ligase (Thermo Scientific). The ligation products were transformed into DH5α-competent *Escherichia coli*, plated on LB agar containing ampicillin (100 μg/mL), and incubated at 37 ℃ overnight. Single colonies were expanded for plasmid extraction, with insert sequences validated by Sanger sequencing. High-purity plasmids were prepared using Qiagen Maxi Prep kits for downstream applications. For RFP-expressing plasmids lacking puromycin resistance, 293T cells with high RFP expression were enriched by fluorescence-

 https://dx.doi.org/10.21037/tlcr-2025-1-1357

activated cell sorting (FACS) and validated by RT-qPCR. The shRNA sequences (5'-3') were as follows:

mEXO1-SH1:TCAAGCCGATTCTCATATTTC
mEXO1-SH2:GATTAGTACTAAAGGGTTAAA
hEXO1-SH1:AATGCAGACTGCTGCAAAGCT
hEXO1-SH2:GCCGTGTTCAAAGAGCAATAT
mArid1a-SH1:CAGGCCCTATGGCCCTAATAT
mArid1a-SH2:CTTTATAGTATGGCGAGTTAA
mMSH2-SH1:CGAGATCATTTCACGGATAAA
mMSH2-SH2:GCAACGAAGATTGGTGCCTTT

For construction of the EXO1 full-length expression plasmid, total RNA from murine tumor tissues was reverse-transcribed using Vazyme Reverse Transcriptase. EXO1 exon sequences were amplified using KOD PLUS high-fidelity DNA polymerase (Toyobo) with gene-specific primers, purified by gel electrophoresis, and cloned into linearized PLVX vectors using the ClonExpress MultiS One Step Cloning Kit (Vazyme). Transformed DH5α colonies were screened on ampicillin-containing LB agar, and plasmid sequences were confirmed by Sanger sequencing. High-purity plasmids were prepared using Qiagen Mini Prep kits for subsequent experiments.

Lentiviral packaging and stable cell line generation

Lentiviral particles were produced using a three-plasmid system (pLVX/pLKO.1:psPAX2:pMD2.G=5:2:1) cotransfected into 293T cells using Easytrans transfection reagent (Yeasen Biotechnology). Cells were seeded in 6-cm dishes and transfected at 70–80% confluency. The post-transfection medium was replaced after 8 h. Viral supernatants were harvested at 24, 48, and 72 h, filtered through 0.45-μm membranes, and concentrated by PEG8000 precipitation (8% final concentration) at 4 ℃ overnight. Viral pellets were collected by centrifugation (4000×*g*, 20 min) and resuspended in DMEM. For stable cell line generation, LLC, LLC-OVA, or B16-F10 cells were seeded in 12-well plates (5×10^5 cells/well) and transduced with viral supernatants containing 8 μg/mL polybrene. Puromycin selection (2 μg/mL) was initiated 48 h post-infection until complete death of untransfected controls. Surviving cells were expanded, and knockdown/overexpression efficiency was confirmed by RT-qPCR.

Western blot

SDS-PAGE gels (8%, 10%, or 12% resolving; 5% stacking) were prepared using 30% acrylamide, Tris-HCl buffers (pH 8.8/6.8), SDS, APS, and TEMED. Total protein was extracted using RIPA lysis buffer (Beyotime) supplemented with protease/phosphatase inhibitors and 1 mM EDTA. Lysates were cleared by centrifugation (12,000 rpm, 15 min, 4 ℃), quantified by BCA assay, and denatured in 2× Laemmli buffer (10 min, 95 ℃). Electrophoresis proceeded at 80 V (stacking gel) and 120 V (resolving gel) until bromophenol blue migration. Proteins were transferred to nitrocellulose membranes (300 mA, 2 h), blocked with 5% BSA, and probed with primary antibodies (4 ℃ overnight) followed by HRP-conjugated secondary antibodies (2 h, room temperature). Signal detection employed ECL reagent (Yeasen) with image acquisition using the eBlot system and densitometric analysis with ImageJ.

Detailed animal methodology

All experimental procedures were conducted in strict accordance with ARRIVE Essential 10 guidelines. Female C57BL/6 mice (6 weeks old, derived from the same litter to minimize genetic variability) were utilized throughout this study. Single mouse was treated as an analysis unit. Using a random number table for randomization, mice were assigned to the following experimental groups: (1) Experiment 1: EXO1-knockdown (EXO1-KD) LLC subcutaneous tumors *vs.* PLKO.1-empty vector LLC control; (2) Experiment 2: EXO1-KD B16 subcutaneous tumors *vs.* PLKO.1-empty vector B16 control; (3) Experiment 3: Group A (EXO1-KD LLC tumors), Group B (EXO1-KD LLC + anti-PD1 therapy), Group C (PLKO.1-empty vector LLC + anti-PD1), and Control (PLKO.1-empty vector LLC); (4) Experiment 4: Group A (ARID1A-KD

https://dx.doi.org/10.21037/tlcr-2025-1-1357

LLC tumors), Group B (MSH2-KD LLC tumors), and Control (PLKO.1-empty vector LLC). Sample sizes of 5–7 mice per group were determined based on established practices in oncology literature, with technical replicates included for selected KD cell lines. The cages were arranged in a Latin square design to balance environmental interference. Blinding was not implemented during procedures or assessment, as personnel were aware of tumor cell lines per cage and treatment assignments. The primary endpoint was tumor volume, with all tumor-bearing mice surviving to predefined endpoints (Day 14 or 21 post-injection) and subsequently euthanized via cervical dislocation under anesthesia (as described in main methods). No animals were excluded from analysis. Statistical comparisons utilized Student's t-test (two groups) or one-way ANOVA with Tukey's post hoc test (≥3 groups) in GraphPad Prism v9.0 ($p<0.05$ significance threshold), with normality testing applied to continuous data.

Moreover, conclusions should be interpreted with caution due to exclusive use of C57BL/6 mice without validation in alternative models.

Humane endpoints and management in tumor-bearing mice

In accordance with approved animal welfare protocols, predefined humane endpoints were established for this study. Mice were immediately removed from the study and humanely euthanized if any of the following criteria were met: the tumor burden exceeded 1.0 cm in any single diameter, the tumor became ulcerated or necrotic, or the tumor significantly interfered with normal mobility, feeding, or behavior.

Throughout the course of this investigation, no animals reached these predefined humane endpoint criteria requiring early intervention. All procedures were conducted in strict compliance with institutional guidelines for the ethical care and use of laboratory animals.

Neoantigen Prediction Using pVACtools

Somatic variants identified from whole-exome sequencing (WES) of mouse tumor samples were filtered to retain only high-confidence nonsynonymous mutations. The filtered variants were annotated using Ensembl VEP against the GRCm39 reference genome to obtain gene symbols and amino acid changes. Neoantigen prediction was performed using the pVACseq pipeline. For each sample, mutant peptide sequences of 8–11 amino acids spanning the mutation site were generated. Binding affinity to the mouse MHC-I alleles (H-2-Db and H-2-Kb, corresponding to the C57BL/6 background) and MHC-II alleles (H2-IAb) was predicted using NetMHCpan integrated into the pVACseq workflow. Clonality of the underlying mutations was inferred from the variant allele frequency (VAF) extracted from the original VCF files, with mutations exhibiting VAF > 0.25 designated as clonal. The number and quality (binding strength, clonality) of predicted neoantigens were compared between experimental groups.

HR/NHEJ reporter assays (pDR-GFP and EJ5-GFP)

To functionally assess the impact of Exo1 knockdown on DNA double-strand break (DSB) repair pathway choice, we employed the well-established DR-GFP (homologous recombination, HR) and EJ5-GFP (non-homologous end joining, NHEJ) reporter systems. Stable LLC cell lines expressing either Exo1-shRNA or a scrambled control were first generated by lentiviral transduction followed by RFP-based flow cytometry sorting; knockdown efficiency was verified by qPCR. Cells were then transfected with 2 μg of pDR-GFP or EJ5-GFP plasmid using Lipofectamine 3000 according to the manufacturer's instructions. After 24 h, puromycin ([1 μg/mL]) was added to select for stable integrants over a period of [3 days]. To induce a site-specific DSB, cells were transiently transfected with an I-SceI expression plasmid. After an additional 48 h, the percentage of GFP-positive cells was quantified by flow cytometry.

Microsatellite Instability (MSI) Analysis by PCR and Capillary Electrophoresis

Genomic DNA was extracted from LLC cells (control, Exo1-sh, and Msh2-sh) using the DNeasy Blood & Tissue Kit (Qiagen).

https://dx.doi.org/10.21037/tlcr-2025-1-1357

A panel of murine microsatellite markers was selected, including D3Mit22, D3Mit278, D13Mit3 and DxMit172. PCR amplification was carried out in a 15 μL reaction volume containing 1 μL DNA, 0.3 μL Taq Green Master Mix (Promega), and 0.15 μL of each primer (forward primers were 5′-end labeled with FAM or HEX fluorophores). Thermal cycling conditions were: initial denaturation at 94 ℃ for 3 min; 35 cycles of 94 ℃ for 15 s, [55 ℃] for 15 s, and 72 ℃ for 30 s; followed by a final extension at 72 ℃ for 3 min. PCR products were diluted 1:10 in Hi-Di formamide and mixed with GeneScan 500 LIZ size standard. Fragment analysis was performed on an Applied Biosystems 3730xl DNA Analyzer, and electropherograms were visualized using GeneMapper software. Shifts in fragment length compared to control samples were considered indicative of microsatellite instability.

Primers of D3Mit22:
AAGGATTGAAGAATGGTTGGG
AATCAGCGATTTCAGCACG
Primers of D3Mit278:
AACTACCATCTAAAACATCCTCTGTG
AGATCCCTAGAGAAACAGAACTGG
Primers of D13Mit3:
TCAGGCTCATCCCAGATACC
TTTTGCAGAGAACACACACC
Primers of DxMit172:
TACCACAGTTTGAATAAAGATGTGTG
GAAGAAACCATGACTCCTCTTTG

https://dx.doi.org/10.21037/tlcr-2025-1-1357

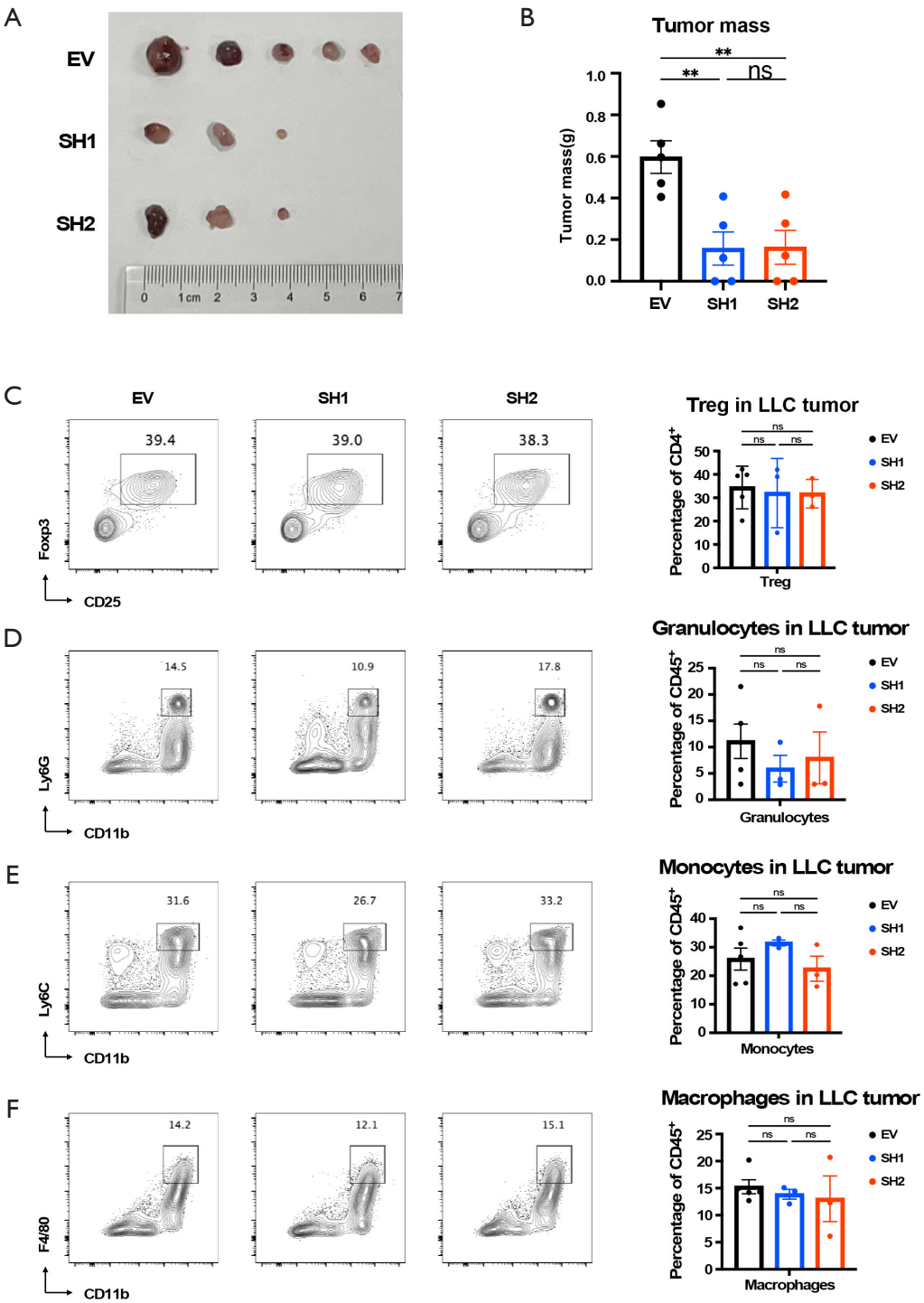
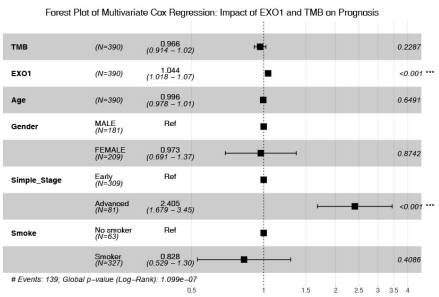


Figure S1 Reduced *Exo1* expression does not affect regulatory T cells (Tregs) and myeloid immune-cell infiltration in subcutaneous tumors. (A) Tumors in C57 mice induced by subcutaneous injection of *Exo1*-knockdown and control LLC cell lines (Repeated experiment). (B) Mass of subcutaneous tumors in *Exo1*-knockdown and control group (Repeated experiment). (C) Proportion of Treg infiltration in *Exo1*-knockdown and control subcutaneous LLC tumors. (D) Proportion of granulocyte infiltration in *Exo1*-knockdown and control subcutaneous LLC tumors. (E) Proportion of monocyte infiltration in *Exo1*-knockdown and control subcutaneous LLC tumors. (F) Proportion of macrophage infiltration in *Exo1*-knockdown and control subcutaneous LLC tumors.

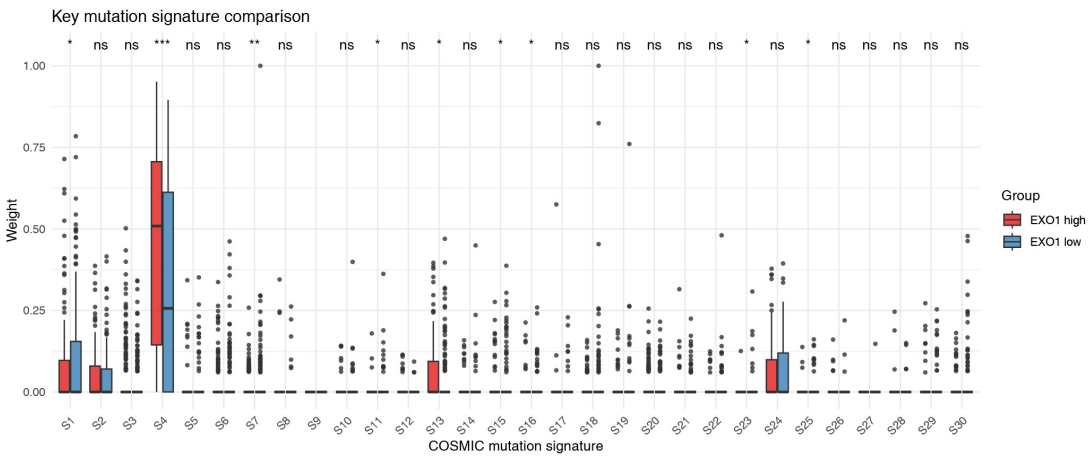
https://dx.doi.org/10.21037/tlcr-2025-1-1357

A

Variable	Coefficient	SE	HR	95% CI	z	p
TMB	-0.03425	0.02846	0.9663	0.9139–1.022	-1.204	0.229
EXO1	0.04268	0.01280	1.0436	1.0178–1.070	3.335	<0.001 ***
Age	-0.00428	0.00940	0.9957	0.9776–1.014	-0.455	0.649
Gender (female vs male)	-0.02758	0.17427	0.9728	0.6913–1.369	-0.158	0.874
Stage (advanced vs early)	0.87759	0.18339	2.4051	1.6789–3.445	4.785	<0.001 ***
Smoking (smoker vs non-smoker)	-0.18882	0.22852	0.8279	0.5290–1.296	-0.826	0.409
TMB × EXO1	0.0005584	0.001688	1.0006	0.9973–1.004	0.331	0.741



B



C

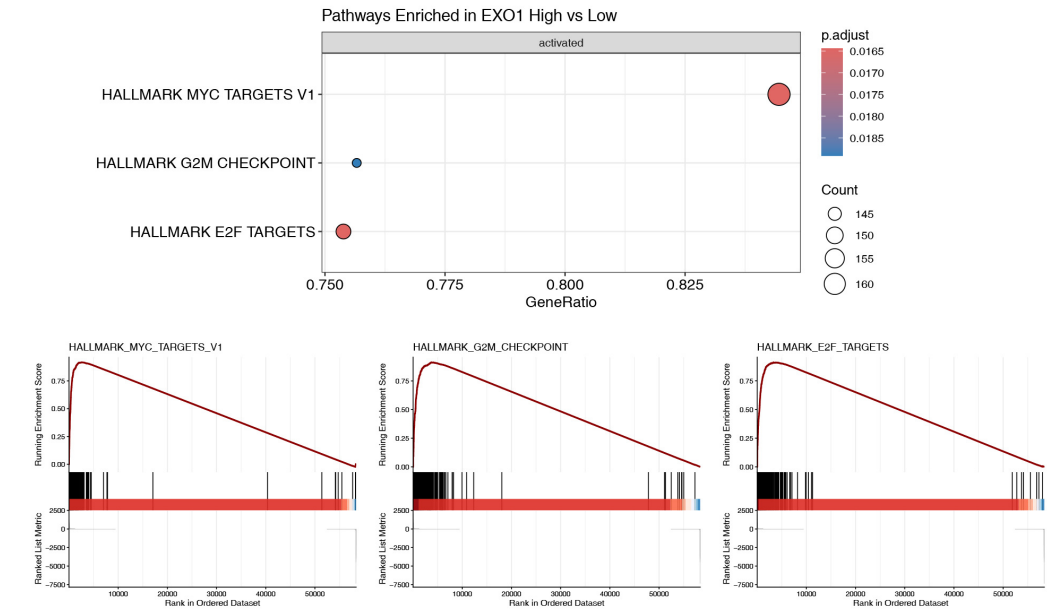


Figure S2 TCGA-LUAD cohort analysis reveals distinct EXO1-associated genomic and transcriptomic features. (A) Forest plot summarizing multivariable Cox regression analysis of overall survival in TCGA-LUAD patients. The model included EXO1 expression, tumor mutational burden (TMB), their interaction term, and clinical covariates (age, gender, stage, smoking status). Hazard ratios (HR) with 95% confidence intervals are shown. (B) Mutational signature analysis of whole-exome sequencing data from TCGA-LUAD tumors stratified by EXO1 expression (high *vs.* low, based on median expression). The bar plot displays the relative contribution of COSMIC mutational signatures in each group. (C) Gene set enrichment analysis (GSEA) of TCGA-LUAD transcriptomic data comparing EXO1-high and EXO1-low tumors. The dot plot shows significantly enriched hallmark gene sets (FDR < 0.25) in the EXO1-high group.

https://dx.doi.org/10.21037/tlcr-2025-1-1357

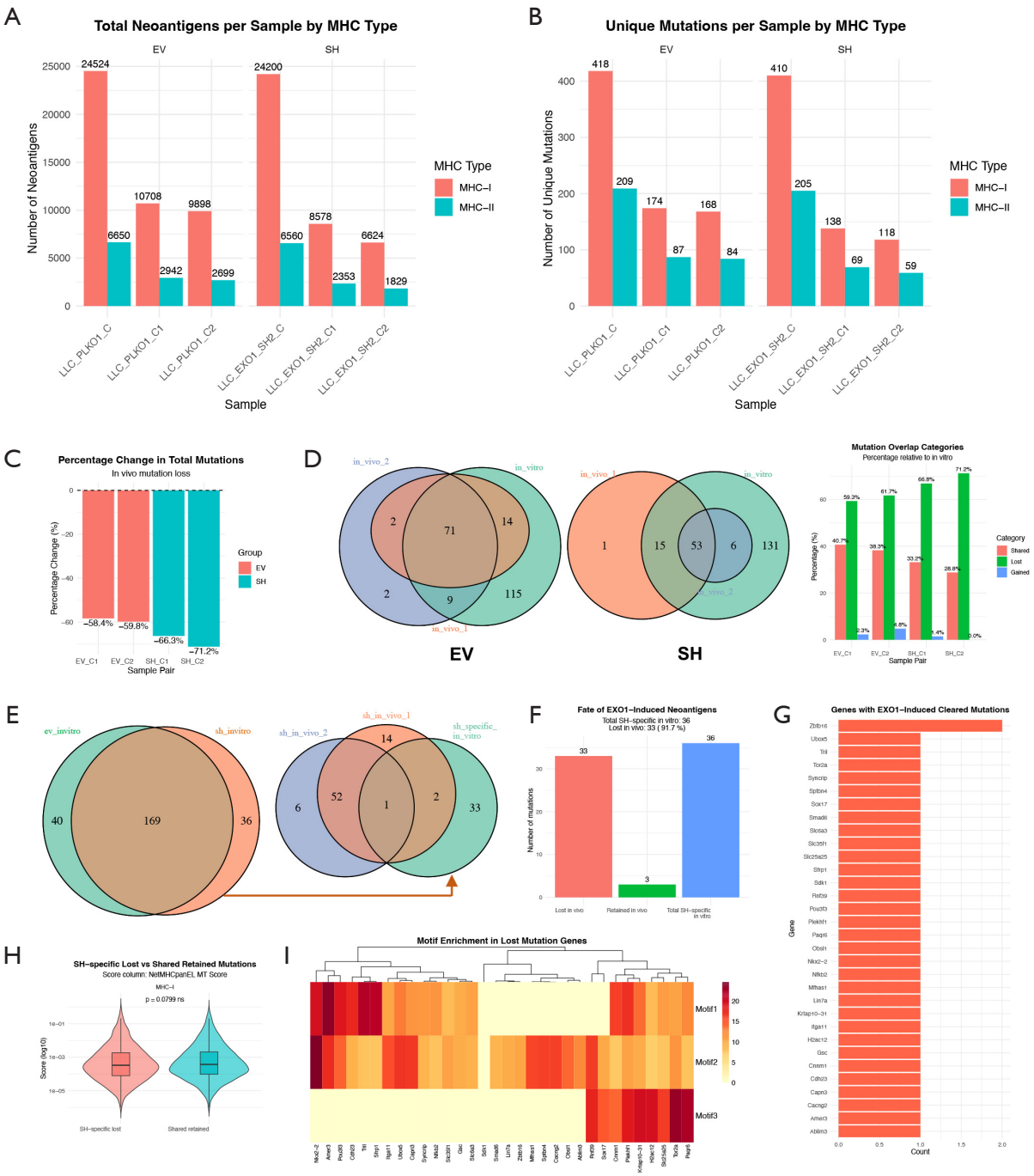


Figure S3 Neoantigen dynamics and characteristics in Exo1-knockdown models. (A) Total number of predicted neoantigens in each experimental group. (B) Number of unique mutations identified per group. (C) Number of neoantigens lost in *in vivo* samples relative to their matched *in vitro* controls. (D) Venn diagrams depicting the overlap of neoantigens between *in vivo* samples and their corresponding *in vitro* counterparts, with the percentage change in neoantigen numbers indicated. (E) Venn diagrams showing neoantigens newly acquired in Exo1-knockdown *in vitro* cells compared to control *in vitro* cells, and their intersection with neoantigens detected in Exo1-knockdown *in vivo* tumors. (F) Classification of the intersecting neoantigens from (E) by mutation type. (G) Gene names of the intersecting neoantigens from (E). (H) Comparison of MHC binding affinity between neoantigens that were eliminated *in vivo* and those that persisted. (I) MEME-derived motif scores for the intersecting neoantigens.

https://dx.doi.org/10.21037/tlcr-2025-1-1357

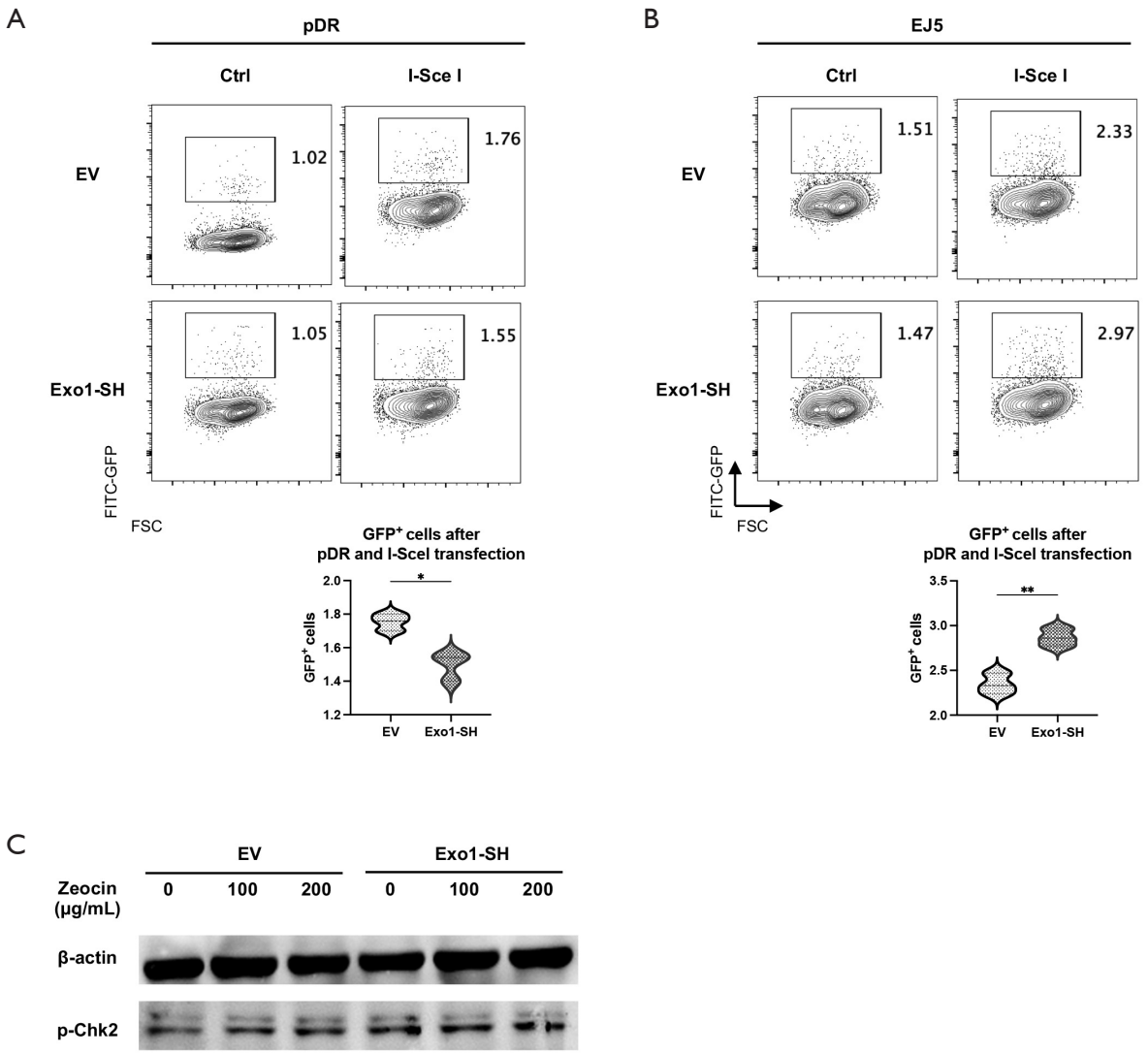


Figure S4 Assessment of DNA repair pathway efficiency using reporter assays. (A) Quantification of GFP-positive cells in Exo1-knockdown versus control LLC cells after transfection with the I-SceI expression plasmid and the pDR-GFP reporter, which measures homologous recombination (HR) activity. (B) Quantification of GFP-positive cells in Exo1-knockdown versus control LLC cells after transfection with the I-SceI expression plasmid and the EJ5-GFP reporter, which measures non-homologous end joining (NHEJ) activity. (C) Western blot analysis of p-Chk2 expression in control and Exo1-sh LLC cells after 24-hour treatment with the indicated concentrations of Zeocin.

https://dx.doi.org/10.21037/tlcr-2025-1-1357

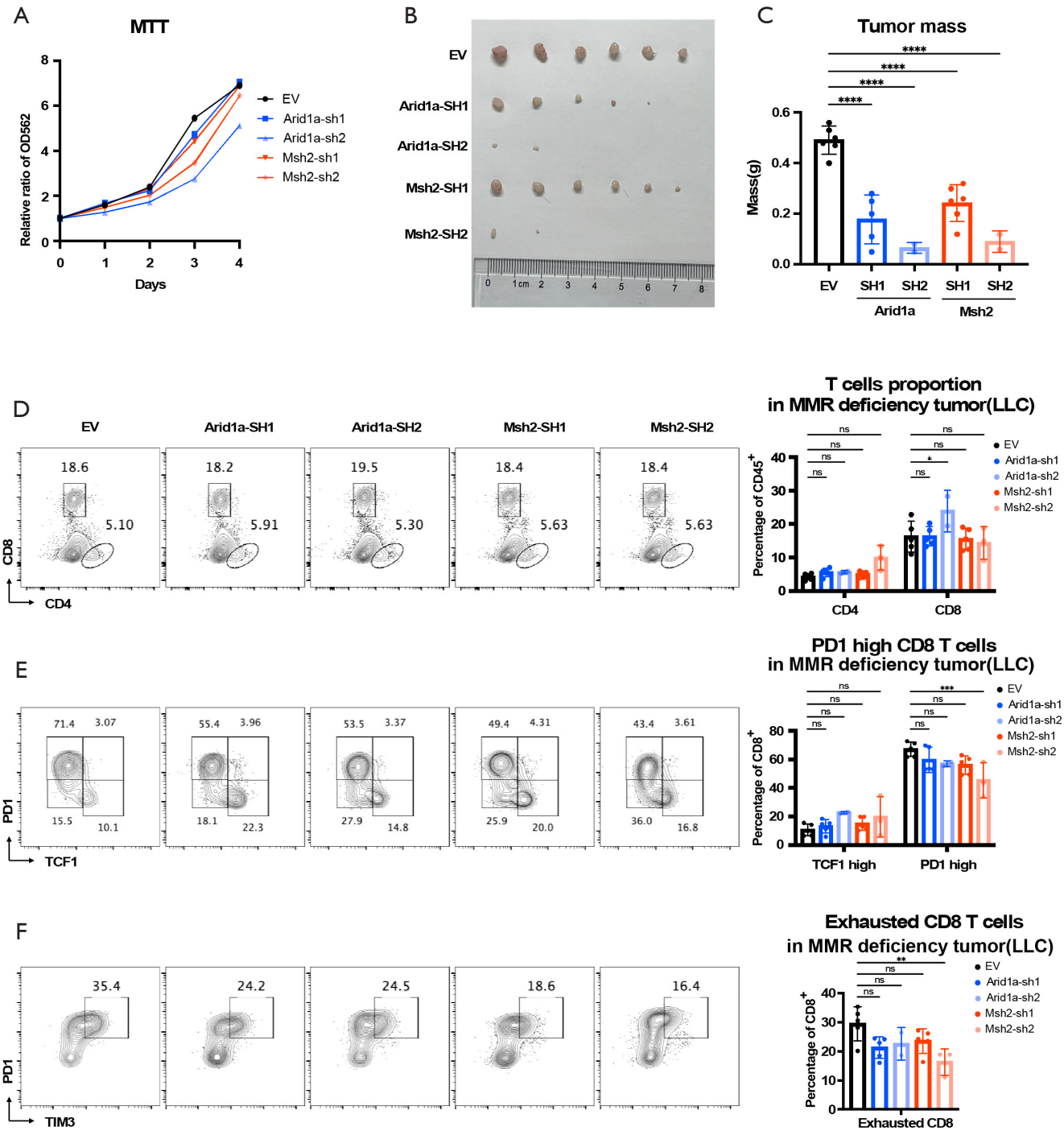


Figure S5 Reduced *Exo1* expression does not cause tumor phenotypes *in vivo* by affecting the mismatch repair pathway. (A) MTT assay shows the growth curves of *Arid1a*-knockdown, *Msh2*-knockdown, and control LLC cells *in vitro*. (B) Tumors were induced in C57 mice by subcutaneous injection of *Arid1a*-knockdown, *Msh2*-knockdown, and control LLC cell lines. (C) Mass of subcutaneous tumors in *Arid1a*-knockdown, *Msh2*-knockdown, and control groups. (D) Proportion of $CD4^+$ and $CD8^+$ T-cell infiltration in *Arid1a*-knockdown, *Msh2*-knockdown, and control subcutaneous LLC tumors. (E) PD-1 and TCF-1 expression in $CD8^+$ T cells in *Arid1a*-knockdown, *Msh2*-knockdown, and control subcutaneous LLC tumors. (F) Proportion of exhausted $CD8^+$ T-cell infiltration in *Arid1a*-knockdown, *Msh2*-knockdown, and control subcutaneous LLC tumors.

https://dx.doi.org/10.21037/tlcr-2025-1-1357

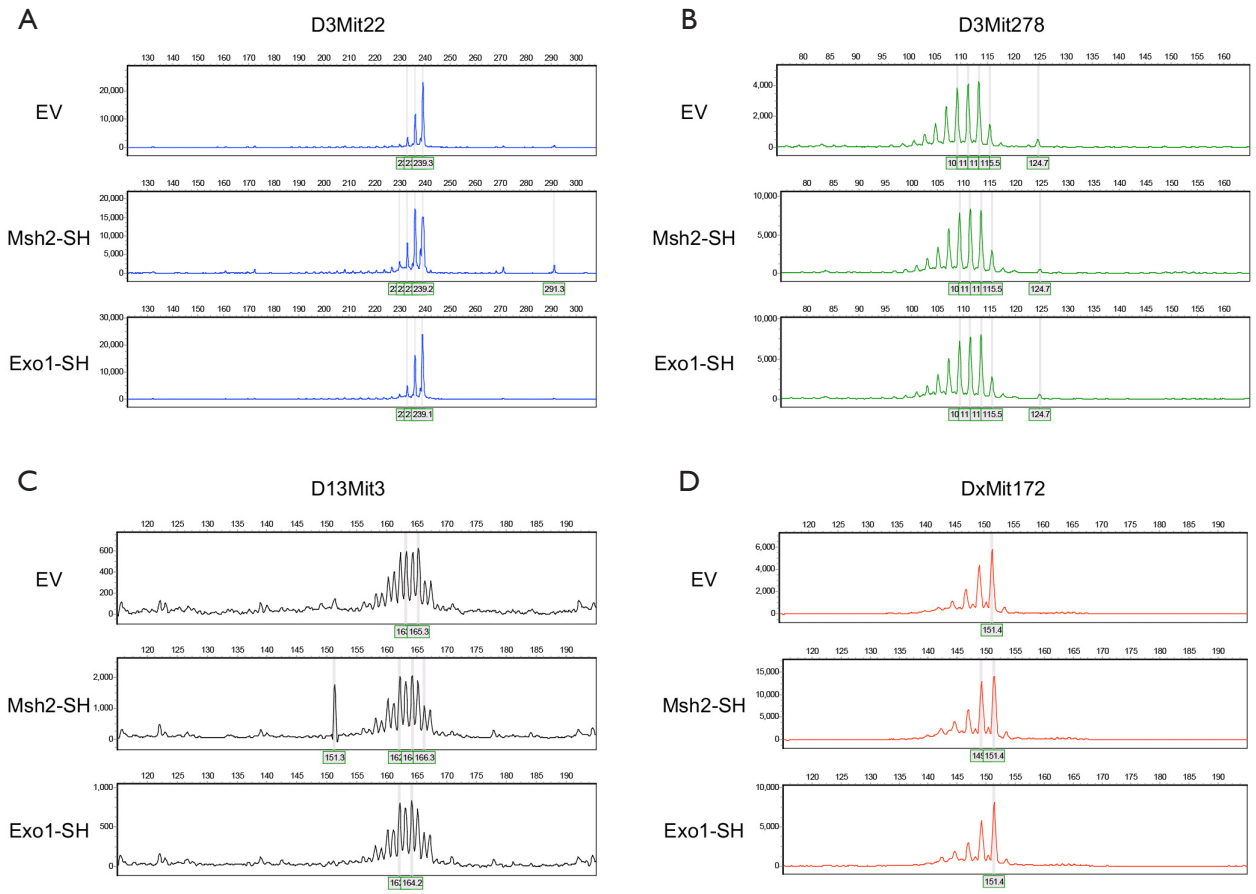


Figure S6 Microsatellite instability analysis by capillary electrophoresis. Representative electropherograms of PCR-amplified microsatellite markers from LLC cells stably expressing control shRNA (Control), Msh2 shRNA (Msh2-sh), or Exo1 shRNA (Exo1-sh). The markers analyzed are: (A) D3Mit22, (B) D3Mit278, (C) D13Mit3, and (D) DxMit172. Fluorescence intensity is plotted against fragment size (base pairs).

https://dx.doi.org/10.21037/tlcr-2025-1-1357