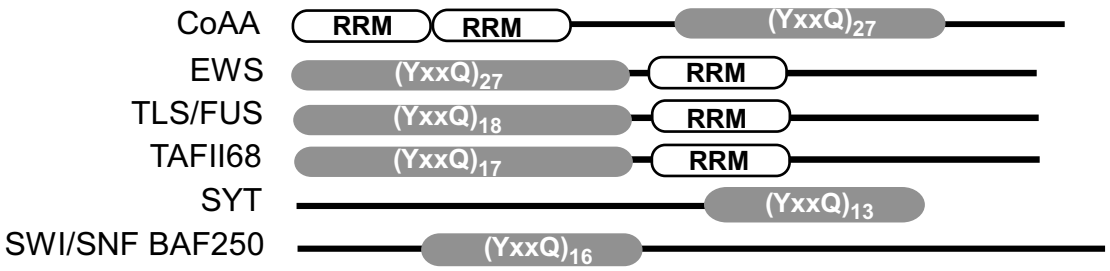
Table S1. The YxxQ motif-containing proteins identified from Swiss-Port/TrEMBL databases

Identity	Name of entry	Accession number	Species	Fragment size (aa)
CoAA	CoAA	Q96PK6	human	669
	SYT-IP	O75932	human	669
	CoAA	Q920B2	rat	278
	RBM14	Q91Z21	mouse	618
	16 kDa protein	Q62019	mouse	573
	RBM14	Q9DBI6	mouse	669
	EST	Q8C2Q3	mouse	669
EWS	EWS	Q01884	human	656
	EWS-Fli	Q9BZD1	human	476
	FLJ31747 (EWS)	Q96MX4	human	661
	Similar to EWS	Q9BWA2	human	354
	FLJ32119 (EWS)	Q96MN4	human	600
	Similar to EWS	Q96FE8	human	655
	EWS fragment	Q7Z6Z3	human	75
	EWS homolog	Q9D2P0	mouse	655
	EWS	Q61545	mouse	655
	EWS	Q6NVA3	mouse	656
FUS	FUS	P35637	human	526
	FUS-ATF1	Q9H4A8	human	226
	Fusion t(12,16)	Q8TBR3	human	526
	FUS-CHOP	Q16273	human	126
	FUS-like	Q13344	human	528
	FUS	P56959	mouse	518
	Similar to FUS	Q91VQ2	mouse	280
	Similar to FUS	Q8CFQ9	mouse	517
	FUS	Q28009	bovine	512
TAFII68	TAFII68	Q92804	human	592
	TAFII68	Q86X94	human	501
	TAFII68	Q8BQ46	mouse	557
SYT	SYT	Q15532	human	418
	SYT homolog-1	Q75177	human	396
	SS18 (SYT)	Q6FGL9	human	391
	SS18L1	Q8NE69	human	396
	SYT fragment	Q8TDQ9	human	414
	SYT	Q62280	mouse	418

	SS18	Q6P1I1	mouse	392
	SS18-like1	Q7TQF3	mouse	402
BAF250	SWI/SNF p270	O14497	human	1902
	SWI/SNF subunit OSA1	Q8NFD6	human	1999
	Chromatin remodeling p250	Q9BY33	human	1939
	SWI related	Q96TA9	human	1208
	OSA1	Q96T89	human	1685
	BRG1-associated 250a	Q9HBJ5	human	2285
	OSA1	Q925Q1	mouse	1902

The protein databases of Swiss-Port (release 44.3; 156998 entries) and TrEMBL (release 27.3; 1379120 entries) were analyzed with ScanProsite program. Prosite format scanned for the YxxQ motifs: Y-{P}(1,2)-Q(1,2)-X(1,4)-Y-{P}(1,2)-Q(1,2)-X(1,4)-Y-{P}(1,2)-Q(1,2). The taxonomic species filter was set as *Homo sapiens, Mus musculus, Rattus norvegicus, and Bos taurus*.

a



b

CoAA *Q96PK6* 669aa (211-570 is shown)

```
GRDRSPLRRS PPRASYVAPL TAQPATYRAQ PSVSLGAAYR AQPSASLGVG YRTQPMTAQA 270
ASYRAQPSVS LGAPYRGQLA SPSSQSAAAS SLGPYGGAQP SASALSSYGG QAAAASSLNS 330
YGAQGSSLAS YGNQPSSYGA QAASSYGVRA AASSYNTQGA ASSLGSYGAQ AASYGAQSAA 390
SSLAYGAQAA SYNAQPSASY NAQSAPYAAQ QAASYSSQPA AYVAQPATAA AYASQPAAYA 450
AQATTPMAGS YGAQPVVQTQ LNSYGAQASM GLSGSYGAQS AAAATGSYGA AAAYGAQPSA 510
TLAAPYRTQS SASLAASYAA QQHPQAAASY RGQPGNAYDG AGQPSAAYLT MSQGAVANAN 570
```

EWS *Q01844* 656aa (1-300 is shown)

```
MASTDYSTYS QAAAQQGYSA YTAQPTQGYA QTTQAYGQQS YGTYGQPTDV SYTQAQTTAT 60
YGQTAYATSY GQPPTGYTTP TAPQAYSQPV QGYGTGAYDT TTATVTTTQA SYAAQSAYGT 120
QPAYPAYGQQ PAATAPTRPQ DGNKPTETSQ PQSSTGGYNQ PSLGYGQSNY SYPQVPGSYP 180
MQPVTAPPSY PPTSYSSTQP TSYDQSSYSQ QNTYGQPSSY GQQSSYGQQS SYGQQPPTSY 240
PPQTGSYSQA PSQYSQQSSS YGQQSSFRQD HPSSMGVYGQ ESGGFSGPGE NRSMSGPDNR 300
```

TLS/FUS *P35637* 526aa (1-240 is shown)

```
MASNDYTQQA TQSYGAYPTQ PGQGYSQQSS QPYGQQSYSG YSQSTDTSGY GQSSYSSYGQ 60
SQNTGYGTQS TPQGYGSTGG YGSSQSSQSS YGQQSSYPGY GQQPAPSSTS GSYGSSSQSS 120
SYGQPQSGSY SQQPSYGGQQ QSYGQQQSYN PPQGYGQQNQ YNSSSGGGGG GGGGGNYGQD 180
QSSMSSGGGS GGGYGNQDQS GGGGSGGYGQ QDRGGRGRGG SGGGGGGGGG GYNRSSGGYE 240
```

TAFII68 *Q92804* 592aa (1-180 is shown)

```
MSDSGSYGQS GGEQQSYSTY GNPGSQGYGQ ASQSYSGYGQ TTDSSYGQNY SGYSSYGQSQ 60
SGYSQSYGGY ENQKQSSYSQ QPYNNQGQQQ NMESSGSQGG RAPSYDQPDY GQQDSYDQQS 120
GYDQHQGSYD EQSNYDQQHD SYSQNQQSYH SQRENYSHHT QDDRRDVSRY GEDNRGYGGS 180
```

SYT *Q15532* 418aa (221-418 is shown)

```
GGQHYQGQQP PMGMMGQVNQ GNHMMGQRQI PPYRPPQQGP PQQYSGQEDY YGDQYSHGGQ 280
GPPEGMNQQY YPDGHNDYGY QQPSYPEQGY DRPYEDSSQH YYEGGNSQYG QQQDAYQGPP 340
PQQGYPPQQQ QYPGQQGYPG QQQGYGPSQG GPGPQYPNYP QGQGQQYGGY RPTQPGPPQP 400
PQQRPYGYDQ GQYGNYQQ
```

BAF250 *Q8NFD6* 1999aa (61-360 is shown)

```
AAASGGAQQR SHHAPMSPGS SGGGGQPLAR TPQPSSPMDQ MGKMRPQPYG GTNPYSQQQG 120
PPSGPQQGHG YPGQPYGSQT PQRYPMTMQG RAQSAMGGLS YTQQIPPYGQ QGPSGYGQQG 180
QTPYYNQQSP HPQQQQPPYS QQPPSQTPHA QPSYQQQPQS QPPQLQSSQP PYSQQPSQPP 240
HQQSPAPYPS QQSTTQQHPQ SQPPYSQPQA QSPYQQQQPQ QPAPSTLSQQ AAYPQPQSQQ 300
SQQTAYSQQR FPPPQELSQD SFGSQASSAP SMTSSKGGQE DMNLSLQSRP SSLPDLSGSI 360
```

Figure S1 The YxxQ-containing proteins. A. Schematic representation of CoAA, EWS, TLS/FUS, TAFII68, SYT, and SWI/SNF BAF250 structures in which RRMs are shown as open boxes and repeated tyrosine- and glutamine-rich sequences (YxxQ) are shown as filled boxes with number of repeats indicated. B. The YxxQ-containing sequences from each protein are shown, with their tyrosine residues (Y) highlighted, and residues within the motif including glutamine residues (Q) shaded. The YxxQ, YxQQ, and YxQ sequence patterns are selected and shown.

Figure S2 Primer pairs used in PCR analysis.

P1-F: AAGAGCGCGGGCACAGCAAC
P1-B: CGGGATGTGGAGCGAAGGTCA

P2-F: AGGAGGCTGAGTTGGGAGGAT
P2-B: AGGAGGCTCTGGGAGGAAAGG

P3-F: CCAGCCCATTTTTATTGAGTTCCT
P3-B: CCAAGGCTGAAAGGTAACCCTCTA

P4-F: GCCGGCATGATCTTCCCTCAGAGGATG
P4-B: AGGACTTGGAGTAATCAAGGGAACTA

P5-F: CGACGGGGCGGATACGACTCCGGAGGA
P5-B: GCACTTCCCCACCCCTCCCA

P6-F: CTTCGACTACCAGCAGGCTTTT
P6-B: ~~GAATACACCGAGGAGACTGCC~~ Erratum: CCGTCAGAGGAGCCACATAAG

P7-F: GTATGGTTCCGACCGGCGTTTAGCCGAG
P7-B: GCTGCTCTCGAGCTACATGCGGCGCTGGT

P8-F: GGTACTGTCCGCAAGGTTCATGTG
P8-B: TCTGTTGAAGGCTTGGGGATGG

P9-F: CTGCTCTGTGTAGGGCTCGTCC
P9-B: CATCAGGCTGCACAGGCGG

P10-F: CCACAAGCTAGACCATCAGGAAGG
P10-B: GCAACAGCAGGTGAATCAAAGTAGG

P11-F: TGAGACCCTGTGCCAGCCCTG
P11-B: GCCGGAATACCATCAGCCACA

P12-F: ACCAAGAACCTGAAGCCTGTACCC
P12-B: GTGAAACAGCTCAGCTCCAGC

P13-F: AGTCGCGCCACCAGCATCTTC
P13-B: GGTGACGTTGGGATCCAGGGTAA

P14-F: CCTCGGTGTCCTACTTCAAAT
P14-B: TCCAGGTGGCGACGATCTTC

P15-F: GCCGGCATGATCTTCCCTCAGAGGATG
P15-B: AACATAGACATGTGACAAACTCAGGGAA

P16-F: CTGCCAGGACCTCAGCAACCCCAGG
P16-B: TGAGTGTACGGGGGAAGGCCGG

P17-F: TCGCTGGACGCGCGGGATGCGAGTCCCC
P17-B: CCACCTTCCTGGAGAGCCCGGACCTCT

TRalpha-F: TCACCTGGCAACTTCAATGC
TRalpha-B: CCTGATTTTCCCAGCGATGT

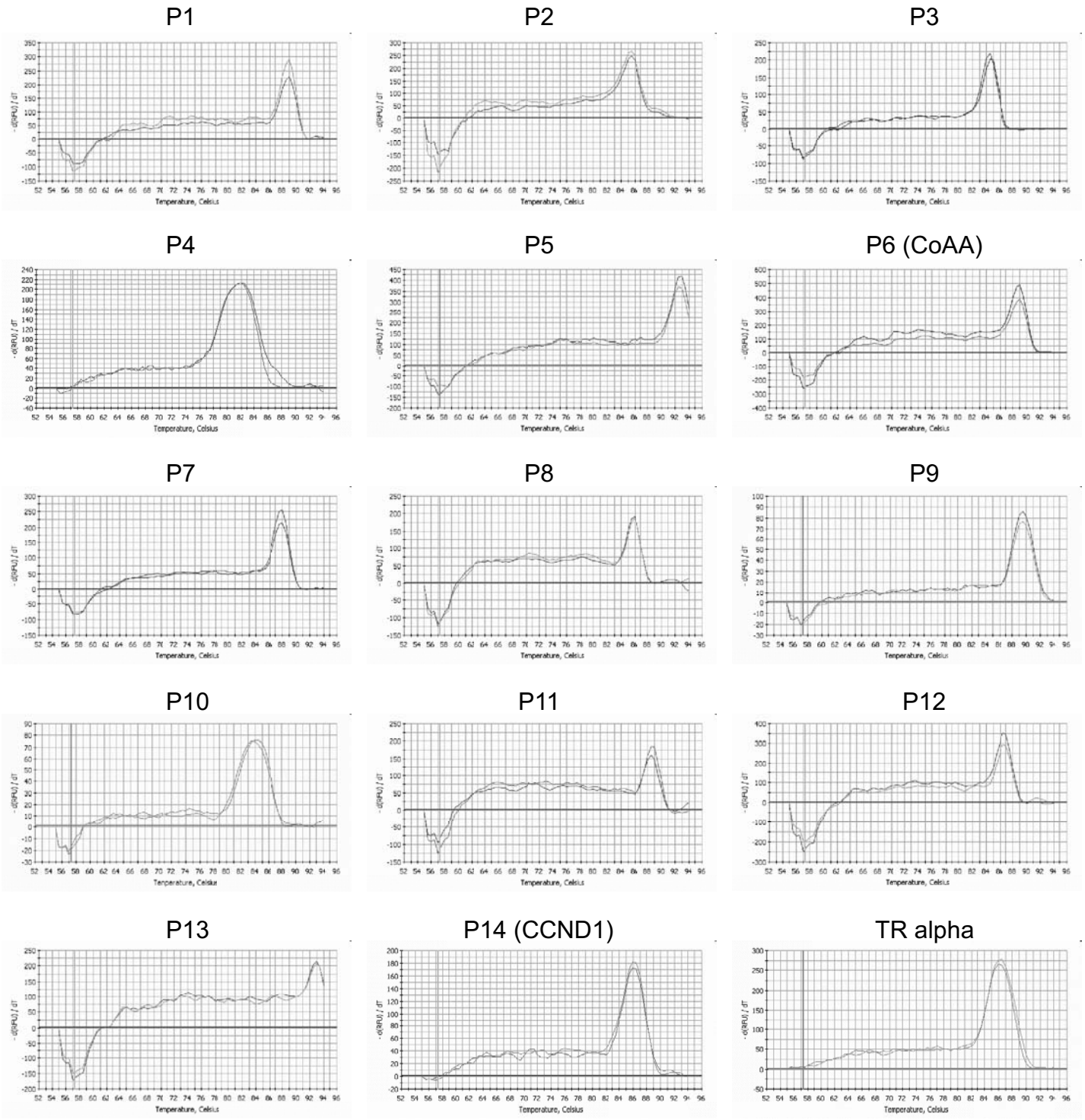


Figure S3 Dissociation curve. Primer pairs used in amplicon mapping were assayed in duplicate using normal genomic DNA as template.

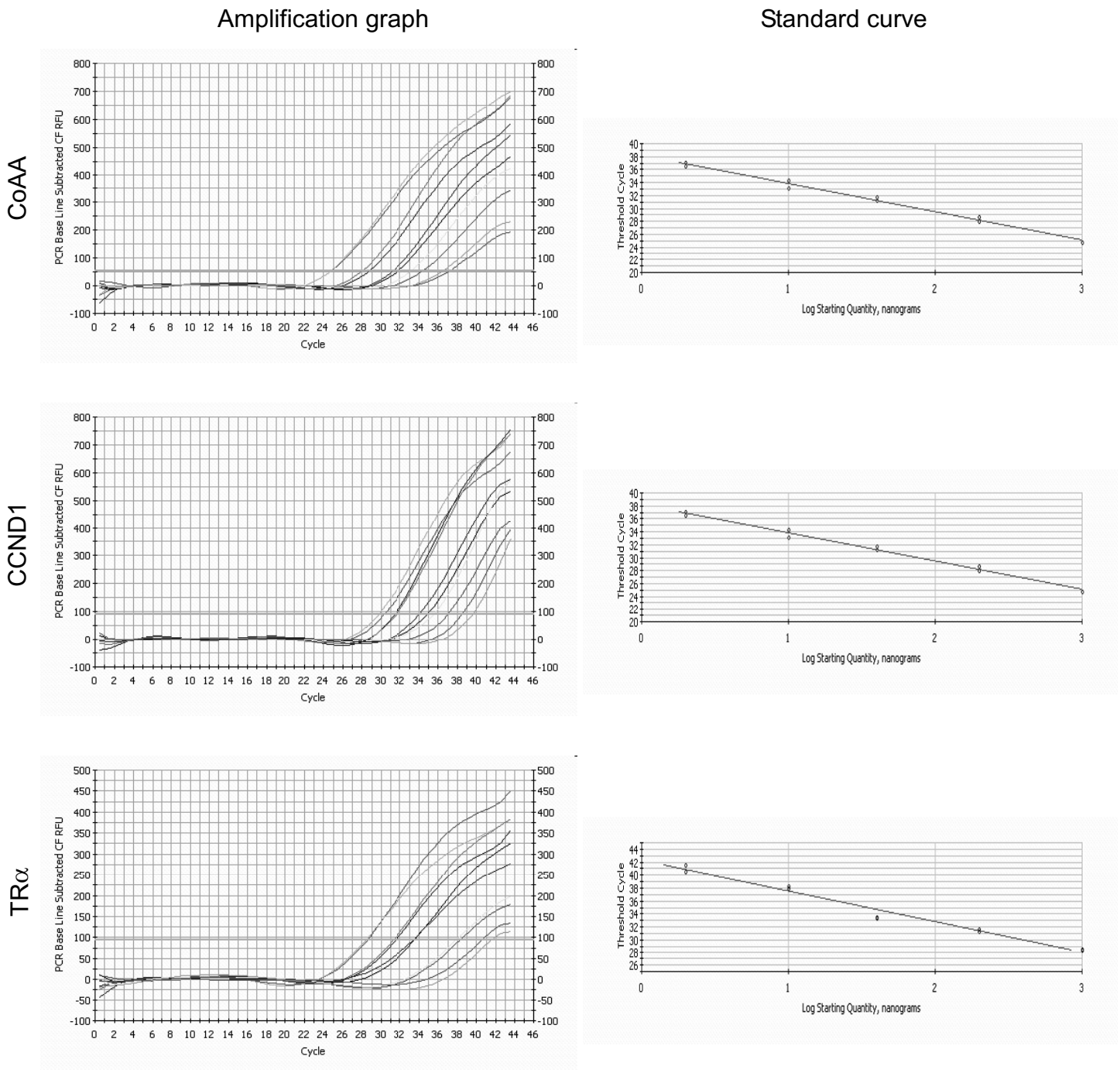


Figure S4 Standard curves of real-time PCR analysis. Standard curves of PCR for CoAA, CCND1 and TRα were generated by using serial dilutions of normal human genomic DNA as template.

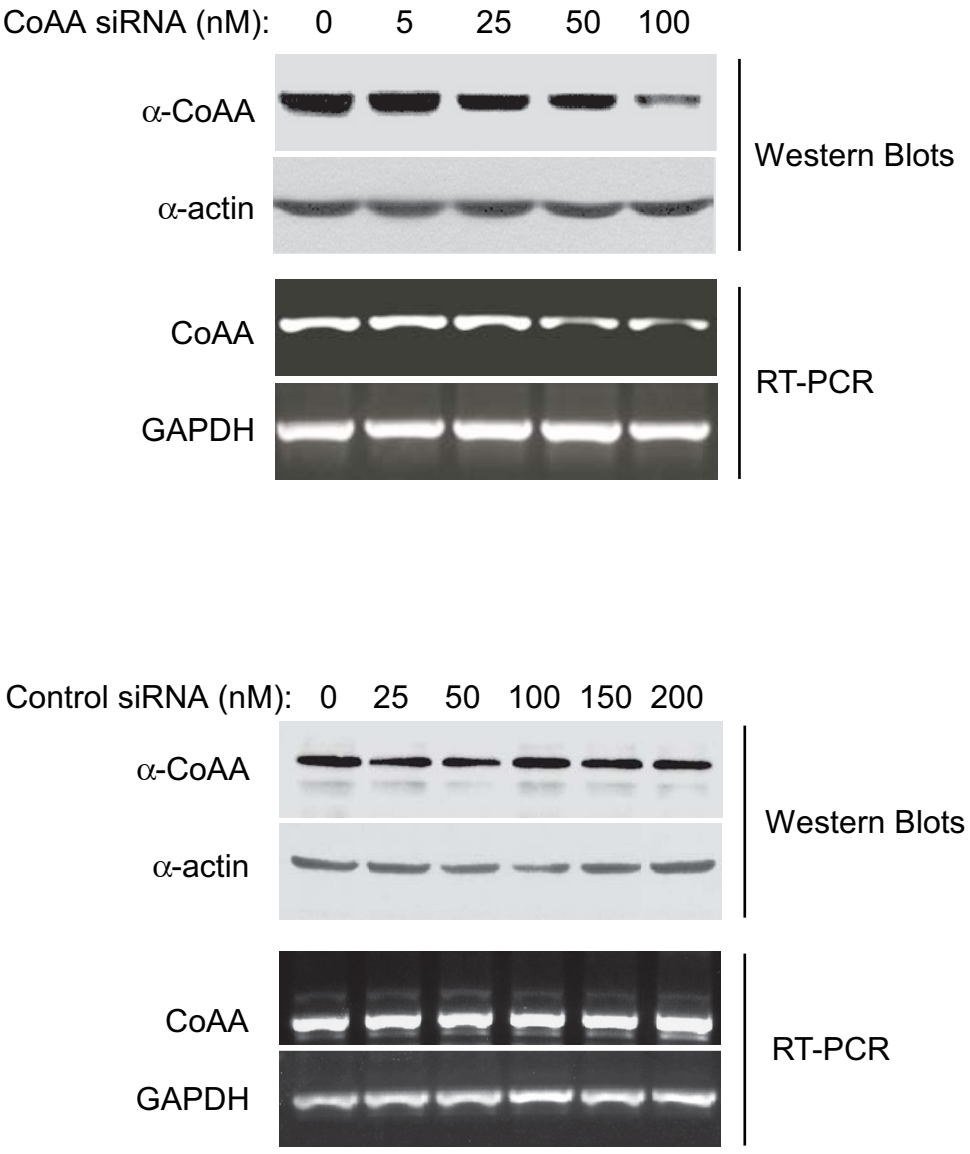


Figure S5 siRNA of CoAA. CoAA or control siRNA was transfected in cells at increasing concentrations as indicated. Western blots were carried out using anti-CoAA and anti-actin antibodies. RT-PCR was performed using CoAA and GAPDH primer pairs. The results suggested that siRNA of CoAA inhibits CoAA mRNA and protein expression.

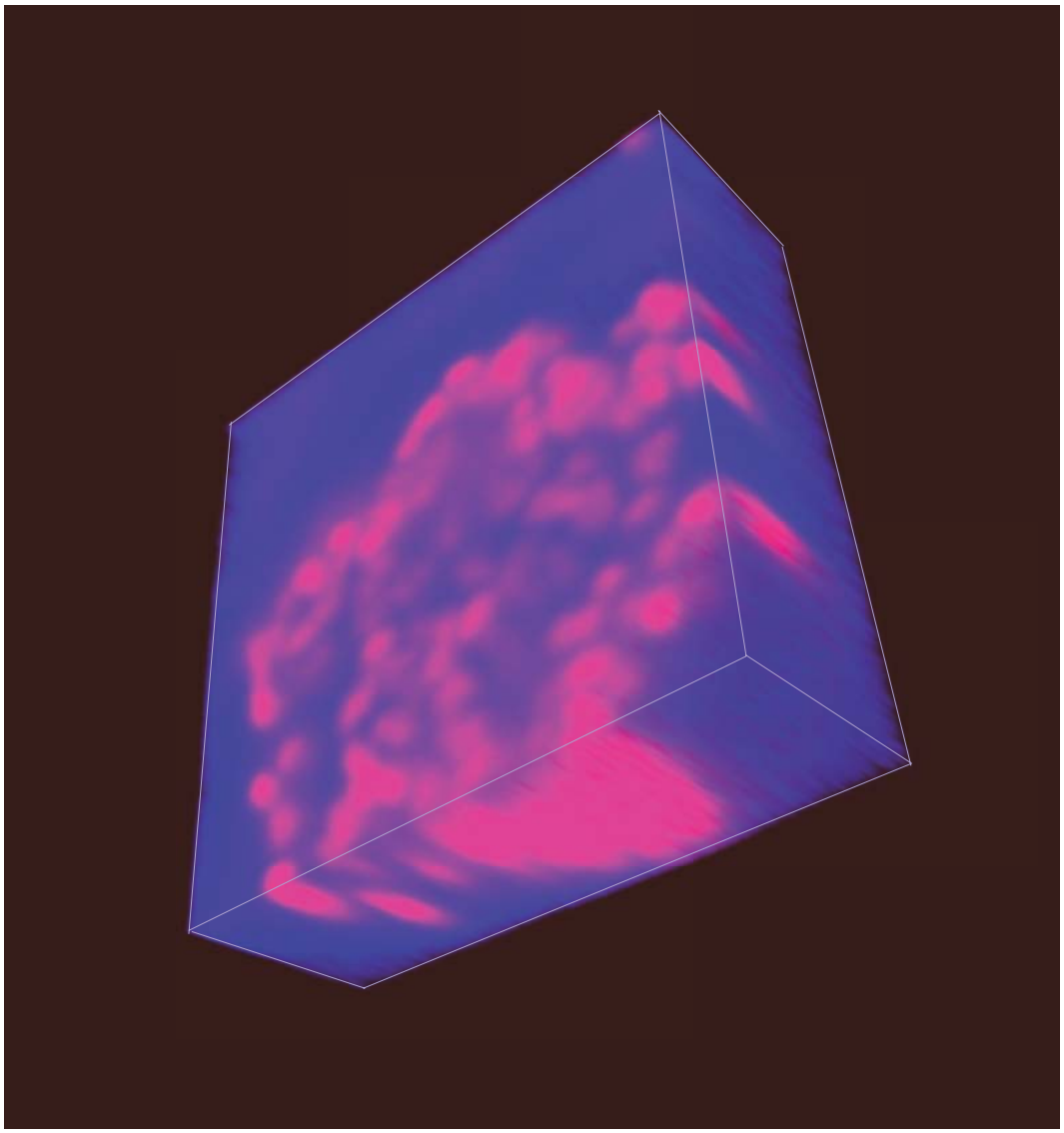


Figure S6 Three dimensional scanning of a single nucleus from a squamous cell lung carcinoma with amplified CoAA gene stained by fluorescent *in situ* hybridization (FISH) (red). The slide is counterstained with DAPI (blue).