

Supporting Information

Hopkins et al. 10.1073/pnas.1418805112

SI Experimental Procedures

LOGO Analysis. Logo sequences in Fig. S1 were generated using WebLogo, version 3.4.

siRNA Sequences. The siRNA sequences were as follows: DCP2, #1 siRNA ID s46660, sense siRNA sequence: CCCUUUAU-CAGACCAUUAAtt; DCP2, #2 siRNA ID s46659, sense siRNA sequence: GGAUAAGAGACUUUGCUAAtt; NUDT16, #1 siRNA ID s43641, sense siRNA sequence: GCCUAUUU-GUUAGGGAGGAtt; and NUDT16, #2 siRNA ID s43640, sense siRNA sequence: GCACGUUAAAGAAGACAGAtt.

Primer Sequences. The primer sequences were as follows: 5'-RACE adaptor outer primer: F: 5'-GCTGATGGCGATGAA-TGAACACTG-3'; 5'-RACE adaptor inner primer: F: 5'-CGC-GGATCCGAACACTGCGTTTGCTGGCTTTGATG-3'; MP12 N: R: 5'-GGGCTTGTTGCCACGAGTTAGA-3'; MP12 N: F: 5'-CAAGCAGTGGACCGCAATGAGA-3'; human HIS3 5'-end forward primer (with T7 linker): 5'-TAATACGACTCACTA-TAGGCTATGGCAGC-3'; human H2A 5'-end forward primer (with T7 linker): 5'-TAATACGACTCACTATAGGATGTCNG-GACG-3'; human RPS3A 5'-end forward primer (with T7 linker): 5'-TAATACGACTCACTATAGGTCTGACCAGCACC-3'; human RPS3A reverse: 5'-TTGAGACCATCAGATGCAATTT-3'; human RPL37A 5'-end forward primer (with T7 linker): 5'-TAATACGACTCACTATAGGCTTTCTGGGCTC-3'; human RPL37A reverse: 5'-TCTTCATGCAGGAACCACAG-3'; human EEF2 5'-end forward primer (with T7 linker): 5'-TAAT-ACGACTCACTATAGGCTCTCCGCCGTC-3'; human EEF2

reverse: 5'-GGATCTGGTCTACCGTGAAGT-3'; human RPL24 5'-end forward primer (with T7 linker): 5'-TAATACGACT-CACTATAGGtcttttctccg-3'; human RPL24 reverse: 5'-GG-TAGATCTTGTACCCCGCTAAA-3'; human RPS8 5'-end forward primer (with T7 linker): 5'-TAATACGACTCACTATAG-GCTCTTTCCAGCCAGC-3'; human RPS8 reverse: 5'-CTTCT-TGTGGTAGGGCTTTCT-3'; human RPS23 5'-end forward primer (with T7 linker): 5'-TAATACGACTCACTATAGGCT-CTTTCCGCTCAGG-3'; human RPS23 R: 5'-GCCCCAAAT-GAGCTTTCTTATACTG-3'; human EEF1A1 5'-end forward primer (with T7 linker): 5'-TAATACGACTCACTATAGG-CTTTTCGCA-3'; human RPL10 5'-end forward primer (with T7 linker): 5'-TAATACGACTCACTATAGGCTCTTTCCCT-TC-3'; human 28S rRNA: F: 5'-GGGTGGTAAACTCCATC-TAAGG-3'; R: 5'-GCCCTCTTGAAGTCTCTCTTC-3'; human GAPDH: F: 5'-ACCAAATCCGTTGACTCCGACCTT-3'; R: 5'-TCGACAGTCAGCCGCATCTTCTTT-3'; NUDT16 qPCR primers (human): F: 5'-GGTAGGCAGCCACTATCATT-3'; R: 5'-GCAGTCCCTGCAGCTATATT-3'; DCP2 qPCR primers (human): F: 5'-TCCTCGAGAGGTGGAGAAA-3'; R: 5'-GAGCCACAGCTTCAGTATAA-3'; mouse RPS3A 5'-end forward primer (with T7 linker): 5'-TAATACGACTCACTATAG-GCTCCCCGCCGCGC-3'; mouse RPS3A reverse: 5'-GCTTG-TTCTTGCCGACCGCCAT-3'; mouse EEF2 5'-end forward primer (with T7 linker): 5'-TAATACGACTCACTATAGGctct-tccgccgag-3'; mouse EEF2 reverse: 5'-TGATGGCACGGATCT-GATCTACT-3'; mouse 28S rRNA: F: 5'-GGTTGAGGGCCA-CCTTATT-3'; R: 5'-GAAGAAAGACCGGGAAGAGAAA-3'.

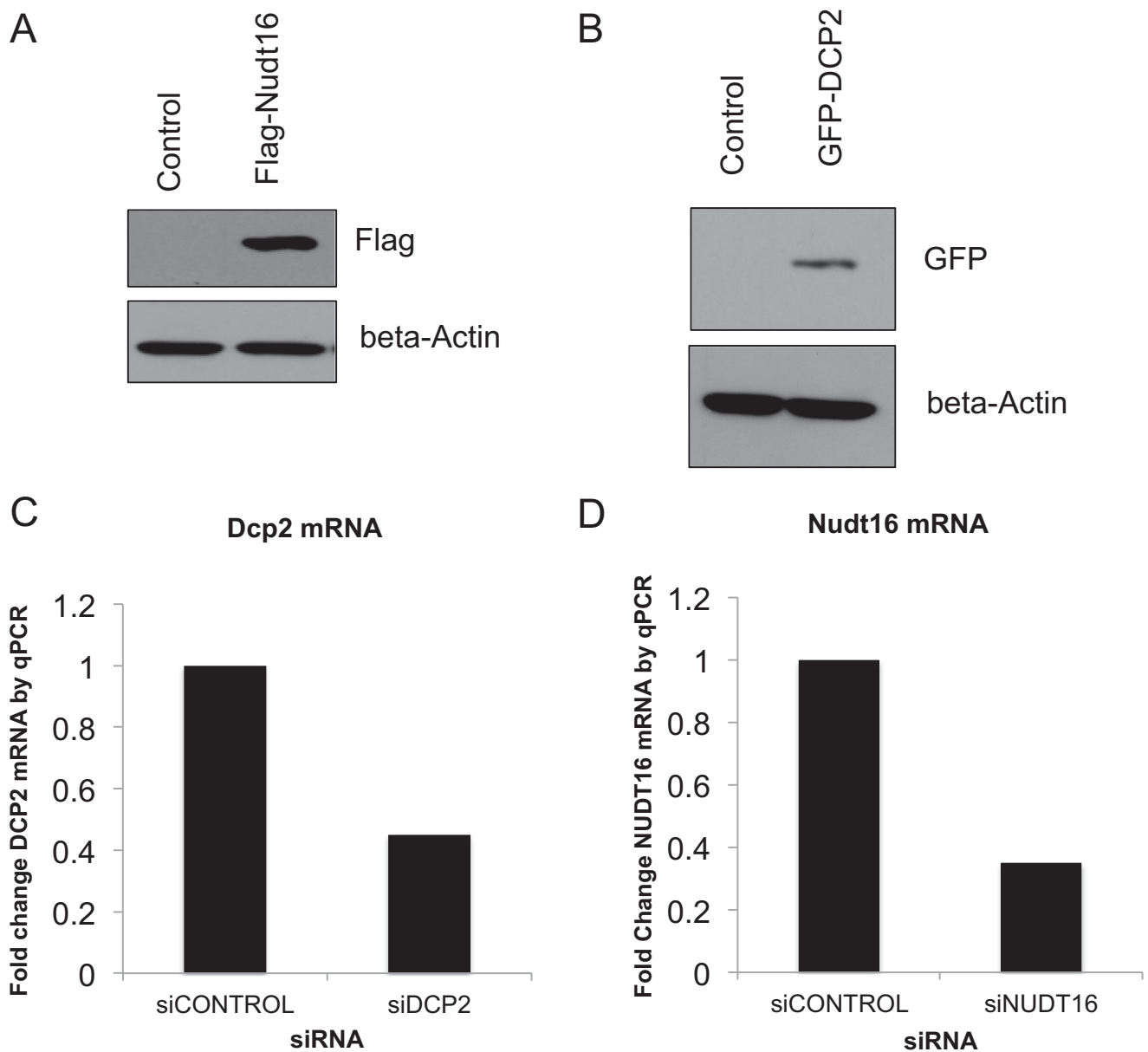


Fig. S3. Overexpression and siRNA knockdown of DCP2 and NUDT16 in U2OS cells. (*A* and *B*) U2OS cells were transfected with the indicated plasmid. (*A*) Control U2OS cells or U2OS cells expressing ectopic Flag-tagged NUDT16. Flag-NUDT16 expression was assessed by immunoblot. (*B*) Total protein was collected from control U2OS cells or U2OS cells expressing ectopic GFP tagged DCP2; GFP-DCP2 expression was assessed by immunoblot. (*C* and *D*) U2OS cells were treated with the indicated siRNA, and total RNA was collected 3 d post knockdown. RNA was analyzed by RT-qPCR and normalized to GAPDH. (*C*) DCP2 mRNA levels normalized to siCONTROL-treated U2OS cells. Mean is shown; $n = 1$. (*D*) NUDT16 mRNA levels normalized to siCONTROL-treated U2OS cells. Mean is shown; $n = 1$.

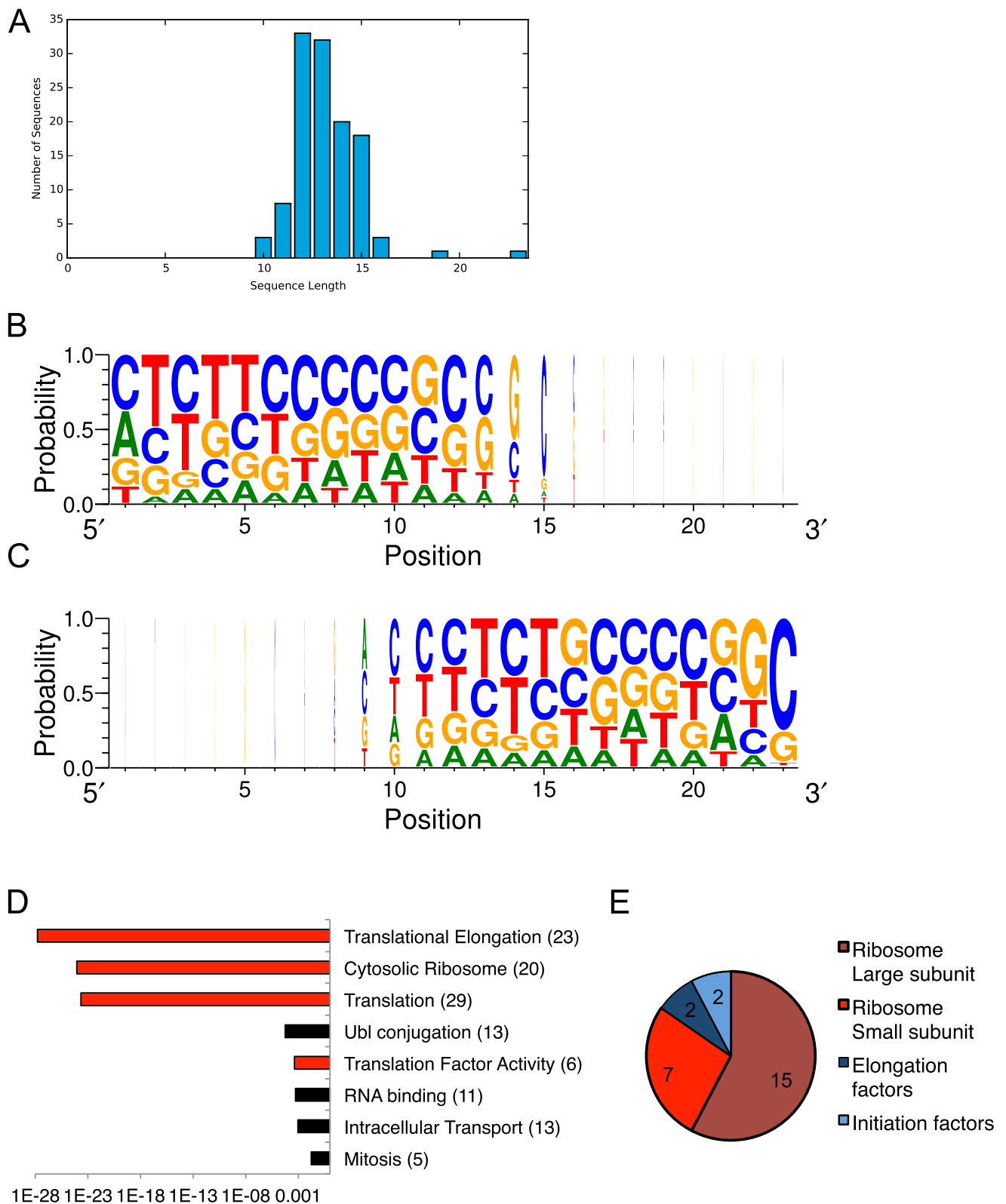


Fig. S4. Analysis of cap-snatched mRNAs. (A) Bar graph of size distributions of snatched sequences. (B) LOGO analysis of nucleotide sequence of full-length cap-snatched 5'-UTRs aligned to the 5' end. (C) LOGO analysis of nucleotide sequences of full-length cap-snatched 5'-UTRs aligned to the 3' end. (D) Analysis of GO enrichment of 119 snatched sequences shown as a function of P value. Translationally related categories are highlighted in red. (E) Pie chart of cap-snatched genes that are part of the core translation machinery, as shown by subunit or function.

Table S1. 5'-RACE results of RVFV cap-snatched sequences in human U2OS cells

Snatched sequence, 5' to 3'	Official gene symbol, human	5'-UTR type	Length, nt	Genomic coordinates
ACCGCCGAGACC	ACTB		12	chr7: 5570220–5570232
GCCGAGACCGC	ACTB		11	chr7: 5570218–5570229
AGAGCGGAGGCGC	ADC		13	chr1: 33546820–33546833
ACGGCCAGGTTGGGT	AGL		15	chr1: 100315691–100315706
AGAGCCGCCATC	ARF1		12	chr1: 228270408–228270420
AGCGGCAGCGGC	ATP1B1		12	chr1: 169075941–169075953
CTTTTCTCTCTC	BET1	TOP	12	chr7: 93633540–93633552
CACTCGTGTCTC	BEX1		12	chrX: 102319088–102319100
AGTGGCCATCTGC	BTF3		14	chr5: 72794272–72794286
CTTTTCATTTCC	C17orf76-AS1	TOP	12	chr17: 16342350–16342362
GTATCTAGCATTTT	C4orf29		14	chr4: 128886618–128886632
GCTGCCGCCGTGCG	CADM2		14	chr3: 85008144–85008158
AGGCCCGGGAGCGGC	CAPNS1		15	chr19: 36631006–36631021
ATGGGCTGTGGGTGCCGCTGAGC	CBX6		23	chr22: 39268198–39268221
ACTTGGCTTCAAAGC	CDK1		15	chr10: 62538237–62538252
ATTTATTACTTC	CHL1		12	chr3: 290267–290279
CTCCATCGCCGCC	CSNK2A3		13	chr11: 11374834–11374847
CTCCCCACCCGCTC	CREBBP	TOP	15	chr16: 3930066–3930081
AGTGCGGGGTCCGGC	CTS2		14	chr20: 57582293–57582307
AGAGTCCGAGCCG	CUL3		13	chr2: 225449913–225449926
CTCTTTGTGAGC	DANCR	TOP	12	chr4: 53578670–53578682
CTCAGCAGCCAG	DDHD2		12	chr8: 38089020–38089032
GTTTCGCCTCAGG	DERL1		13	chr8: 124054539–124054552
CCCCTCTGTCTT	DESI1	TOP	14	chr22: 42017036–42017050
GTTTTTCAAAG	DLEU7-AS1	TOP-LIKE	12	chr13: 51423650–51423662
CTTTTTCGCA	EEF1A1	TOP	10	chr6: 74230745–74230755
CTCTCCGCCGTAC	EEF2	TOP	15	chr19: 3985446–3985461
AGTCACTGAGCCG	EIF1		13	chr17: 39845141–39845154
TTCTCTCTCGGC	EIF3C	TOP-LIKE	12	chr16: 28415152–28415164
ACTGATCGGTGAGGC	ERLIN1		15	chr10: 101945711–101945726
GCAGCCGGCGGCGC	ESRRA		14	chr11: 64073051–64073065
GCACAGAGTCTGC	ESX1		13	chrX: 103499564–103499577
ACCGACAAAATGGC	ETAA1		15	chr2: 67624464–67624479
AGGAGCGTAGAGGC	FA2		14	chr5: 175875385–175875399
AGACGTTCTTCGC	FTH1		13	chr11: 61735090–61735103
GCAGTTCGGC	FTL		10	chr19: 49468565–49468575
ACACAACCAGC	FXVD2		11	chr11: 117698738–117698749
CGCCTGGACGCGAG	FXVD5		13	chr19: 35645640–35645653
CTTGAGAAAGCAAG	GPR61		14	chr1: 110082516–110082530
CGCTACCCGGTTGC	GRN		15	chr17: 42422656–42422671
GCTCAATTTCTC	HAX1		12	chr1: 154245112–154245124
GGCGCAGAGGCCTGC	HBS1L		15	chr6: 135375945–135375960
GCGGAGCCGCG	HDAC1		11	chr1: 32757709–32757720
ATTATTTTCTC	HIST1H2BD	TOP-LIKE	11	chr6: 26158349–26158360
GACTTTCCCGATCGC	HIST2H2AA4	TOP-LIKE	15	chr1: 149814302–149814317
CTCAGGACTCAG	HLA-E		12	chr6: 30457285–30457297
GCGGACTGTCCTTTC	HMGCS1		15	chr5: 43313499–43313514
TTTCTGCCGTGGACG	HNRNPA1	TOP-LIKE	16	chr12: 54674532–54674548
ACAAAGCCCAGACG	HNRNPM		14	chr19: 8509860–8509874
AGTTGCTTCAGC	HSP90AA1		12	chr14: 102553369–102553381
CTCATTGAACTCGC	HSPA8		14	chr11: 122932830–122932844
AGGCGCTGTCTC	HTATIP2		12	chr11: 20385297–20385309
GAGGTCGCTG	IGSF21		11	chr1: 18434403–18434414
AGTAGCAGCAGC	IMPDPH1		12	chr7: 128045984–128045996
TGGCCCCAGCGGTG	IWS1	TOP-LIKE	14	chr2: 128284040–128284054
CGCCACCGTCTGTC	KRT8		13	chr12: 53342903–53342916
AAGCGTTGGGTGAC	LOC339240		14	chr17: 18327348–18327362
AGCTGGCCAGGTC	LOC541473		13	chr7: 72442106–72442119
TTCTGGCGCGGAG	MARCKSL1		13	chr1: 32801817–32801830
CCCGTTTCCGGTGG	MLX	TOP-LIKE	14	chr17: 40719116–40719130
ACTTCCGGTTCCCG	MTRF1L	TOP-LIKE	14	chr6: 153323846–153323860
CGCGTCGCTAGC	NDUF55		12	chr1: 39492008–39492020
AGTGTGAGGGGC	NOMO1		12	chr16: 14927651–14927663

Table S1. Cont.

Snatched sequence, 5' to 3'	Official gene symbol, human	5'-UTR type	Length, nt	Genomic coordinates
CAGACCGCCGAGG	PCBP1		13	chr2: 70314607–70314620
CCCAGACCAGC	PCBP2		11	chr12: 53845885–53845896
TTTTTCATTCCC	PDGFB	TOP-LIKE	12	chr22: 39640871–39640883
GTTTTGCAGACGC	PPIA		13	chr7: 44836279–44836292
GTTTTGCAGACGC	PPIA		12	chr7: 44836279–44836291
GTTTTGCAGACGC	PPIA		13	chr7: 44836279–44836292
GTTTTGCAGACGC	PPIA		13	chr7: 44836279–44836292
GTTTTGCAGACGC	PPIA		13	chr7: 44836279–44836292
ATCGCAGAGTGGAGC	PRKAR1A		15	chr17: 66508590–66508605
CTCTTCCGCGCC	RABGGTA	TOP	13	chr14: 24740820–24740833
AGTCGGACGGC	RAN		12	chr12: 131356640–131356652
TTTTTCCTCAGC	RBBP5	TOP-LIKE	12	chr1: 205074295–205074307
CTCGTTGCGCAGT	RBMX		13	chrX: 135962871–135962884
GCGGCCGCACATGTGTTTC	RCC2		19	chr1: 17766206–17766225
TTAGCGACTATTGC	REXO2		14	chr11: 114310183–114310197
CTCTTCCCTTC	RPL10	TOP	12	chrX: 153626717–153626729
CTCTTCCGGTTAGC	RPL10A	TOP	16	chr6: 35436180–35436196
CTTCCGCTCGGC	RPL13	TOP	13	chr16: 89627089–89627102
CTTCCGACCTGGC	RPL18	TOP	15	chr19: 49122418–49122433
TTCTCTTCGGC	RPL21	TOP-LIKE	12	chr13: 27825691–27825703
TCTTTTCGCCATC	RPL24	TOP-LIKE	13	chr3: 101405546–101405559
CTTTTCGCCATC	RPL24	TOP	12	chr3: 101405546–101405558
CTTTTTCGTGGGC	RPL27A	TOP	15	chr11: 8704338–8704353
CTCTTCCGTCTC	RPL28	TOP	13	chr19: 55897299–55897312
TCTTACGCCATC	RPL35A	TOP-LIKE	13	chr3: 197677064–197677077
CTTCTGGTCTCGGC	RPL37	TOP	15	chr5: 40835299–40835314
CTTCTGGGCTCAC	RPL37A	TOP	14	chr2: 217363566–217363580
CTTTCCGCCCGC	RPL39L	TOP	13	chr3: 186857066–186857079
CTCTTCCGCGCG	RPL8	TOP	13	chr8: 146017762–146017775
CTCTCGCCAGGC	RPLP0	TOP	12	chr12: 120638898–120638910
CTCTCGCCAGGC	RPLP0	TOP	12	chr12: 120638898–120638910
TCTCTGCCAGGC	RPLP0	TOP-LIKE	13	chr12: 120638898–120638911
CTCTCGCCAGGC	RPLP0	TOP	12	chr12: 120638898–120638910
CCCTTTCCTCAGC	RPLP1	TOP	13	chr15: 69745157–69745170
CTCTTCCCTGCC	RPS12	TOP	13	chr6: 133135707–133135720
GACCACAGGCTGGC	RPS18P9		14	chr6: 149915630–149915644
CTCTTTCGCTCAGG	RPS23	TOP	14	chr5: 81574159–81574173
TCCTTTTGTCCG	RPS25	TOP-LIKE	13	chr11: 118889042–118889055
CTTTTACCTC	RPS29	TOP	10	chr14: 50053084–50053094
TCTGACCAGCACC	RPS3A		13	chr4: 152020791–152020804
CTCTTTCCTTGC	RPS4X	TOP	12	chrX: 71497056–71497068
CTCTTTCAGCCAGC	RPS8	TOP	15	chr1: 45241245–45241260
CTCTTCCGCGCC	SCAF8	TOP	13	chr6: 155054995–155055008
ACAGAGTCCGGC	SFN		12	chr1: 27189639–27189651
ATTCAACATGGAGGC	SKA2		15	chr17: 57232516–57232531
AGGCCGGAAGCGC	SLC39A1		13	chr1: 153936000–153936013
CTCGGCAGCCGC	SNX29		12	chr16: 12070598–12070610
AGGTCGCGCGC	SRP14		12	chr15: 40331317–40331329
ATCCATGGCTTCCG	SRPRB		14	chr3: 133524688–133524702
TTTACGTGTC	STRAP		11	chr12: 16035316–16035327
AGCCGCTGCGCCCGAG	THBS1		16	chr15: 39873279–39873295
GCTAGGCCGGC	TOR3A		11	chr1: 179051221–179051232
AGTAGCAGCAGCG	TPRXL		13	chr3: 14106237–14106250
CTCTGCCACTCTCG	TTC28		14	chr22: 28398515–28398529
ACTGCGAAGCCAGC	UCP2		15	chr11: 73693873–73693888
CTTTTCCTCAGC	YARS	TOP	12	chr1: 33283514–33283526

Snatched sequences are shown, along with their length, their location in the human genome (as matched by BOWTIE), the gene from which the UTR was snatched, and their TOP or TOP-like status.