

SUPPLEMENTARY INFORMATION

Analysis of the Coding Transcriptome in NSCLC Highlights Variant-Specific Gene Expression and Signaling in CD74-ROS1 Fusions

Jasmine Vargas¹, Julia Olivieri^{2,*}, and Georgios Pantouris^{1*}

¹Department of Chemistry, University of the Pacific, Stockton, CA 95211, USA

²Department of Computer Science, University of the Pacific, Stockton, CA 95211, USA

*Correspondence: jolivieri@pacific.edu; gpantouris@pacific.edu

Supplementary Table S1: Differential Gene Expression Analysis Across the Four Groups*

	A549	CD74-ROS1	CD74^{Δ2-41}-ROS1	CD74-ROS1^{K439M}
A549	-	62	549	877
CD74-ROS1	62	-	433	696
CD74^{Δ2-41}-ROS1	549	433	-	509
CD74-ROS1^{K439M}	877	696	509	-

*The total number of genes in the dataset is 23,342

Supplementary Table S2: Equivalent Gene Expression Analysis Across the Four Groups*

	A549	CD74-ROS1	CD74^{Δ2-41}-ROS1	CD74-ROS1^{K439M}
A549	-	10082	8118	8559
CD74-ROS1	10082	-	6996	8174
CD74^{Δ2-41}-ROS1	8118	6996	-	7335
CD74-ROS1^{K439M}	8559	8174	7335	-

*The total number of genes in the dataset is 23,342

Supplementary Table S3: Differentially expressed genes (n=62) identified between A549 and CD74-ROS1

CES1	HDAC9
LOC107987423	NPY1R
PRICKLE1	IL1RL1
CEACAM6	PTPRD
NTS	FGF12
IL6	SUCNR1
AR	C5orf38
PKIB	SPINK6
CDH19	SMIM31
LUM	DLGAP1
F2RL2	UGT2B7
FAAH2	DOC2B
IGFL2-AS1	UNC13D
SCEL	MMP16
SVEP1	C4BPB
PPARGC1A	DOK7
TSPAN7	GOLGA8S
ADAM28	ST6GALNAC1
CES1P1	LINC01895
VCAN	CRISPLD1
SLC16A7	KCNJ16
LAMA4	KRT20
FN1	NOTCH3
SULT1E1	DHRS2
NOS1	LHFPL4
LOC105372424	NRAD1
RASSF9	PRR15
TRPC6	LHX8
HAGLR	ZEB2
CEACAM5	GRIK2
IGSF10	ARL14EPL

Supplementary Table S4. DNA oligos used for mutagenesis*.

Variant Generated	Primer	Primer Sequence (5' to 3')
CD74 ^{Δ2-41} -ROS1	Forward (PCR reaction 1)	AAGCTGGCTAGCGCCACCGAGAGCAAATGC
	Reverse (PCR reaction 1)	GCATTTGCTCTCGGTGGCGCTAGCCAGCTT
	Forward (PCR reaction 2**)	GCTAGCGCCACCATGAGCAAATGCAGC
	Reverse (PCR reaction 2**)	GCTGCATTTGCTCATGGTGGCGCTAGC
	Forward	AAGGTGGCCGTTATGACCCTGAAGAAG
	Reverse	CTTCTTCAGGGTCATAACGGCCACCTT
CD74-ROS1 ^{K439M}	Forward	AAGGTGGCCGTTATGACCCTGAAGAAG
	Reverse	CTTCTTCAGGGTCATAACGGCCACCTT

* CD74-ROS1 C6-R32 was used as a template for PCR with the exception of **

** The confirmed plasmid of CD74^{Δ2-41}-ROS1 (Post-PCR reaction 1 product) was used as a template for PCR reaction 2

Supplementary Table S5. DNA oligos used for Sanger sequencing*

Variant	Primer	Primer Sequence (5' to 3')
CD74-ROS1	Reverse	GTGCATCCAGCTCTCGAACA
	Forward	CACCATGCACCGGAGAAGAA
	Forward	CGTGCCTAACAAGCCCGGCA
	Forward	TAGAACTGATGGAAGGCGGC
	Reverse	GGGTTTAAACGGGCCCCTAGT
	Forward	TAGAACTGATGGAAGGCGGC
CD74 ^{A2-41} -ROS1	Reverse	GTGCATCCAGCTCTCGAACA
	Forward	ATGCAGAACGCCACCAAGTA
	Forward	ACATCATACTCGTGGGCGAC
	Forward	TGACACTGGGCCACCAGCCT
CD74-ROS1 ^{K439M}	Reverse	GTGCATCCAGCTCTCGAACA
	Forward	CACCATGCACCGGAGAAGAA
	Forward	CGTGCCTAACAAGCCCGGCA
	Reverse	GGGTTTAAACGGGCCCCTAGT
	Forward	TGACACTGGGCCACCAGCCT

* Due to the nucleic acid length of CD74-ROS1 variants (>2000 bp) and the reading limitations of Sanger sequencing (<1000 bp), multiple primers were employed to confirm the entire nucleic acid sequence of the fusion proteins.

ATGCACCGGAGAAGAAGCCGATCTTGCAGAGAGGATCAGAAACCAGTGATGGACGACCAGCGG
GACCTCATTAGCAACAACGAACAGCTGCCCATGCTGGGCAGACGGCCTGGCGCTCCAGAGAGCA
AATGCAGCCGGGGCGCCCTGTACACCGGCTTCAGCATCCTGGTTACCCTGCTGCTGGCTGGCCAG
GCCACCACTGCCTATTTCTTTATCAGCAACAGGGCAGACTGGACAAACTGACAGTGACCAGCC
AGAACCTGCAGCTGGAAAACCTGCGGATGAAACTGCCCAAACCTCCTAAACCTGTCAGCAAGAT
GAGAATGGCCACCCCTCTGCTGATGCAGGCCCTGCCTATGGGCGCCCTGCCTCAGGGACCTATGC
AGAACGCCACCAAGTACGGCAACATGACCGAGGACCACGTGATGCACCTGCTGCAAAAACGCCG
ACCCCTGAAGGTGTACCCTCCTCTGAAGGGCAGCTTTCTGAGAATCTGAGGCACCTGAAGAA
CACCATGGAACCATCGACTGGAAAGTGTTGAGAGCTGGATGCACCACTGGCTGCTATTCGAG
ATGAGCAGACATTCTCTGGAACAGAAGCCAACCGACGCTCCACCCAAGGCCGGCGTGCCTAACA
AGCCCGGCATCCCAAGCTGTTGGAGGGCAGCAAGAACAGCATCCAGTGGGAGAAGGCCGAGG
ACAACGGCTGTAGAATCACATACTACATCCTGGAAATCCGGAAGAGCACCAGCAACAATCTGCA
GAACCAGAACCTGAGATGGAAGATGACATTCAACGGCTCCTGTAGTAGCGTGTGCACCTGAAAA
AGCAAGAACCTGAAGGGCATCTTCCAGTTCAGAGTGGTGGCTGCTAACAACCTGGGATTTCGGCG
AATACAGCGGCATCAGCGAGAACATCATACTCGTGGGCGACGATTTCTGGATCCCTGAAACCAG
CTTTATCCTAACCATCATCGTGGGAATTTTCTGGTGGTGACAATCCCTCTGACCTTTGTATGGCA
CCGGCGCTTAAAGAATCAGAAGTCCGCCAAGGAAGGCGTGACCGTGCTGATCAACGAGGACAA
GGAAGTGGCAGAGCTGAGAGGCCTGGCCGCTGGCGTCGGCCTGGCTAACGCCTGCTACGCCATC
CACACCCTGCCACCCAAGAAGAGATCGAGAATCTGCCTGCCTTTCCACGGGAGAAGTTGACAC
TGAGGCTGCTTCTGGGCTCCGGCGCATTCCGGCGAGGTGTACGAGGGTACAGCCGTGGATATCCTG
GGAGTCGGGAGCGGCGAGATCAAGGTGGCCGTTAAGACCCTGAAGAAGGGCTCTACCGATCAG
GAGAAGATCGAGTTCCTGAAAGAGGCCACCTGATGAGTAAGTTCAACCACCCAATATCCTGA
AGCAGCTGGGAGTTTGTCTGCTGAACGAGCCCCAGTACATCATCCTAGAACTGATGGAAGGCGG
CGACCTGCTGACCTACCTGCGGAAGGCCAGAATGGCCACATTCTACGGACCTCTGCTGACACTG
GTGGATCTTGTGGACCTCTGTGTGGACATCTCCAAGGGCTGCGTGTACCTGGAAAGAATGCACTT
CATCCATAGAGATCTTGCTGCCAGAACTGCCTGGTGTCCGTTAAGGACTACACATCTCCTAGAA
TCGTGAAAATCGGAGATTTCCGCCCTGGCCAGAGATATCTACAAGAACGACTACTACAGAAAGCG
GGGAGAAGGCCTGCTCCCTGTGCGGTGGATGGCCCCTGAGAGCCTGATGGATGGCATCTTCACC
ACCCAGAGCGACGTGTGGTCCTTCGGCATCCTGATCTGGGAAATCCTGACACTGGGCCACCAGC
CTTACCCCGCCCACTCTAATCTCGACGTGCTGAATTACGTGCAGACCGGAGGAAGACTGGAACC
ACCTAGAAATTGCCCTGACGATCTGTGGAACCTGATGACCCAGTGCTGGGCTCAGGAGCCTGATC
AAAGACCTACCTTCCACAGAATTCAAGACCAGCTGCAACTGTTTAGAACTTCTTCTGAAACAGC
ATCTACAAGTCCAGAGACGAGGCCAACAATAGCGGCGTGATCAACGAGTCTTTGAGGGGGGAG
GACGGCGATGTGATCTGCCTGAACTCTGATGACATCATGCCTGTGGCCCTGATGGAAACAAAGA
ACCGGGAAGGCCTGAACTACATGGTGTGGCCACAGAGTGCGGCCAGGGAGAGGAAAAAAGCG
AAGGCCCTCTGGGCTCTCAGGAGAGCGAGTCATGCGGCCTGCGGAAGGAAGAAAAGGAGCCCC
ACGCCGATAAGGACTTTTGCCAGGAAAAACAGGTGGCCTACTGCCCCAGCGGCAAGCCTGAAGG
CCTGAACTATGCCTGTCTGACCCACAGCGGCTACGGCGACGGCAGCGAC

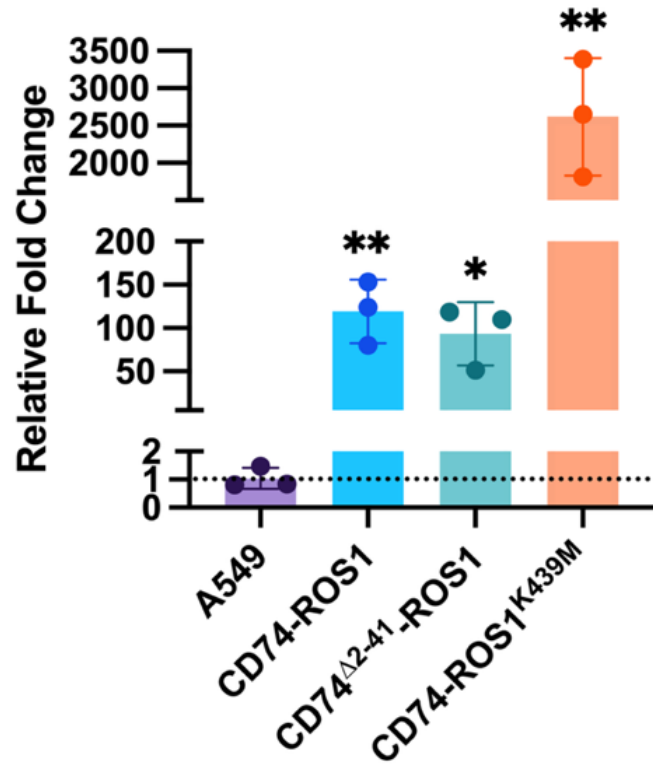
Supplementary Figure S1: Nucleic acid sequence of CD74-ROS1 (C6-R32) confirmed by Sanger sequencing.

ATGAGCAAATGCAGCCGGGGCGCCCTGTACACCGGCTTCAGCATCCTGGTTACCCTGCTGCTGGC
TGGCCAGGCCACCACTGCCTATTTCTTTATCAGCAACAGGGCAGACTGGACAAACTGACAGTG
ACCAGCCAGAACCTGCAGCTGGAAAACCTGCGGATGAAACTGCCCAAACCTCCTAAACCTGTCA
GCAAGATGAGAATGGCCACCCCTCTGCTGATGCAGGCCCTGCCTATGGGCGCCCTGCCTCAGGG
ACCTATGCAGAACGCCACCAAGTACGGCAACATGACCGAGGACCACGTGATGCACCTGCTGCAA
AACGCCGACCCCCTGAAGGTGTACCCTCCTCTGAAGGGCAGCTTTCCTGAGAATCTGAGGCACCT
GAAGAACACCATGGAAACCATCGACTGGAAAGTGTTGAGAGCTGGATGCACCACTGGCTGCTA
TTCGAGATGAGCAGACATTCTCTGGAACAGAAGCCAACCGACGCTCCACCCAAGGCCGGCGTG
CTAACAAGCCCCGGCATCCCCAAGCTGTTGGAGGGCAGCAAGAACAGCATCCAGTGGGAGAAGG
CCGAGGACAACGGCTGTAGAATCACATACTACATCCTGGAAATCCGGAAGAGCACCAGCAACA
ATCTGCAGAACCGAACCTGAGATGGAAGATGACATTCAACGGCTCCTGTAGTAGCGTGTGCAC
CTGGAAGCAAGAACCTGAAGGGCATCTTCCAGTTCAGAGTGGTGGCTGCTAACAACCTGGGA
TTCGGCGAATACAGCGGCATCAGCGAGAACATCATACTCGTGGGCGACGATTTCTGGATCCCTG
AAACCAGCTTTATCCTAACCATCATCGTGGGAATTTTCTGGTGGTGACAATCCCTCTGACCTTTG
TATGGCACCGGCGCTTAAAGAATCAGAAGTCCGCCAAGGAAGGCGTGACCGTGCTGATCAACGA
GGACAAGGAAGTGGCAGAGCTGAGAGGCCTGGCCGCTGGCGTCGGCCTGGCTAACGCCTGCTAC
GCCATCCACACCCTGCCCACCAAGAAGAGATCGAGAATCTGCCTGCCTTTCCACGGGAGAAGT
TGACACTGAGGCTGCTTCTGGGCTCCGGCGCATTCGGCGAGGTGTACGAGGTACAGCCGTGGAT
ATCCTGGGAGTCGGGAGCGGCGAGATCAAGGTGGCCGTTAAGACCCTGAAGAAGGGCTCTACCG
ATCAGGAGAAGATCGAGTTCTGAAAGAGGGCCACCTGATGAGTAAGTTCAACCACCCAATAT
CCTGAAGCAGCTGGGAGTTTGTCTGCTGAACGAGCCCCAGTACATCATCCTAGAACTGATGGAA
GGCGGCGACCTGCTGACCTACCTGCGGAAGGCCAGAATGGCCACATTCTACGGACCTCTGCTGA
CACTGGTGGATCTTGTGGACCTCTGTGTGGACATCTCCAAGGGCTGCGTGTACCTGGAAAGAATG
CACTTCATCCATAGAGATCTTGCTGCCAGAACTGCCTGGTGTCCGTTAAGGACTACACATCTCC
TAGAATCGTGAAAATCGGAGATTTCCGGCCTGGCCAGAGATATCTACAAGAACGACTACTACAGA
AAGCGGGGAGAAGGCCTGCTCCCTGTGCGGTGGATGGCCCCCTGAGAGCCTGATGGATGGCATCT
TCACCACCCAGAGCGACGTGTGGTCTTCGGCATCCTGATCTGGGAAATCCTGACACTGGGCCAC
CAGCCTTACCCCGCCCACTCTAATCTCGACGTGCTGAATTACGTGCAGACCGGAGGAAGACTGG
AACCACCTAGAAATTGCCCTGACGATCTGTGGAACCTGATGACCCAGTGCTGGGCTCAGGAGCC
TGATCAAAGACCTACCTTCCACAGAATTCAAGACCAGCTGCAACTGTTTAGAAACTTCTTCCTGA
ACAGCATCTACAAGTCCAGAGACGAGGCCAACAATAGCGGCGTGATCAACGAGTCTTTCGAGGG
GGAGGACGGCGATGTGATCTGCCTGAACTCTGATGACATCATGCCTGTGGCCCTGATGGAAACA
AAGAACCAGGAAGGCCTGAACTACATGGTGTGGCCACAGAGTGCGGCCAGGGAGAGGAAAAA
AGCGAAGGCCCTCTGGGCTCTCAGGAGAGCGAGTCATGCGGCCTGCGGAAGGAAGAAAAGGAG
CCCCACGCCGATAAGGACTTTTGCCAGGAAAAACAGGTGGCCTACTGCCCCAGCGGCAAGCCTG
AAGGCCTGAACTATGCCTGTCTGACCCACAGCGGCTACGGCGACGGCAGCGAC

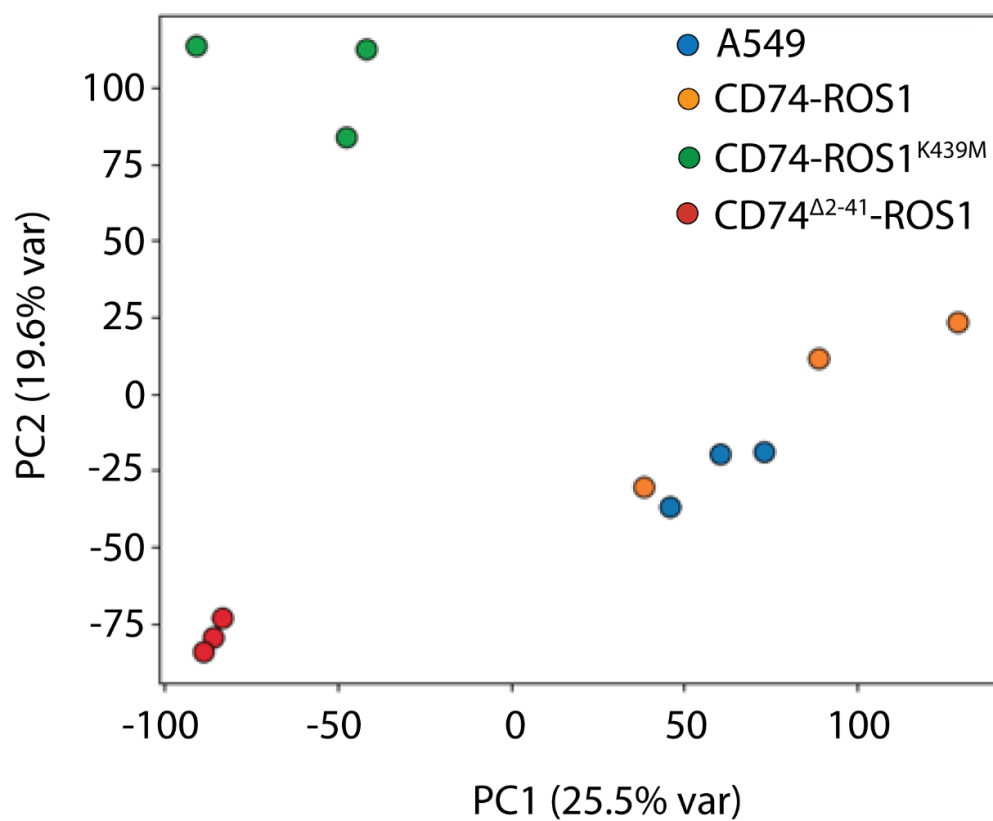
Supplementary Figure S2: Nucleic acid sequence of CD74^{A2-41}-ROS1 confirmed by Sanger sequencing.

ATGCACCGGAGAAGAAGCCGATCTTGCAGAGAGGATCAGAAACCACTGATGGACGACCAGCGG
GACCTCATTAGCAACAACGAACAGCTGCCCATGCTGGGCAGACGGCCTGGCGCTCCAGAGAGCA
AATGCAGCCGGGGCGCCCTGTACACCGGCTTCAGCATCCTGGTTACCCTGCTGCTGGCTGGCCAG
GCCACCACTGCCTATTTCTTTATCAGCAACAGGGCAGACTGGACAAACTGACAGTGACCAGCC
AGAACCTGCAGCTGGAAAACCTGCGGATGAAACTGCCCAAACCTCCTAAACCTGTCAGCAAGAT
GAGAATGGCCACCCCTCTGCTGATGCAGGCCCTGCCTATGGGCGCCCTGCCTCAGGGACCTATGC
AGAACGCCACCAAGTACGGCAACATGACCGAGGACCACGTGATGCACCTGCTGCAAAAACGCCG
ACCCCTGAAGGTGTACCCTCCTCTGAAGGGCAGCTTTCCTGAGAATCTGAGGCACCTGAAGAA
CACCATGGAACCATCGACTGGAAAGTGTTGAGAGCTGGATGCACCACTGGCTGCTATTCGAG
ATGAGCAGACATTCTCTGGAACAGAAGCCAACCGACGCTCCACCCAAGGCCGGCGTGCCTAACA
AGCCCGGCATCCCAAGCTGTTGGAGGGCAGCAAGAACAGCATCCAGTGGGAGAAGGCCGAGG
ACAACGGCTGTAGAATCACATACTACATCCTGGAAATCCGGAAGAGCACCAGCAACAATCTGCA
GAACCAGAACCTGAGATGGAAGATGACATTCAACGGCTCCTGTAGTAGCGTGTGCACCTGGA
AGCAAGAACCTGAAGGGCATCTTCCAGTTCAGAGTGGTGGCTGCTAACAACCTGGGATTTCGGCG
AATACAGCGGCATCAGCGAGAACATCATACTCGTGGGCGACGATTTCTGGATCCCTGAAACCAG
CTTTATCCTAACCATCATCGTGGGAATTTTCCTGGTGGTGACAATCCCTCTGACCTTTGTATGGCA
CCGGCGCTTAAAGAATCAGAAGTCCGCCAAGGAAGGCGTGACCGTGCTGATCAACGAGGACAA
GGAAGTGGCAGAGCTGAGAGGCCTGGCCGCTGGCGTCGGCCTGGCTAACGCCTGCTACGCCATC
CACACCCTGCCCACCAAGAAGAGATCGAGAATCTGCCTGCCTTTCCACGGGAGAAGTTGACAC
TGAGGCTGCTTCTGGGCTCCGGCGCATTCCGGCGAGGTGTACGAGGGTACAGCCGTGGATATCCTG
GGAGTCGGGAGCGGCGAGATCAAGGTGGCCGTTATGACCCTGAAGAAGGGCTCTACCGATCAG
GAGAAGATCGAGTTCCTGAAAGAGGCCACCTGATGAGTAAGTTCAACCACCCAATATCCTGA
AGCAGCTGGGAGTTTGTCTGCTGAACGAGCCCCAGTACATCATCCTAGAACTGATGGAAGGCGG
CGACCTGCTGACCTACCTGCGGAAGGCCAGAATGGCCACATTCTACGGACCTCTGCTGACACTG
GTGGATCTTGTGGACCTCTGTGTGGACATCTCCAAGGGCTGCGTGTACCTGGAAAGAATGCACTT
CATCCATAGAGATCTTGCTGCCAGAACTGCCTGGTGTCCGTTAAGGACTACACATCTCCTAGAA
TCGTGAAAATCGGAGATTTCCGCCTGGCCAGAGATATCTACAAGAACGACTACTACAGAAAGCG
GGGAGAAGGCCTGCTCCCTGTGCGGTGGATGGCCCCTGAGAGCCTGATGGATGGCATCTTCACC
ACCCAGAGCGACGTGTGGTCTTCGGCATCCTGATCTGGGAAATCCTGACACTGGGCCACCAGC
CTTACCCCGCCCACTCTAATCTCGACGTGCTGAATTACGTGCAGACCGGAGGAAGACTGGAACC
ACCTAGAAATTGCCCTGACGATCTGTGGAACCTGATGACCCAGTGCTGGGCTCAGGAGCCTGATC
AAAGACCTACCTTCCACAGAATTCAAGACCAGCTGCAACTGTTTAGAACTTCTTCCTGAACAGC
ATCTACAAGTCCAGAGACGAGGCCAACAATAGCGGCGTGATCAACGAGTCTTTCGAGGGGGAG
GACGGCGATGTGATCTGCCTGAACTCTGATGACATCATGCCTGTGGCCCTGATGGAAACAAAGA
ACCGGGAAGGCCTGAACTACATGGTGTGCTGGCCACAGAGTGCGGCCAGGGAGAGGAAAAAAGCG
AAGGCCCTCTGGGCTCTCAGGAGAGCGAGTCATGCGGCCTGCGGAAGGAAGAAAAGGAGCCCC
ACGCCGATAAGGACTTTTGCCAGGAAAAACAGGTGGCCTACTGCCCCAGCGGCAAGCCTGAAGG
CCTGAACTATGCCTGTCTGACCCACAGCGGCTACGGCGACGGCAGCGAC

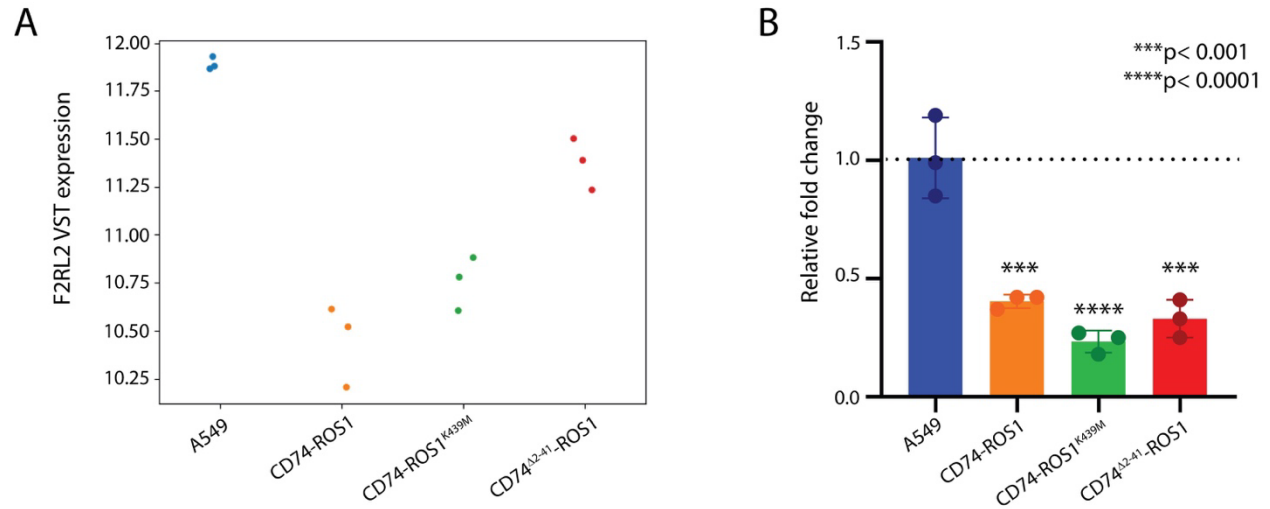
Supplementary Figure S3: Nucleic acid sequence of CD74-ROS1^{K439M} confirmed by Sanger sequencing.



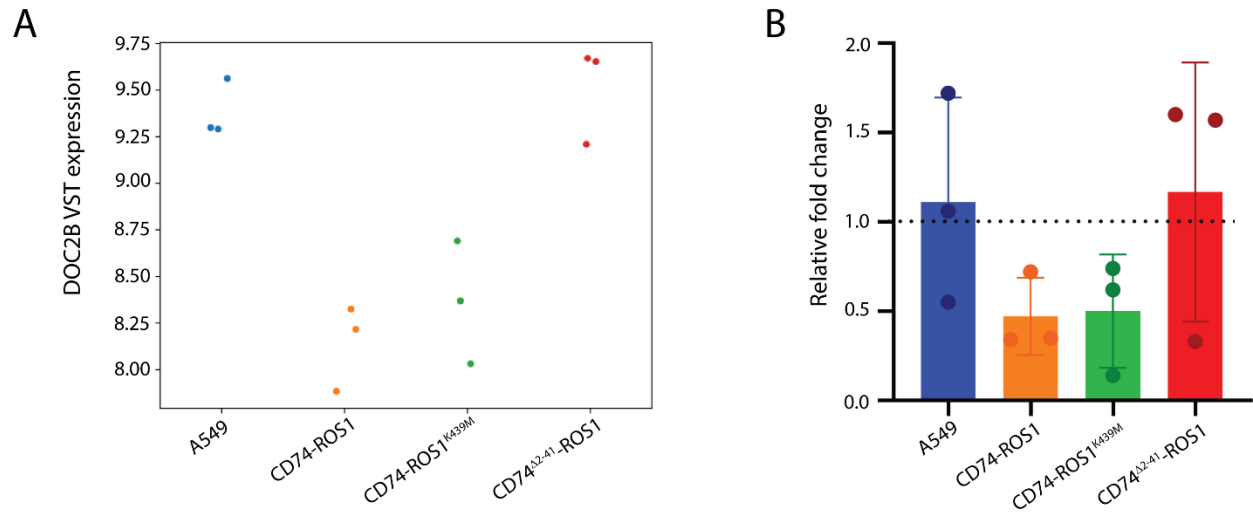
Supplementary Figure S4. CD74-Ros1 variants confirmed at the mRNA level. Relative fold change indicates gene expression levels of CD74-Ros1, CD74 Δ 2-41-Ros1, and CD74-Ros1^{K439M}, compared to non-transfected A549 cells. Experiments were carried out in biological triplicates. Data are shown as the mean $\pm$ SD (n = 3) after normalization to the control (A549). The dotted line represent no fold change in comparison to the control; *P < 0.05 and **P < 0.01



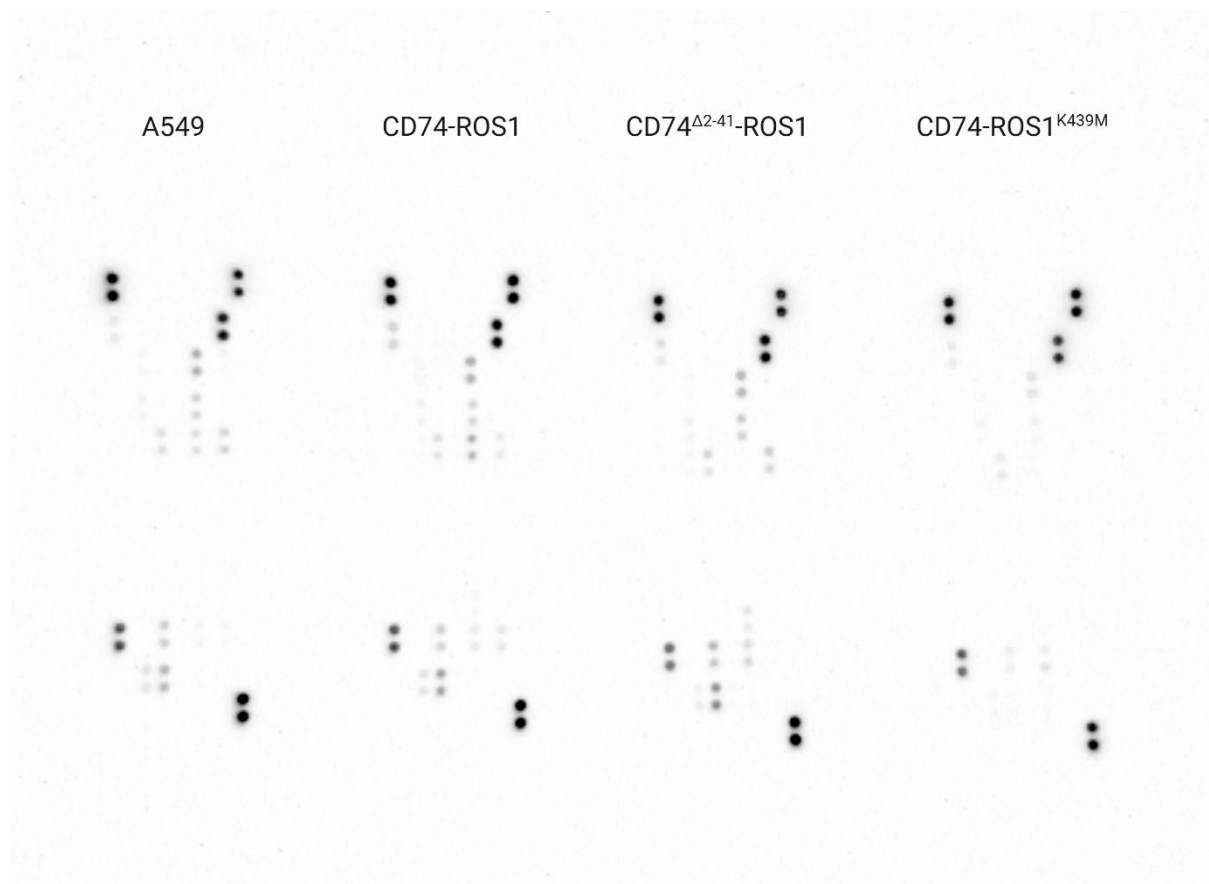
Supplementary Figure S5. Principal component analysis (PCA) analysis, demonstrating the variation in global gene expression between A549 and CD74-ROS1 variants.



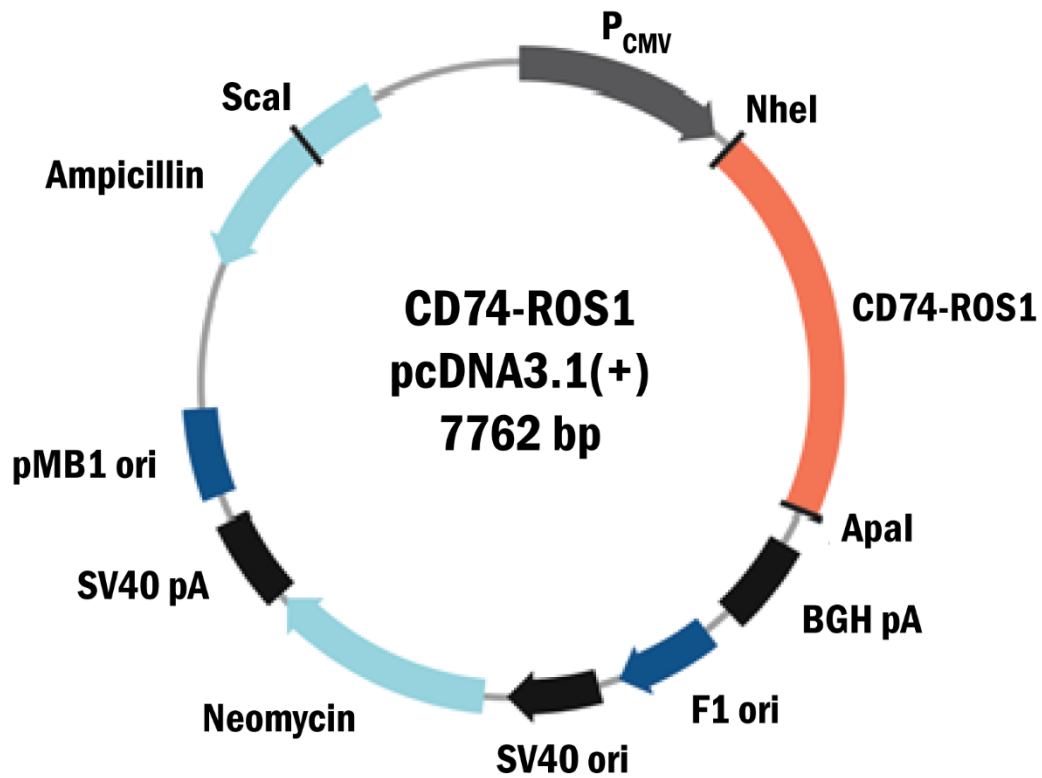
Supplementary Figure S6: F2RL2 mRNA analysis A) F2RL2 high throughput mRNA differences between different variants B) F2RL2 qPCR confirmation of high throughput screening in biological triplicate. The results are presented as mean $\pm$ standard deviation (SD). ***p<0.001 ****p<0.0001



Supplementary Figure S7: DOC2B mRNA analysis A) DOC2B high throughput mRNA differences between different variants B) DOC2B qPCR confirmation of high throughput screening in biological triplicate. The results are presented as mean $\pm$ standard deviation (SD).



Supplementary Figure S8: Raw protein phosphorylation data obtained from the human phospho-kinase array kit (R&D Systems, Cat No. ARY003C) for A549 and CD74-ROS1 variants. Phosphorylation of 37 human kinases, divided into membrane A (top) and B (bottom) were measured simultaneously.



Supplementary Figure S9: CD74-ROS1 plasmid design. CD74-ROS1 gene and its two mutants were cloned into the pcDNA3.1(+) vector between the **NheI** and **ApaI** cloning sites.