SUPPLEMENTAL INFORMATION

Identification and characterization of a novel PPARα-regulated and 7α-hydroxyl bile acid-preferring cytosolic sulfotransferase mL-STL (Sult2a8)

Lu Feng, [1,*] Yee-Lok Yuen, [1,*] Jian Xu, [1,*] Xing Liu,[*] Martin Yan-Chun Chan,[*] Kai Wang,[*] Wing-Ping Fong,[*] Wing-Tai Cheung, [2,§] and Susanna Sau-Tuen Lee[2,*]

School of Life Sciences, Faculty of Science,[*] and School of Biomedical Sciences, Faculty of Medicine,[§] The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR

[1]These authors contributed equally to the work.

[2]Co-corresponding authors

To whom correspondence should be addressed

Susanna Sau-Tuen Lee, PhD
School of Life Sciences,
Faculty of Science,
The Chinese University of Hong Kong,
Shatin, New Territories,
Hong Kong SAR
Phone: (852) 3943-6333
Fax number: (852) 2603-5646
E-mail: lee2022@cuhk.edu.hk

Wing-Tai Cheung, PhD
School of Biomedical Sciences,
Faculty of Medicine,
The Chinese University of Hong Kong,
Shatin, New Territories,
Hong Kong SAR
Phone: (852) 3943-6104
E-mail: wtcheung@cuhk.edu.hk

Running title: PPARα regulates bile acid sulfonation during fasting

SUPPLEMENTARY METHODS

Mass spectrometry analysis

The column-purified recombinant His-mL-STL and preparative SDS-PAGE gel semi-purified recombinant mL-STL proteins were confirmed by mass spectrometry analysis. The proteins were first resolved in a 14% SDS-PAGE gel and stained with Coomassie blue R for 1 h. After the gel was destained, the bands corresponding to the His-mL-STL and mL-STL proteins were excised and punched into small discs using a Monoject 202 blunt-end needle (Covidien, Mansfield, MA; Cat. No. 8881-202314). The gel discs were equilibrated with 25 mM NH_4HCO_3 for 1 h with vortexing at room temperature. They were then dehydrated and destained by three consecutives washes each with 50% acetonitrile (ACN)/25 mM ammonium bicarbonate (NH_4HCO_3) for 10 min followed by a final wash with of 100% ACN for 10 min. The gel was dried completely by a CentriVap vacuum concentrator (Labconco, Kansas City, MO) and then the proteins in the gel discs were reduced by 10 mM DTT/25 mM NH_4HCO_3 at 60^{o}C for 30 min. After that the proteins were alkylated in 55 mM iodoacetamide/25 mM NH_4HCO_3 for 30 min in dark and the gel discs were then washed, dehydrated, dried, and digested with 400 ng trypsin/20 µl of 25 mM NH_4HCO_3 (Promega, Madison, WI; Cat. No. V5111) at 37^{o}C overnight. The digested peptides were extracted with 25 mM NH_4HCO_3, 5% trifluoroacetic acid (TFA)/50% ACN followed by 100% ACN for 10 min in a Tru-sweep^{TM} ultrasonic cleaner (Crest, Trenton, NJ). The extracted peptides were dried using a speed vacuum and resuspended in 10 µl of 0.1% TFA. The peptide solution was purified with a ZipTip C_{18} column (Millipore, Billerica) pre-wetted with 50% ACN/0.1% TFA and equilibrated with 0.1% TFA. The peptides were then loaded onto the pre-washed ZipTip C_{18} column and it was washed with 0.1% TFA. Finally, the purified peptides were eluted in 3 µl of 50% ACN/0.1% TFA and stored frozen at -80°C. Each peptide sample (0.5 µl) was spotted on a MALDI plate and was allowed to air dry. The sample spot was covered with 0.5 µl of 10 mg/ml matrix (α-cyano-4-hydroxy-cinnamic acid) and the sample was analyzed by an Applied Biosystems 4700 proteomics analyzer. The peptide

mass fingerprint spectrum of each peptide spot was generated automatically by plotting the % intensity of positively charged ions against mass-to-charge (m/z) ratio of the ions in the range of 500 - 4000. Peptide analysis was performed using data-dependent acquisition of one MS scan followed by MS/MS scans of the ten most abundant ions in each MS scan (1). Data from peptide mass fingerprinting was processed by the Mascot software using the GPS Explorer^{TM} Workstation. The database searching was performed using the NCBI non-redundant database with no missed cleavage allowed and peptide tolerance of 100 ppm and MS/MS tolerance of 0.5 Da (2).

SUPPLEMENTARY REFERENCES

1. Tong, W. Y., Y. M. Liang, V. Tam, H. K. Yip, Y. T. Kao, K. M. Cheung, K. W. Yeung, and Y. W. Lam. 2010. Biochemical characterization of the cell-biomaterial interface by quantitative proteomics. *Mol. Cell. Proteomics.* **9**: 2089-2098.

2. Poetsch, A., D. Schlusener, C. Florizone, L. Eltis, C. Menzel, M. Rogner, K. Steinert, and U. Roth. 2008. Improved identification of membrane proteins by MALDI-TOF MS/MS using vacuum sublimated matrix spots on an ultraphobic chip surface. *J. Biomol. Tech.* **19**: 129-138.

Supplementary Table S1. Primers used in this study

Name	Sequences (5’ - 3’)	Application
AP8	ACGACTCACTATAGGGCTTTTTTTTTTTTAA	FDD
ARP1	ACAATTTCACACAGGACGACTCCAAG	
5’-RACE primer R	TGGTGCATTTGTGGAAGTCTAAGAA	5’-RACE (5’/3’ RACE)
5’-RACE primer I	CTGAGGCCAATCTGATTAGCTCTGT	5’-RACE (GeneRacer)
3’-RACE primer S	AAATGCACCATGATGGCTCCACCCATA	3’-RACE (GeneRacer)
M13 reverse	GTCATAGCTGTTTCCTG	Sequencing of 5’- and 3’-RACE clones
M13 forward	CTGGCCGTCGTTTTAC	
mL-STL-SEQ110	CAATGAGGCAGAAACACT	
mL-STL-SEQ380	ATTGTCTGCTTGATTCTG	
mL-STL-SEQ638	CAGGGAAAGGAACAAGTG	
mL-STL-SEQ898	AAGGATGATATCTCAGTC	
mL-STL-SEQ6	CTTGGACCACTTCTGTAC	
mL-STL-SEQ7	CAAGCCATGAAGCAGTAG	
mL-STL-FP278-Pst I	*aaaa*ctgcag*c*ATGACAGATGAATTTCTGTGGA (Pst I)	Cloning of His-mL-STL-mpRSETA expression vector
mL-STL-RP1319-Hind III	*ccc*aagctt GTGCCAGGATTGAACTCAGA (Hind III)	
mL-STL-SEQ-FP379	GCCAATGAGCGCACACCA	Sequencing of His-mL-STL-mpRSETA expression vector
mL-STL-SEQ-RP658	GCCACATGCATGTTCGAA	
mL-STL-SEQ-RP1231	GCTTCCTTTCGGGCTTTG	
mL-STL-SEQ-FP3895(-26)	GCGCGTTGGCCGATTCAT	
mL-STL-FP278-Nde I	*gggaatttc*catATGACAGATGAATTTCTGTGGA (Nde I)	Cloning of mL-STL-mmpRSET expression vector
mL-STL-RP1319-Hind III	*ccc*aagctt GTGCCAGGATTGAACTCAGA (Hind III)	
T7	TAATACGACTCACTATAGGG	Sequencing of mL-STL-mmpRSET expression vector
mL-STL-SEQ-380	ATTGTCTGCTTGATTCTG	
mL-STL-SEQ-638	CAGGGAAAGGAACAAGTG	
mL-STL-SEQ-898	AAGGATGATATCTCAGTC	
mL-STL-SEQ5	GTATTGTGCTAGAGAAGT	
mL-STL-FP19	ACAGATTCTCCCGACCTTA	Northern blot analysis
mL-STL-RP1884	CTGAGGCCAATCTGATTAGC	

The flanking regions of the restriction enzyme sites are in lower cases and in italics.

A

ARP1

```
ACAATTTCAC ACAGGACGAC TCCAAGGATA AAGAATCTTG TTAACGACCA TGAGGATCTG AGTTCAATCC TGGCACCAAC 80
ATAGTTGAAG GAAAGAAGTA GTTCCTGAAA TTATTCTTAG ACTTCCACAA ATGCACCATG ATGGCTCCAC CCATAAAATT 160
TCTGTAATTA AAATATTTCT TGATGTCAAA TCCTTAAATA ATGAAATAAT TATACAAATT TGATTTCTGA GTGTTTCAGA 240
CATTTTAAAG AAAAGTAAAA TTCAGAAACA ATCATAACTC TGCTCTTGGA CCACTTCTGT ACACCTCTAT GTCTGGCCCT 320
GTGGGTCACA GGTGACTTTC AAATCTTAGT AAGGATTTTC CTAAAATATG ACCAGCTGTT GAAAATCCAG GTCATGAGCC 400
TACTCAGTTT TTATGGTTCT CTACTTATGG AAGTAGAAAT TTGTACAGTT TGATAAGAAA GAGATTCAGC TGTTTTTGAT 480
GTTTCTTAAC TTCTCCCACC CCTTTCCAGG GAAGCTTCAT GCCTGGACAA GCCATGAAGC AGTAGAGTGC TCCTCATACC 560
TTTGAAGTAG AAAAGATGCC TGCCTGTTTG TGTAACAGGC TGGGGAAGAT TCCAATACAG AGCTAATCAG ATTGGCCTCA 640
GAAAAAATAC TAACTTATTC GTTTTGTTCC TGTCTTTCAG TGTAGAAGAC TTCTGTATTT TTTAAAATAC AATTTTATTT 720
CTTTCAACGA ATTTAAAAAA AACACCTTTG GAACAACAAG AACAACAAAA GTATAATTAC TTCTTCTATT GCTTGCATTG 800
AAGAAATGCT TTAAAGTATC ATCATATTTA ATATTTCTCC ATCATTTCAC TTATAATCAA TAATGTCGGT AAACAAAAAT 880
TAAAAAAAAA AAAGCCCTAT AGTGAGTCGT                                                    910
```

AP8

B

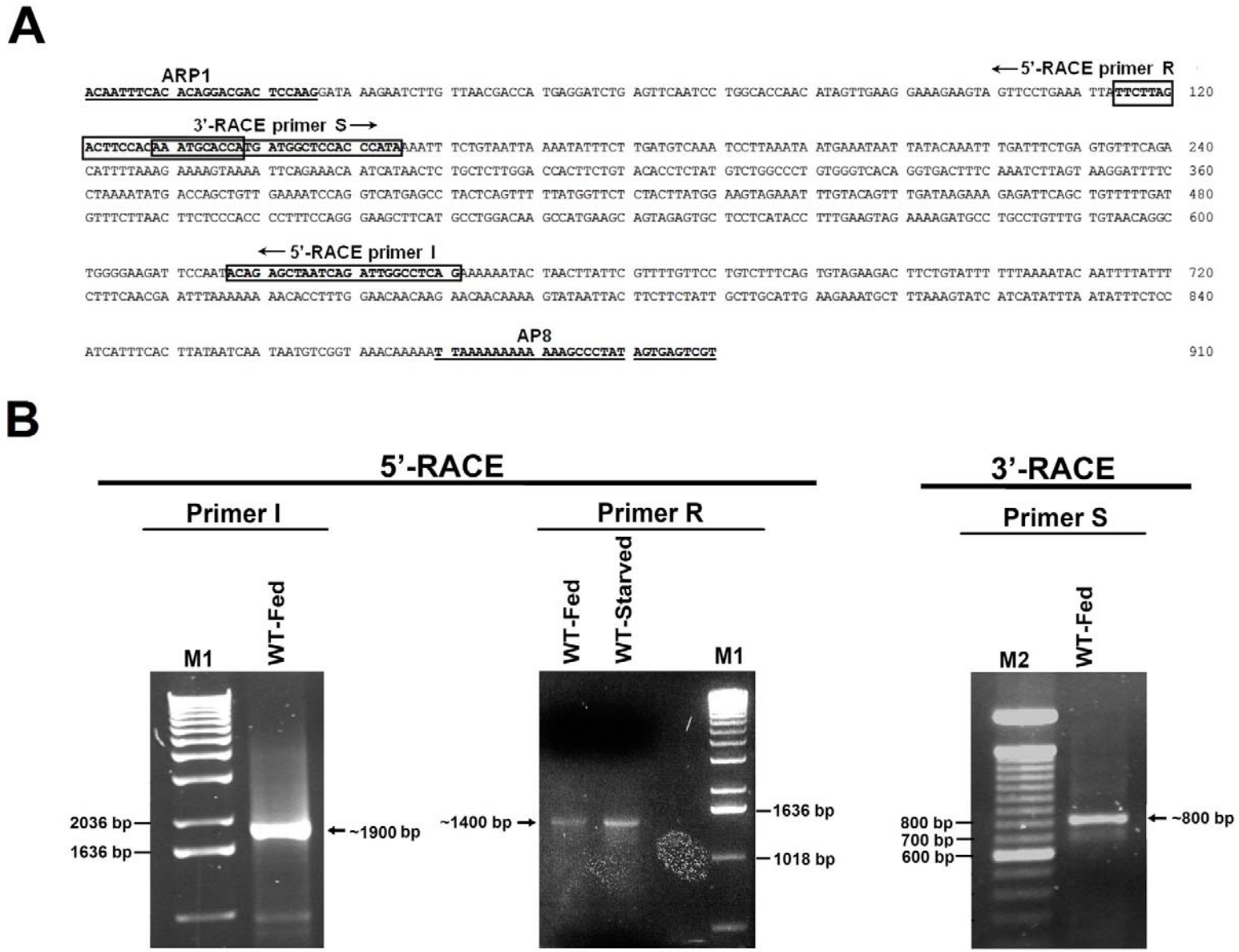
ref|NM_175250.5| *Mus musculus* RIKEN cDNA 2810007J24 gene (2810007J24Rik), transcript variant 1, mRNA

Length: 2147
Identities = 874/875 (99%)

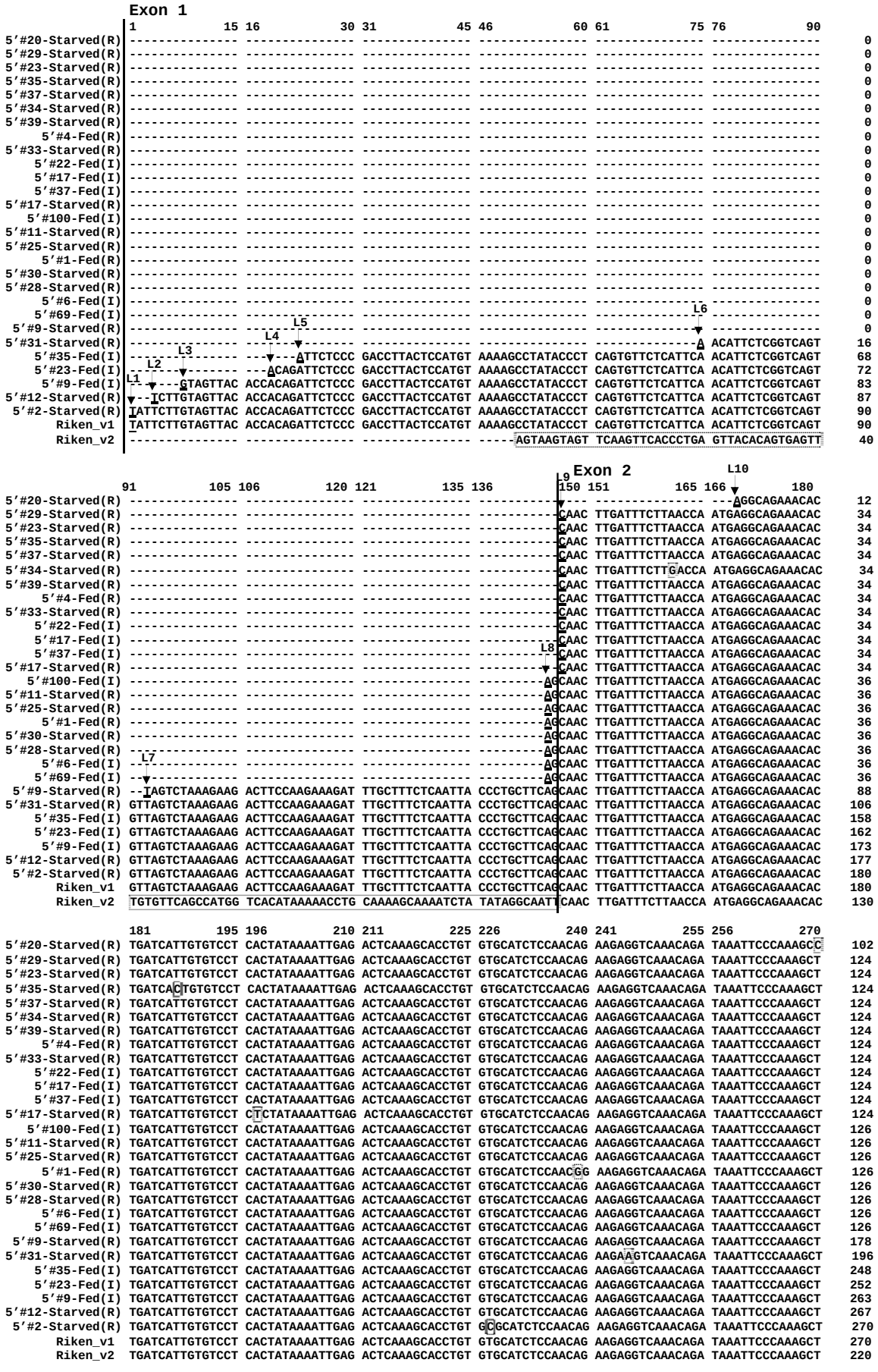
```
mL-STL  18    GACTCCAAGGATAAAGAATCTTGTTAACGACCATGAGGATCTGAGTTCAATCCTGGCACC  77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Riken   1261  GACTCCAAGGATAAAGAATCTTGTTAACGACCATGAGGATCTGAGTTCAATCCTGGCACC  1320

mL-STL  78    AACATAGTTGAAGGAAAGAAGTAGTTCCTGAAATTATTCTTAGACTTCCACAAATGCACC  137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Riken   1321  AACATAGTTGAAGGAAAGAAGTAGTTCCTGAAATTATTCTTAGACTTCCACAAATGCACC  1380

mL-STL  138   ATGATGGCTCCACCCATAAAATTTCTGTAATTAAAATATTTCTTGATGTCAAATCCTTAA  197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Riken   1381  ATGATGGCTCCACCCATAAAATTTCTGTAATTAAAATATTTCTTGATGTCAAATCCTTAA  1440

mL-STL  198   ATAATGAAATAATTATACAAATTTGATTTCTGAGTGTTTCAGACATTTTAAAGAAAAGTA  257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Riken   1441  ATAATGAAATAATTATACAAATTTGATTTCTGAGTGTTTCAGACATTTTAAAGAAAAGTA  1500

mL-STL  258   AAATTCAGAAACAATCATAACTCTGCTCTTGGACCACTTCTGTACACCTCTATGTCTGGC  317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Riken   1501  AAATTCAGAAACAATCATAACTCTGCTCTTGGACCACTTCTGTACACCTCTATGTCTGGC  1560

mL-STL  318   CCTGTGGGTCACAGGTGACTTTCAAATCTTAGTAAGGATTTTCCTAAAATATGACCAGCT  377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Riken   1561  CCTGTGGGTCACAGGTGACTTTCAAATCTTAGTAAGGATTTTCCTAAAATATGACCAGCT  1620

mL-STL  378   GTTGAAAATCCAGGTCATGAGCCTACTCAGTTTTTATGGTTCTCTACTTATGGAAGTAGA  437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Riken   1621  GTTGAAAATCCAGGTCATGAGCCTACTCAGTTTTTATGGTTCTCTACTTATGGAAGTAGA  1680

mL-STL  438   AATTTGTACAGTTTGATAAGAAAGAGATTCAGCTGTTTTTGATGTTTCTTAACTTCTCCC  497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Riken   1681  AATTTGTACAGTTTGATAAGAAAGAGATTCAGCTGTTTTTGATGTTTCTTAACTTCTCCC  1740

mL-STL  498   ACCCCTTTCCAGGGAAGCTTCATGCCTGGACAAGCCATGAAGCAGTAGAGTGCTCCTCAT  557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Riken   1741  ACCCCTTTCCAGGGAAGCTTCATGCCTGGACAAGCCATGAAGCAGTAGAGTGCTCCTCAT  1800

mL-STL  558   ACCTTTGAAGTAGAAAAGATGCCTGCCTGTTTGTGTAACAGGCTGGGGAAGATTCCAATA  617
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Riken   1801  ACCTTTGAAGTAGAAAAGATGCCTGCCTGTTTGTGTAACAGGCTGGGGAAGATTCCAATA  1860

mL-STL  618   CAGAGCTAATCAGATTGGCCTCAGAAAAAATACTAACTTATTCGTTTTGTTCCTGTCTTT  677
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Riken   1861  CAGAGCTAATCAGATTGGCCTCAGAAAAAATACTAACTTATTCGTTTTGTTCCTGTCTTT  1920

mL-STL  678   CAGTGTAGAAGACTTCTGTATTTTTTAAAATACAATTTTATTTCTTTCAACGAATTT-AA  736
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Riken   1921  CAGTGTAGAAGACTTCTGTATTTTTTAAAATACAATTTTATTTCTTTCAACGAATTTAAA  1980

mL-STL  737   AAAAAACACCTTTGGAACAACAAGAACAACAAAAGTATAATTACTTCTTCTATTGCTTGC  796
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Riken   1981  AAAAAACACCTTTGGAACAACAAGAACAACAAAAGTATAATTACTTCTTCTATTGCTTGC  2040

mL-STL  797   ATTGAAGAAATGCTTTAAAGTATCATCATATTTAATATTTCTCCATCATTTCACTTATAA  856
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Riken   2041  ATTGAAGAAATGCTTTAAAGTATCATCATATTTAATATTTCTCCATCATTTCACTTATAA  2100

mL-STL  857   TCAATAATGTCGGTAAACAAAAATTAAAAAAAAAA  891
              ||||||||||||||||||||||||||||||||||
Riken   2101  TCAATAATGTCGGTAAACAAAAATTAAAAAAAAAA  2135
```

Supplementary Figure S1

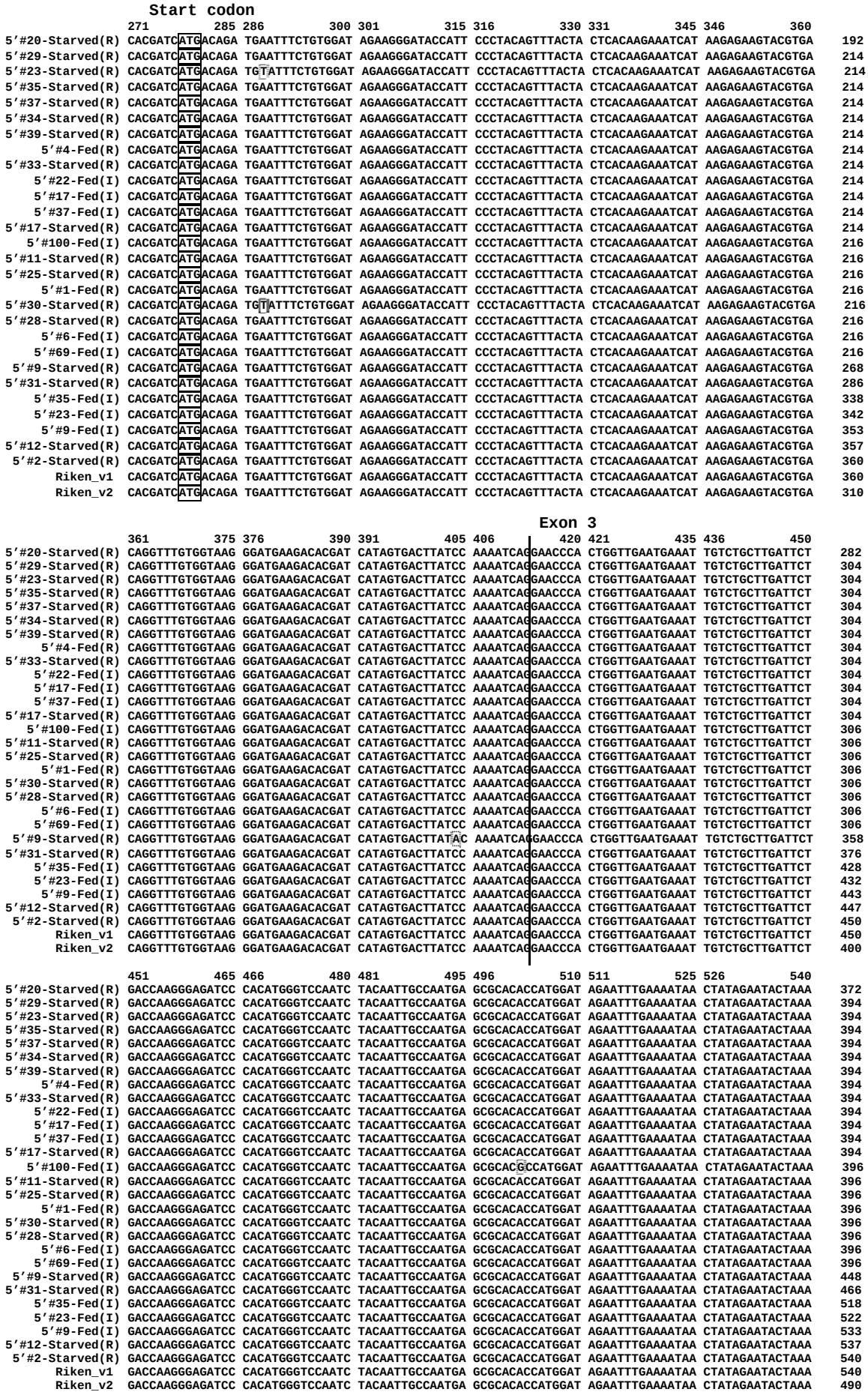
Supplementary Figure S1. DNA sequence of a 910 bp mL-STL FDD fragment and its alignment to 2810007J24Rik cDNA sequence. (A) DNA sequence of the partial mL-STL cDNA fragment isolated from the fluorescent differential display (FDD) analysis. The sequences of ARP1 and AP8 primers are underlined and bolded. (B) Alignment of the partial mL-STL cDNA sequence with the 3'-end of Riken cDNA 2810007J24. The mL-STL cDNA sequence from nucleotides 18 to 891 showed 99% similarity to the 3'-end of Riken cDNA sequence from nucleotides 1261 to 2135. An extra A nucleotide (boxed) was found in Riken cDNA sequence compared with the mL-STL.



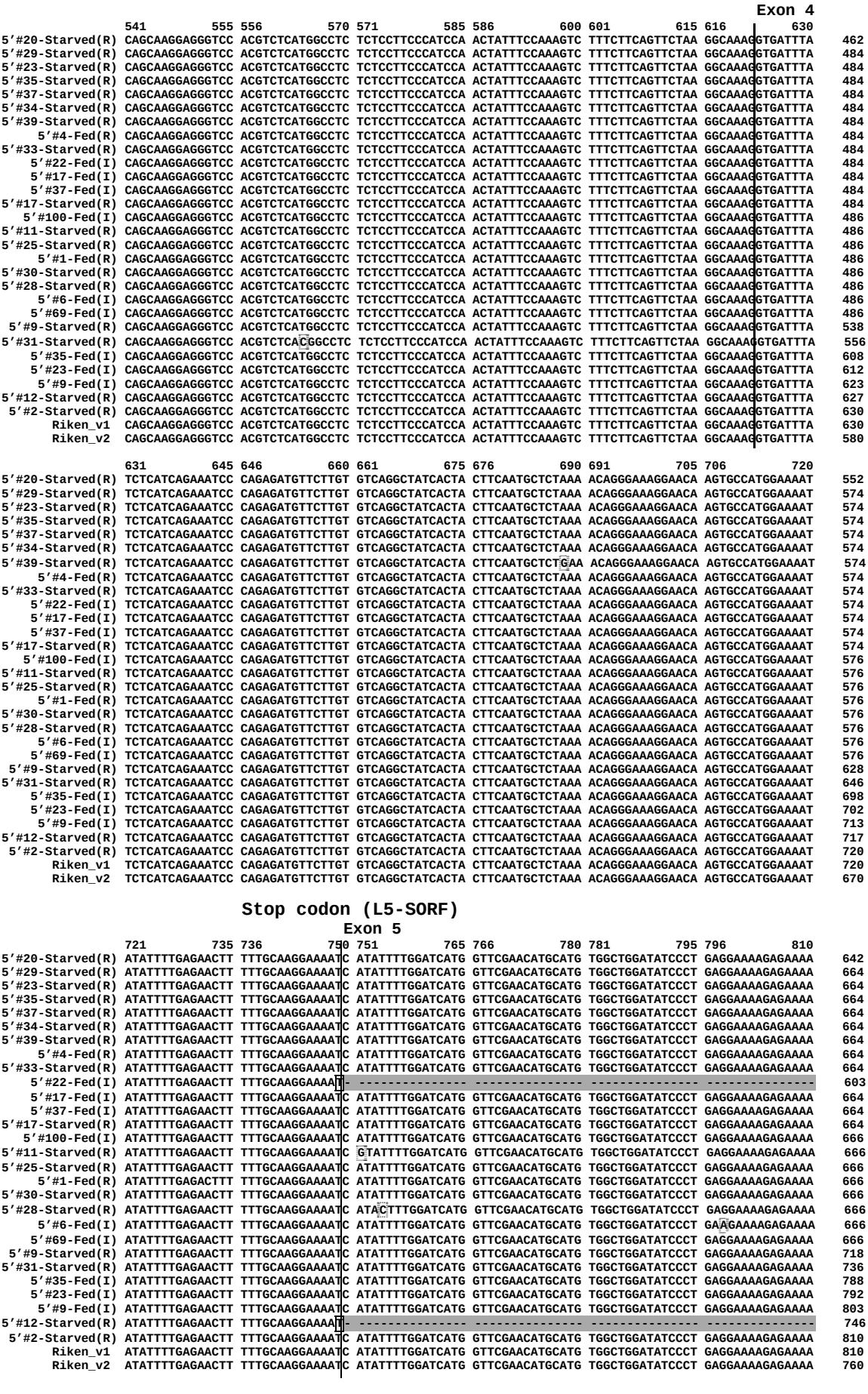
Supplementary Figure S2. 5'- and 3'-RACE primer design and amplicon products. (A) Location of the 5'- and 3'-mL-STL gene-specific rapid amplification of cDNA ends (RACE) primers. The sequences of 5'- and 3'-RACE primers are boxed and bolded, and the PCR amplification orientations are indicated. The arbitrary ARP1 and anchored AP8 primer sequences used for amplification of the partial mL-STL cDNA fragment in the FDD analysis are bolded and underlined. (B) 5'- and 3'-RACE mL-STL PCR products. Total RNA (5 μg/reaction) from the livers of a wild-type fed (WT-Fed) and a 72 h-starved (WT-Starved) mice was used in the RACE experiment.



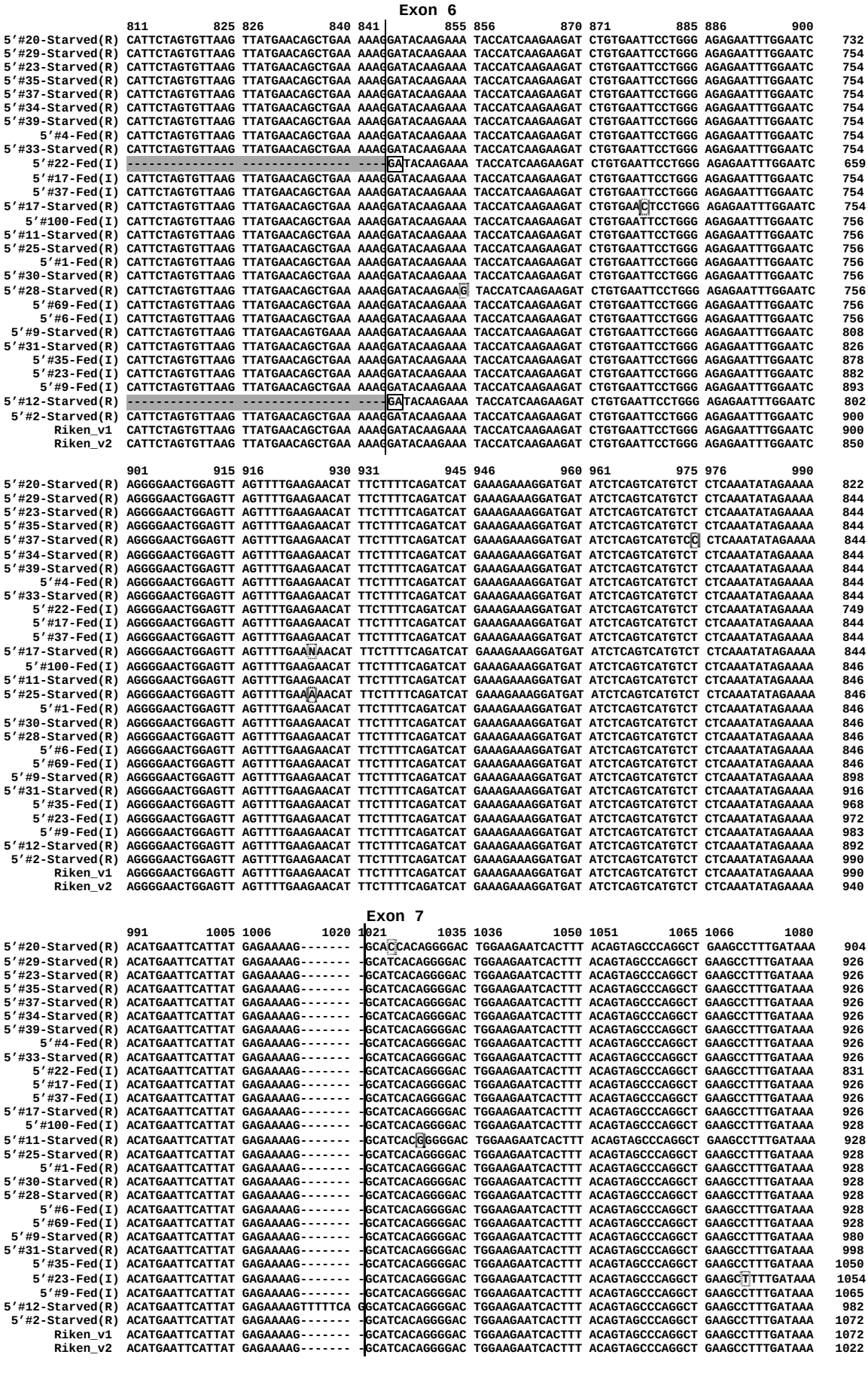
Supplementary Figure S3 (p. 1)



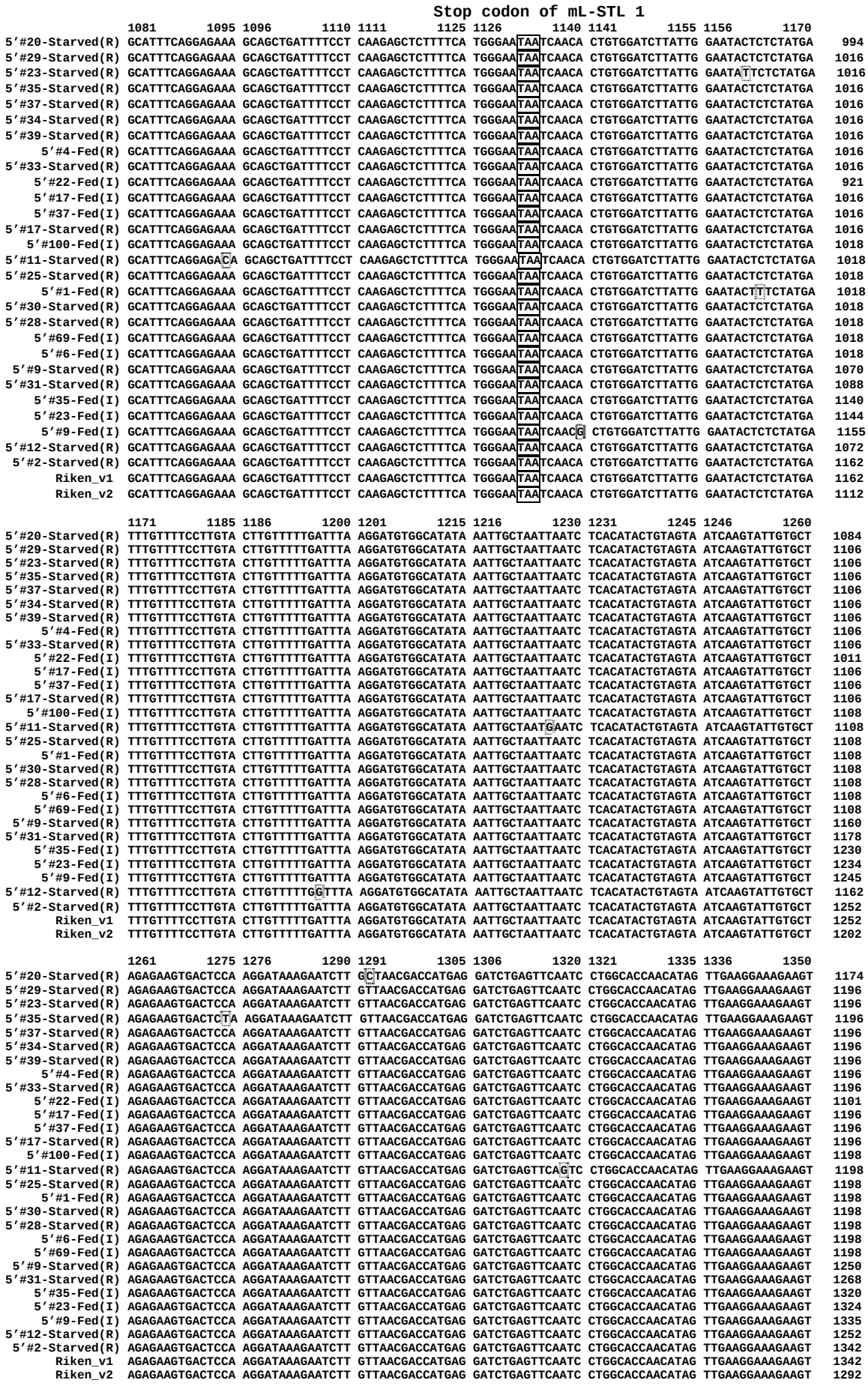
Supplementary Figure S3 (p. 2)



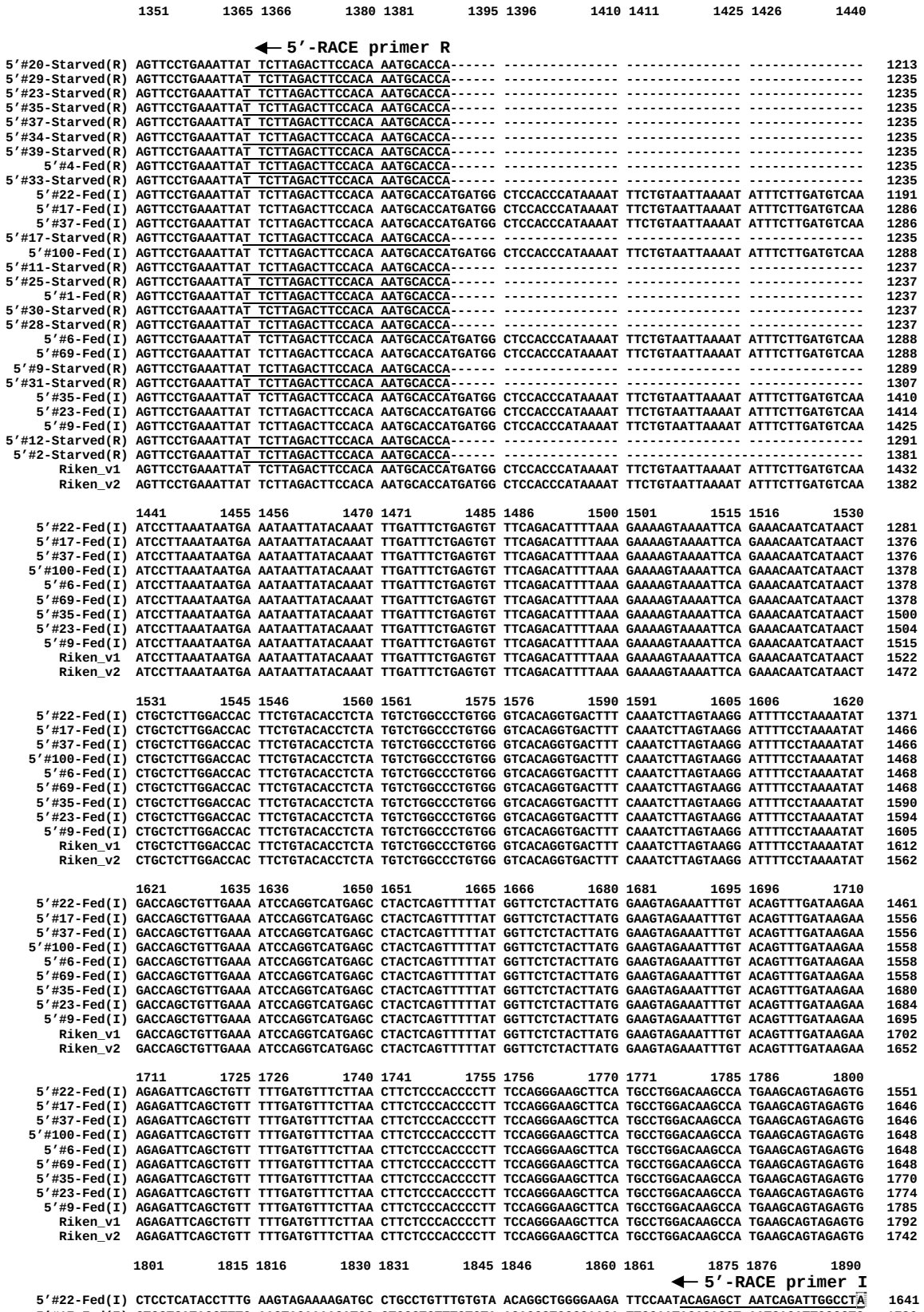
Supplementary Figure S3 (p. 3)



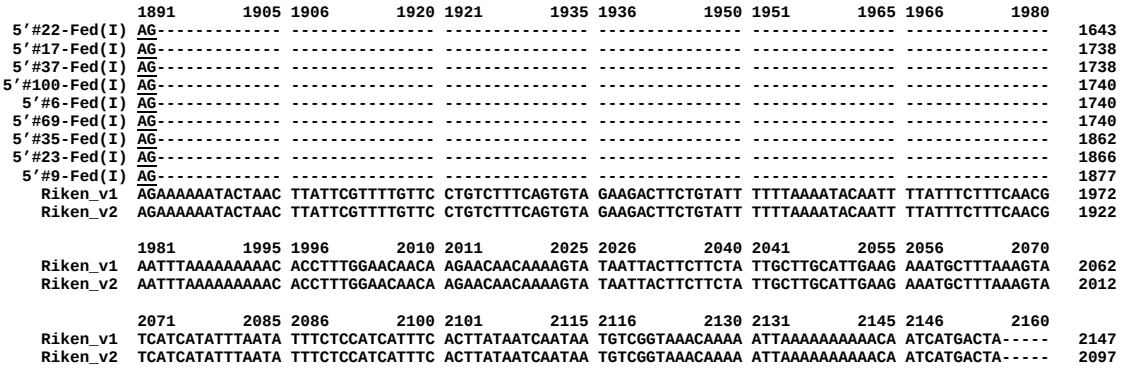
Supplementary Figure S3 (p. 4)



Supplementary Figure S3 (p. 5)



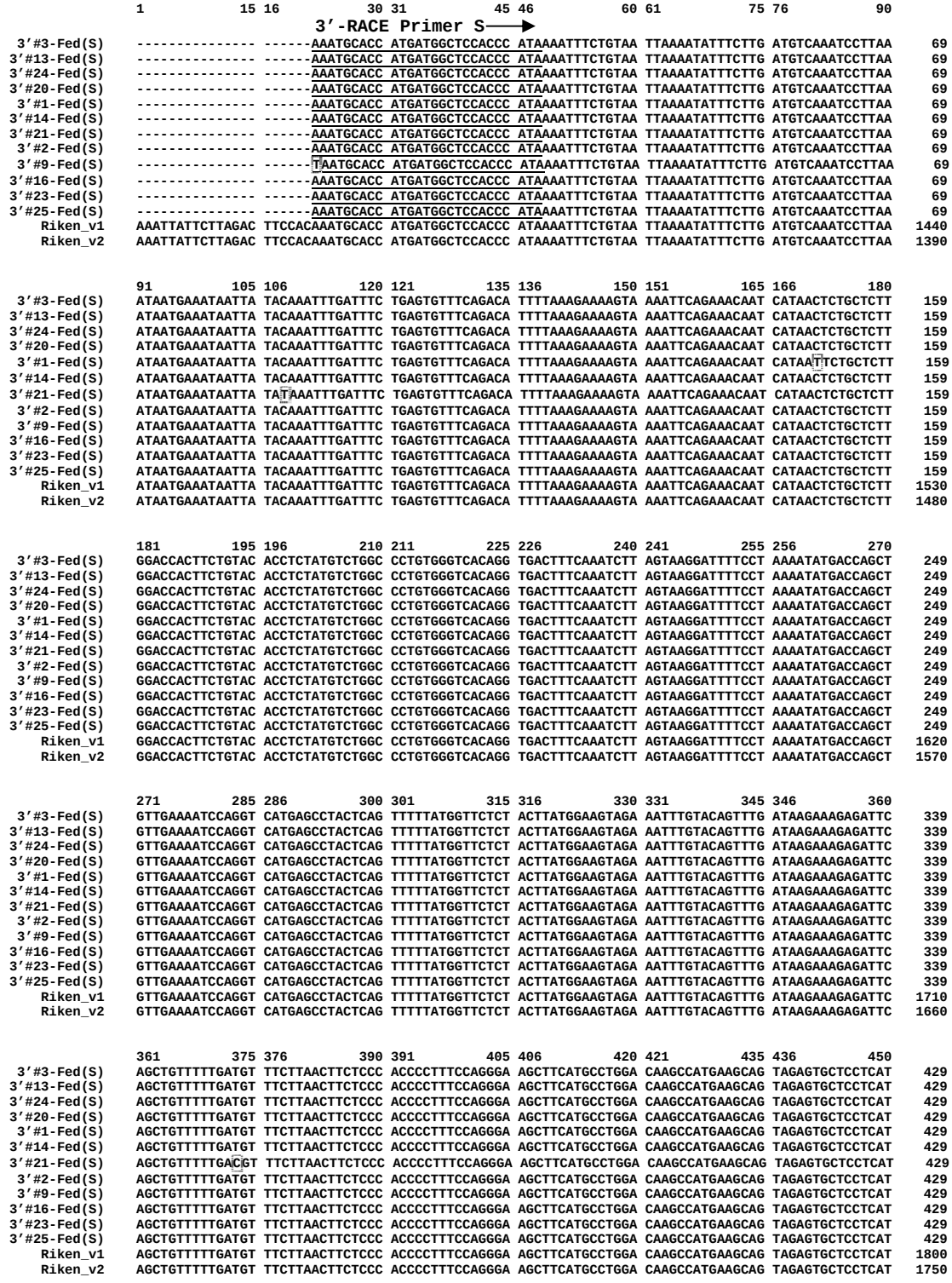
Supplementary Figure S3 (p. 6)



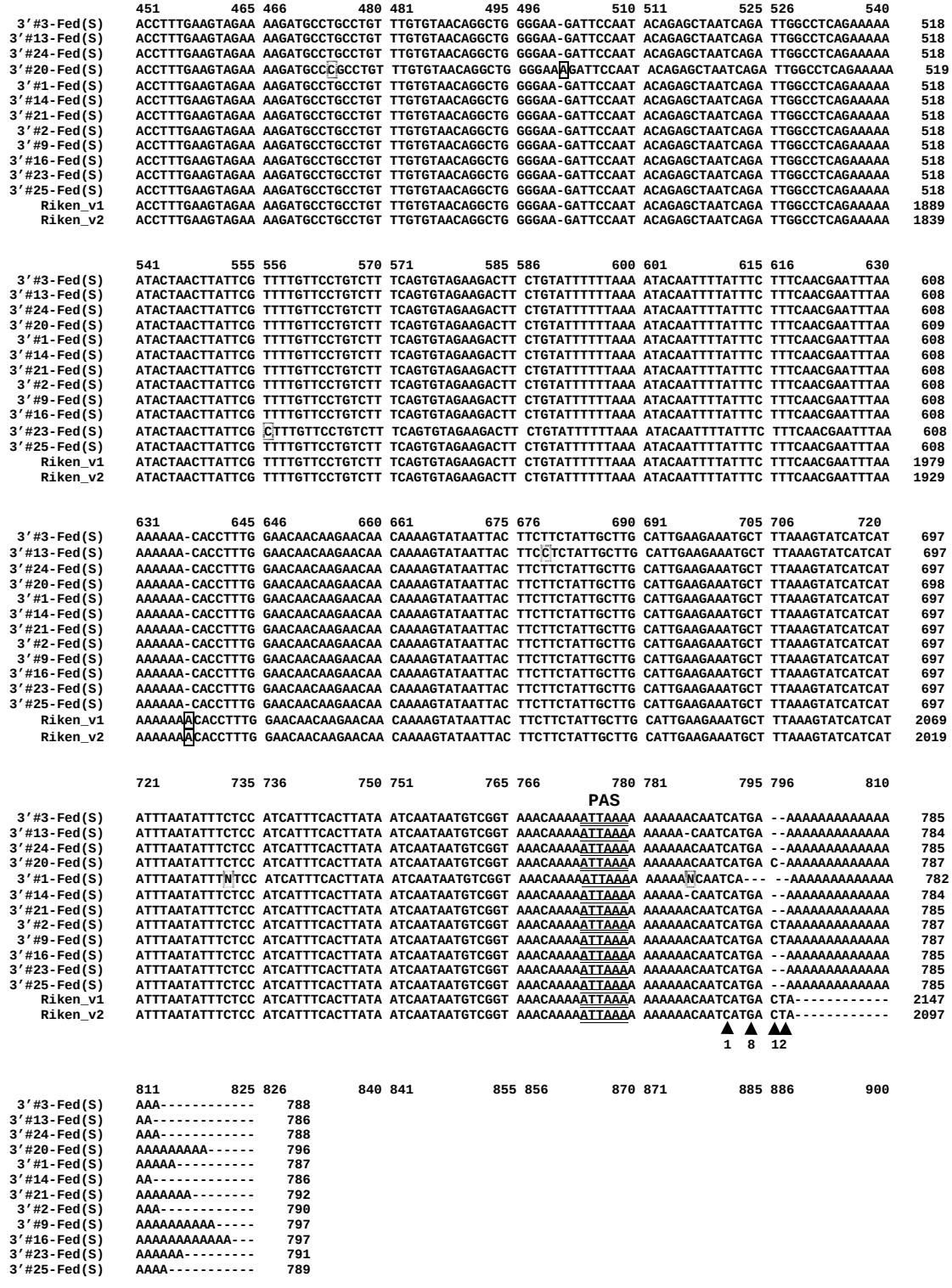
```
            1891       1905 1906       1920 1921       1935 1936       1950 1951       1965 1966       1980
 5′#22-Fed(I) AG------------ --------------- --------------- --------------- --------------- ---------------   1643
 5′#17-Fed(I) AG------------ --------------- --------------- --------------- --------------- ---------------   1738
 5′#37-Fed(I) AG------------ --------------- --------------- --------------- --------------- ---------------   1738
5′#100-Fed(I) AG------------ --------------- --------------- --------------- --------------- ---------------   1740
  5′#6-Fed(I) AG------------ --------------- --------------- --------------- --------------- ---------------   1740
 5′#69-Fed(I) AG------------ --------------- --------------- --------------- --------------- ---------------   1740
 5′#35-Fed(I) AG------------ --------------- --------------- --------------- --------------- ---------------   1862
 5′#23-Fed(I) AG------------ --------------- --------------- --------------- --------------- ---------------   1866
  5′#9-Fed(I) AG------------ --------------- --------------- --------------- --------------- ---------------   1877
    Riken_v1  AGAAAAAATACTAAC TTATTCGTTTTGTTC CTGTCTTTCAGTGTA GAAGACTTCTGTATT TTTTAAAATACAATT TTATTTCTTTCAACG   1972
    Riken_v2  AGAAAAAATACTAAC TTATTCGTTTTGTTC CTGTCTTTCAGTGTA GAAGACTTCTGTATT TTTTAAAATACAATT TTATTTCTTTCAACG   1922

            1981       1995 1996       2010 2011       2025 2026       2040 2041       2055 2056       2070
    Riken_v1  AATTTAAAAAAAAAC ACCTTTGGAACAACA AGAACAACAAAAGTA TAATTACTTCTTCTA TTGCTTGCATTGAAG AAATGCTTTAAAGTA   2062
    Riken_v2  AATTTAAAAAAAAAC ACCTTTGGAACAACA AGAACAACAAAAGTA TAATTACTTCTTCTA TTGCTTGCATTGAAG AAATGCTTTAAAGTA   2012

            2071       2085 2086       2100 2101       2115 2116       2130 2131       2145 2146       2160
    Riken_v1  TCATCATATTTAATA TTTCTCCATCATTTC ACTTATAATCAATAA TGTCGGTAAACAAAA ATTAAAAAAAAAACA ATCATGACTA-----   2147
    Riken_v2  TCATCATATTTAATA TTTCTCCATCATTTC ACTTATAATCAATAA TGTCGGTAAACAAAA ATTAAAAAAAAAACA ATCATGACTA-----   2097
```

Supplementary Figure S3 (p. 7)

Supplementary Figure S3. Alignment of 28 mL-STL 5’-RACE clone nucleotide sequences. All 5’-RACE clones showed identical nucleotide sequence except that a 95 bp fragment (boxed and shaded) was absent in two clones 5’#22-Fed(I) and 5’#12-Starved(R), suggesting the presence of alternative splicing variants. Ten (L1 – L10) different 5’-termini are indicated by downward arrows from nucleotides 1 to 169 relative to clone 5’#2-Starved(R). An extra short fragment (TTTTTCAG) was found at the end of exon 6 in clone 5’#12-Starved(R). The mismatched nucleotides are indicated by boxed nucleotides throughout the sequences.

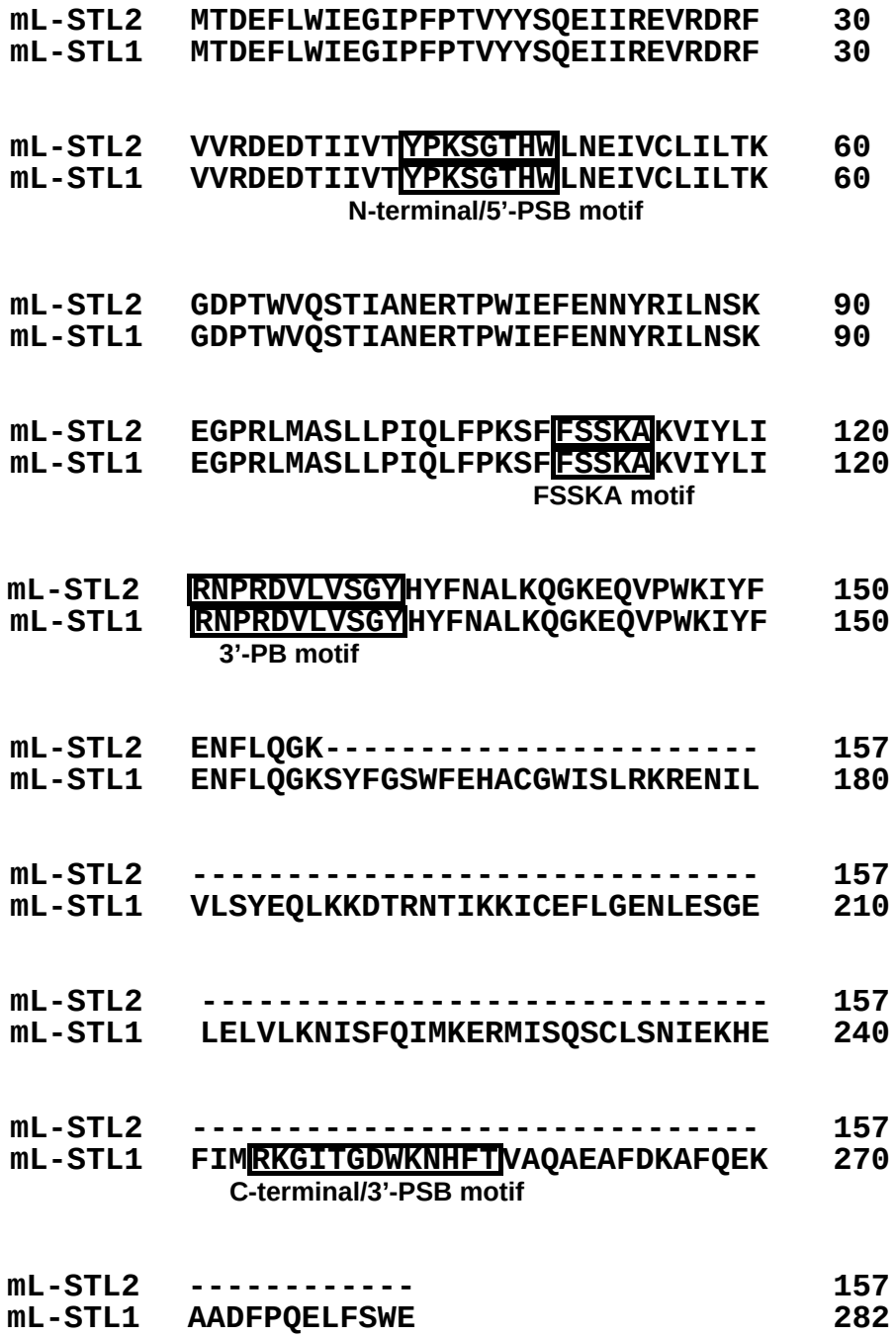


Supplementary Figure S4 (p. 1)

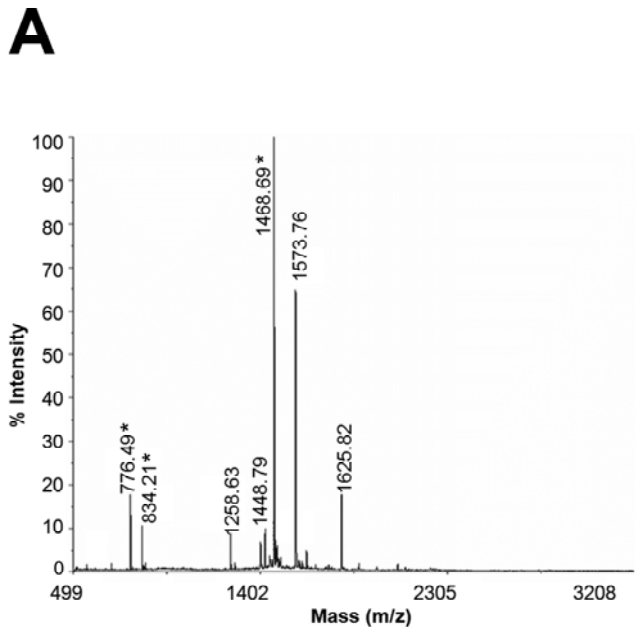


Supplementary Figure S4 (p. 2)

Supplementary Figure S4. Alignment of 12 mL-STL 3'-RACE clone nucleotide sequences. All 3'-RACE clones showed identical nucleotide sequence except the presence of some mismatched nucleotides as indicated by boxed nucleotides throughout the sequences. The longest 3'-RACE cDNA was 797 bp and found in clones 3'#9-Fed(S) and 3'#16-Fed(S), while the shortest 786 bp in clones 3'#13-Fed(S) and 3'#14-Fed(S). The solid triangles show the four different polyadenylation cleavage site (PACS) variants. The polyadenylation signal (PAS) ATTAAA is double-underlined.

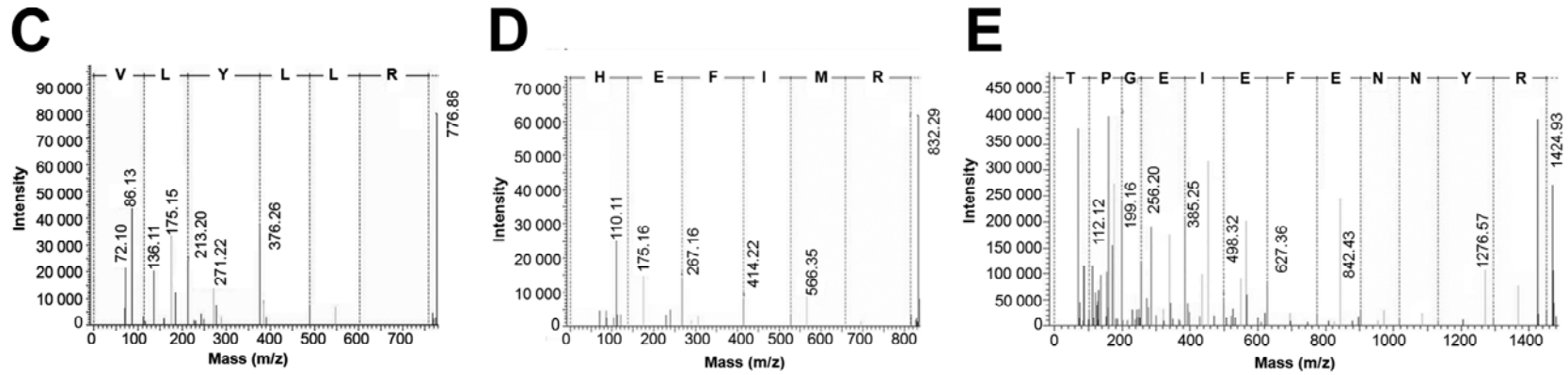


Supplementary Figure S5. Comparison of mL-STL1 and mL-STL2 amino acid sequences. The amino acid residues of mL-STL1 and mL-STL2 are 282 and 157, respectively. The N-terminal conserved sequence [N-terminal/5'-phosphosulfate binding (PSB) motif], SULT2 FSSKA signature motif sequence (FSSKA motif), 3'-phosphate binding (PB) motif, and C-terminal conserved sequence (C-terminal/3'-PSB motif) are boxed. The C-terminal/3'-PSB motif is missing in the mL-STL2 isoform.

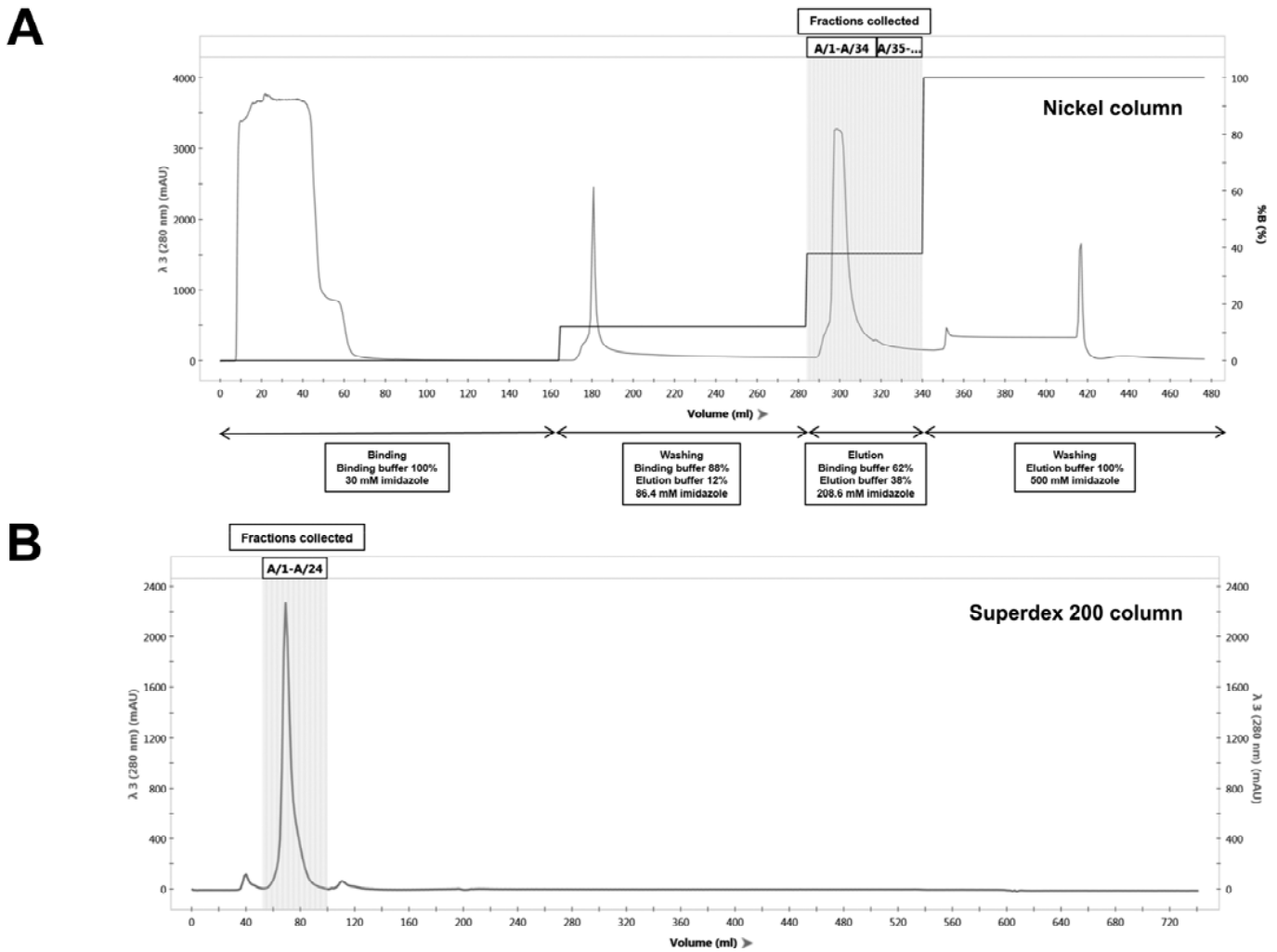


B

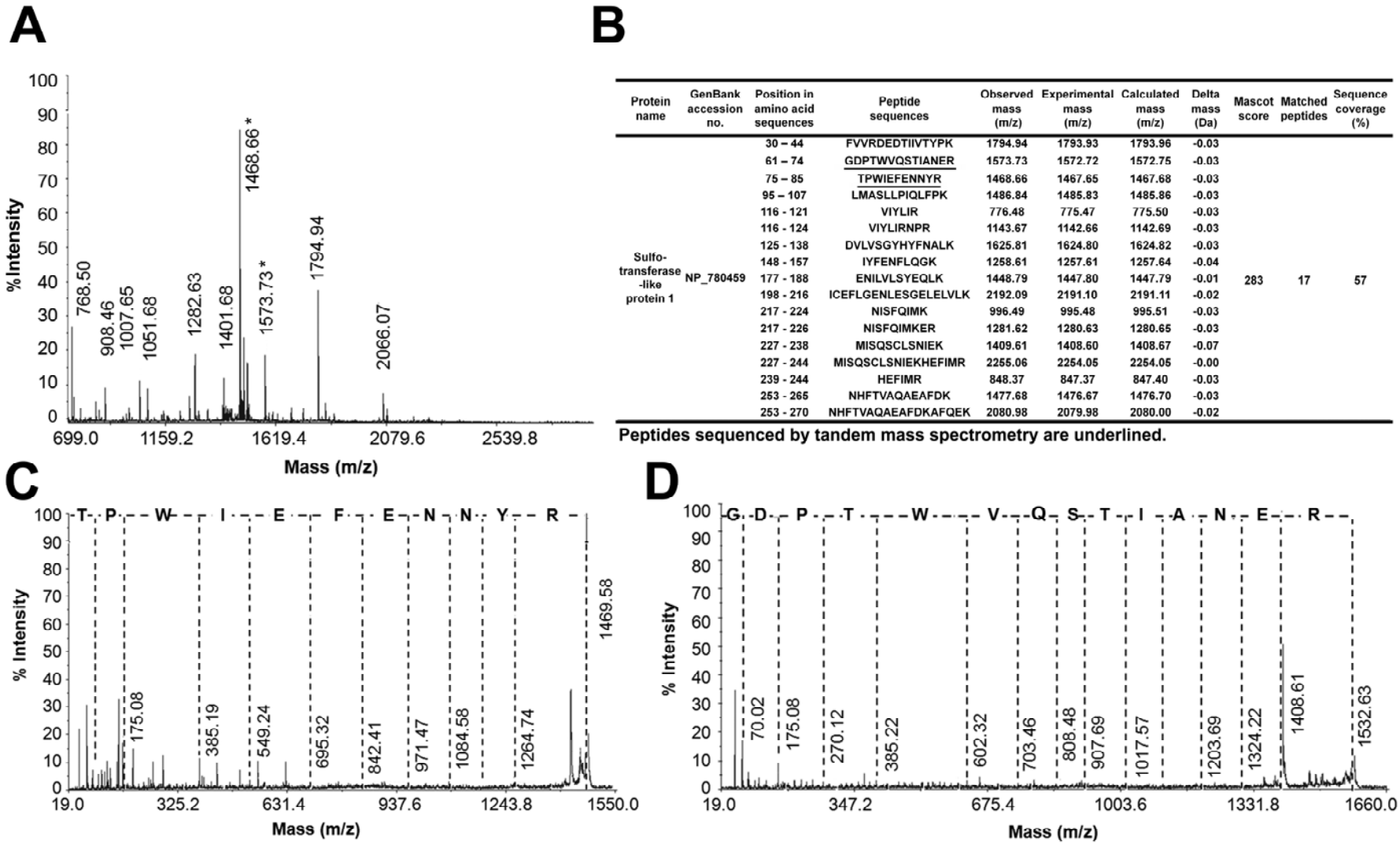
Protein name	GenBank accession no.	Position in amino acid sequences	Peptide sequences	Observed mass (m/z)	Experimental mass (m/z)	Calculated mass (m/z)	Delta mass (Da)	Mascot score	Matched peptides	Sequence coverage (%)
Sulfo-transferase-like protein 1	NP_780459	30 – 33	FVVR	520.32	519.31	519.32	-0.01	593	10/23	34
		116 – 121	VIYLIR	776.49	775.48	775.50	-0.01			
		142 – 147	EQVPWK	786.42	785.41	785.41	0.01			
		239 – 244	HEFIMR	832.41	831.40	831.41	-0.01			
		148 – 157	IYFENFLQGK	1258.63	1257.63	1257.64	-0.01			
		177 – 188	ENILVLSYEQLK	1448.80	1447.79	1447.79	-0.01			
		75 – 85	TPWIEFENNYR	1468.69	1467.68	1467.68	0.00			
		253 – 265	NHFTVAQAEAFDK	1477.69	1476.68	1476.70	-0.02			
		61 - 74	GDPTWVQSTIANER	1573.76	1572.76	1572.75	0.00			
		125 - 138	DVLVSGYHYFNALK	1625.82	1624.82	1624.82	-0.01			



Supplementary Figure S6. Mass spectrum of trypsin-digested recombinant mL-STL protein. The protein band corresponding to the expected recombinant mL-STL protein was excised from the 14% SDS-PAGE, digesed with trypsin, and analyzed by MALDI-TOF peptide sequencing. The tryptic peptides of the excised protein bands were submitted for searching in NCBI database (http://www.matrixscience.com/search_form_ select.html). The peptides that matched with the theoretical predicted peptides of the mouse liver sulfotransferase-like protein 1 (NP_780459) are shown in the MS spectrum in (A) and summarized in a table in (B). The peptides that are indicated with an asterisk (*) in (A) or underlined in (B) were analyzed by MS/MS sequencing. The MS/MS spectra of doubly charged peptides at mass-to-charge (m/z) ratio of 776.49 (C), 834.21 (D), and 1468.69 (E) were sequenced as VLYLLR, HEFIMR, and TPGEIEFENNYR, respectively.



Supplementary Figure S7. Purification of recombinant His-mL-STL protein. A two-step column chromatographic method using a Bio-Rad NGC™ Discover 10 chromatography system was used. (A) The supernatant collected from crude cell lysate was subject to two 5-ml nickel affinity HiTrap columns coupled in series. The columns were charged with nickel ion and pre-equilibrated with a binding buffer (50 mM Tris-HCl, pH 8, 500 mM NaCl, and 30 mM imidazole). The samples were then loaded onto the column and 170 ml of binding buffer was passed through the column at a flow rate of 4 ml/min to remove unbound proteins. Then the column was washed with 120 ml of 86.4 mM imidazole in a mixed binding and elution buffer (50 mM Tris-HCl, pH 8, 500 mM NaCl, and 500 mM imidazole) to remove non-specific protein binding. The bound His-mL-STL proteins were eluted from the column with 28 ml of 208.6 mM imidazole in elution buffer. (B) The eluents containing the His-mL-STL proteins were pooled, concentrated to 3 ml and then loaded, at a flow rate of 1 ml/min, onto a HiLoad Superdex 200 prep grade size exclusion column pre-equilibrated with a size-exclusion buffer (50 mM Tris-HCl, pH 8, and 150 mM NaCl). The His-mL-STL proteins were eluted with 100 ml of the same buffer and the eluted fractions containing the His-mL-STL proteins were pooled and collected for sulfotransferase activity assays.



Protein name	GenBank accession no.	Position in amino acid sequences	Peptide sequences	Observed mass (m/z)	Experimental mass (m/z)	Calculated mass (m/z)	Delta mass (Da)	Mascot score	Matched peptides	Sequence coverage (%)
Sulfo-transferase-like protein 1	NP_780459	30 – 44	FVVRDEDTIIVTYPK	1794.94	1793.93	1793.96	-0.03	283	17	57
		61 – 74	GDPTWVQSTIANER	1573.73	1572.72	1572.75	-0.03			
		75 – 85	TPWIEFENNYR	1468.66	1467.65	1467.68	-0.03			
		95 – 107	LMASLLPIQLFPK	1486.84	1485.83	1485.86	-0.03			
		116 - 121	VIYLIR	776.48	775.47	775.50	-0.03			
		116 - 124	VIYLIRNPR	1143.67	1142.66	1142.69	-0.03			
		125 - 138	DVLVSGYHYFNALK	1625.81	1624.80	1624.82	-0.03			
		148 - 157	IYFENFLQGK	1258.61	1257.61	1257.64	-0.04			
		177 - 188	ENILVLSYEQLK	1448.79	1447.80	1447.79	-0.01			
		198 - 216	ICEFLGENLESGELELVLK	2192.09	2191.10	2191.11	-0.02			
		217 - 224	NISFQIMK	996.49	995.48	995.51	-0.03			
		217 - 226	NISFQIMKER	1281.62	1280.63	1280.65	-0.03			
		227 - 238	MISQSCLSNIEK	1409.61	1408.60	1408.67	-0.07			
		227 - 244	MISQSCLSNIEKHEFIMR	2255.06	2254.05	2254.05	-0.00			
		239 - 244	HEFIMR	848.37	847.37	847.40	-0.03			
		253 - 265	NHFTVAQAEAFDK	1477.68	1476.67	1476.70	-0.03			
		253 - 270	NHFTVAQAEAFDKAFQEK	2080.98	2079.98	2080.00	-0.02			

Supplementary Figure S8. Mass spectrum of trypsin-digested recombinant His-mL-STL protein. The protein band corresponding to the expected recombinant His-mL-STL protein was excised from the 14% SDS-PAGE, digesed with trypsin, and analyzed by MALDI-TOF peptide sequencing. The tryptic peptides of the excised protein bands were submitted for searching in NCBI datatase (http://www.matrixscience.com/search_form_ select.html). The peptides that matched with the theoretical predicted peptides of the mouse liver sulfotransferase-like protein 1 (NP_780459) are shown in the MS spectrum in (A) and summarized in a table in (B). The peptides that are indicated with an asterisk (*) in (A) or underlined in (B) were analyzed by MS/MS sequencing. The MS/MS spectra of doubly charged peptides at mass-to-charge (m/z) ratio of 1468.66 (C) and 1573.73 (D) were sequenced as TPWIEFENNYR and GDPTWVQSTIANER, respectively.