SUPPLEMENTAL INFORMATION

Redefining the sRNA Transcriptome in *Streptococcus pneumoniae* Serotype 2 Strain D39

Dhriti Sinha[1,2], Kurt Zimmer[3], Todd A. Cameron[2], Douglas B. Rusch[4], Malcolm E. Winkler[1#], and Nicholas R. De Lay[2,5#]

[1]Department of Biology, Indiana University Bloomington, Bloomington, IN 47405, USA

[2]Department of Microbiology and Molecular Genetics, McGovern Medical School, University of Texas Health Science Center, Houston, TX 77030, USA

[3]Indiana University School of Informatics, Computing and Engineering, Indiana University Bloomington, Bloomington, IN 47408, USA

[4]Center for Genomics and Bioinformatics, Indiana University Bloomington, Bloomington, IN 47405, USA

[5]MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, University of Texas Health Science Center, Houston, TX 77030, USA

Running title: Detection of New sRNAs in *S. pneumoniae* D39

Keywords: pneumococcus; small RNA identification; RNA-seq and dRNA-seq

[#]Correspondence should be sent to Nicholas R. De Lay (Tel: +1 (713) 500-6293; Email: nicholas.r.delay@uth.tmc.edu) and Malcolm E. Winkler (Tel: +1 (812) 856-1318; E-mail: winklerm@indiana.edu)

TABLE OF CONTENTS

TABLE S1: List showing the overlap between the sRNAs and riboswitches that were identified in *S. pneumoniae* D39V and D39W

sRNA ID in D39V	Coordinates in D39V	Flanking genes	sRNA ID in D39W	Coordinates in D39W	Identification in D39W
ccnC[a]	23967-24065 (+)	*spd_0024*, *spd_0025*	*ccnC*	23967-24065 (+)	dRNA-seq
scRNA	24632-24707 (+)	*spd_0026*	*scRNA*	24632-24707 (+)	
srf-01	29658-29873 (+)	*spd_0033*, *spd_0038*	ND	ND	ND
srf-02[a]	39980-40186 (+)	*spd_0047*, *spd_0048*	*spd_sr5*	39980-40082 (+)	sRNA-seq & dRNA-seq
srf-03[a]	41719-41886 (+)	*spd_0048*, *comA*	*spd_sr6*	41494-41559 (+)	dRNA-seq
srf-04[b]	13229-132308 (+)	*tmrU*, *spd_0128*	*spd_sr37*	131773-131842 (+)	dRNA-seq
srf-05[b]	149645 – 149877 (+)	*spd_0143*, *spd_0144*	*spd_sr14*	149223-149341 (+)	sRNA-seq
srf-06	150711-150914 (-)	*spd_0145*, *spd_0144*	ND	ND	ND
ccnE[a]	212734-212881 (+)	*spd_0221*, *spd_0222*	*ccnE*	212278 – 212425 (+)	sRNA-seq & dRNA-seq
ccnA[b]	231599 – 231691 (+)	*spd_0240*, *ruvB*	*ccnA*	231143 – 231235 (+)	dRNA-seq
ccnB[a]	231786-231883 (+)	*spd_0240*, *ruvB*	*ccnB*	231331 – 231427 (+)	dRNA-seq
srf-07[a]	232279-232354 (+)	*spd_0240*, *ruvB*	*spd_sr23*	231823 – 232091 (+)	sRNA-seq & dRNA-seq
ccnD[a]	234171-234264 (+)	*spd_0242*, *uppS*	*ccnD*	233715 – 233809 (+)	dRNA-seq
srf-08	282194-282479 (+)	*spd_0284*, *spd_0285*	ND	ND	ND
rnpB	344607-345007 (+)	*spd_0339*, *spd_0341*	*rnpB*	344150 – 344553 (+)	
srf-10	508697-508842 (+)	*spd_0500*, *bglG*	*spd_sr33*	508237 – 508335 (+)	sRNA-seq
srf-11[a]	587896-587989 (+)	*spd_0563*, *spd_0564*	*spd_sr42*	587439 – 587538 (+)	dRNA-seq
srf-12[d]	742022-742141	*spd_0726*, *spd_0729*	ND	ND	ND
ssrA[a]	781595-781939 (+)	*spd_0768*, *spd_0770*	*spd_sr46*	781128 – 781485 (+)	sRNA-seq & dRNA-seq
srf-13[a,c]	826260 – 826587 (+)	*spd_0808*, *cad*	*spd_sr49*	825802 – 826129 (+)	sRNA-seq & dRNA-seq
srf-16	1037649-1037755 (-)	*spd_0914*, *spd_0915*	ND	ND	ND
srf-17[b]	1051910-1052049 (-)	*spd_0899*, *asd*	*spd_sr17*	912572 – 912715 (+)	sRNA-seq & dRNA-seq

srf-18[c]	1079561-1079658 (-)	*lacD*, *lacT*	*spd_sr60*	1079135 – 1079199 (-)	dRNA-seq
srf-19	1170746 – 1170923 (+)	*spd_1136*, *spd_1137*	*spd_sr63*	1170288 – 1170385 (+)	sRNA-seq & dRNA-seq
srf-21[a]	1528520-1528643 (-)	*recG*, *spd_1506*	*spd_sr83*	1528044 – 1528186 (-)	sRNA-seq & dRNA-seq
srf-22	1548804-1549088 (-)	*spd_1527*, *spd_1526*	ND	ND	ND
ssrS[a]	1598326 – 1598522 (+)	*spd_1580*, *spd_1581*	*spd_sr85*	1597869 – 1598142 (+)	sRNA-seq & dRNA-seq
srf-24	1873736-1873781 (-)	*mutS*, *spd_1902*	*spd_sr105*	1873276 - 1873324 (-)	dRNA-seq
srf-25[a]	1892857-1893007 (-)	*spd_1924*, *spd_1923*	*spd_sr106*	1892400 – 1892550 (-)	sRNA-seq & dRNA-seq
srf-26	1949385-1949547(+)	*spd_1974*, *arcA*	ND	ND	ND
srf-27[d]	1973172-1973570 (-)	*adcA*, *spd_1996*	*spd_sr110*	1973000 - 1973113 (-)	dRNA-seq
srf-28	1973509-1973913 (-)	*adcA*, *spd_1996*	*spd_sr112*	1973342 - 1973456 (-)	dRNA-seq
srf-29	2008242-2008356 (-)	*spd_2029*, *cbpD*	ND	ND	ND
srf-30[a,c]	2020587-2020685 (-)	*spd_2043*, *rpsB*	*spd_sr116*	2020112-2020228 (-)	dRNA-seq
riboswitch-1[a]	174291-174470 (-)	*ruvA*, *ribD*	*spd-sr22*	173841 – 174014 (-)	sRNA-seq & dRNA-seq
riboswitch-2	376842-377047 (+)	*spd_0371*, *spd_0372*	*spd_sr28*	376371 – 376530 (+)	sRNA-seq & dRNA-seq
riboswitch-3	381487-381662 (-)	*spd_0376*, *serS*	*spd_sr29*	381067 - 381205 (-)	sRNA-seq & dRNA-seq
riboswitch-4	441436-441663 (+)		ND	ND	ND
riboswitch-5[a]	497356-497561 (+)	*spd_0490*, *spd_0491*	*spd_sr32*	496899 – 497104 (+)	dRNA-seq
riboswitch-6	513346-513544 (+)	*bglA-2*, *pheS*	*spd_sr34*	512889 – 513087 (+)	sRNA-seq & dRNA-seq
riboswitch-7	632244-632347 (+)	*spd*_0436, *spd*_0437	ND	ND	ND
riboswitch-8[a]	644329-644497 (+)	*lctO*, *spd_0622*	*spd_sr43*	643872 – 644040 (+)	dRNA-seq
riboswitch-9	647211-647381 (+)	*thiE-1*, *spd_0625*	*spd_sr44*	646754 – 646926 (+)	sRNA-seq & dRNA-seq
riboswitch-10[d]	653914-654012 (-)	*cspR*, *ribU*	ND	ND	ND
riboswitch-11	849418-849636 (+)	*spd_0831*, *spd_0833*	ND	ND	ND
riboswitch-12	864986-865160 (+)	*spd_0850*, *pyrK*	ND	ND	ND
riboswitch-13[a]	996442-996678 (-)	*ppC*, *spd_0954*	*spd_sr12*	967941 – 968177 (+)	dRNA-seq

riboswitch-14[d]	1052832-1052928 (-)	*spd_0897*, *spd_0898*	ND	ND	ND
riboswitch-15[a]	1168835-1168970 (-)	*nth*, *pyrR*	*spd_sr62*	1168377-1168512 (-)	dRNA-seq
riboswitch-16	1175106-1175233 (+)	*gidB*, *uraA*	*spd_sr64*	1174647 – 1174790 (+)	sRNA-seq & dRNA-seq
riboswitch-17[a]	1250018-1250194 (-)	*spd_1216*, *spd_1217*	*spd_sr70*	1249557 – 1249736 (-)	sRNA-seq & dRNA-seq
riboswitch-18[a]	1326523-1326709 (-)	*spd_1308*, *spd_1307*	*spd_sr74*	1326066 – 1326252 (-)	sRNA-seq & dRNA-seq
riboswitch-19	1404449-1404618 (-)	*mntE*, *spd_1383*	ND	ND	ND
riboswitch-20[a]	1459261-1459468 (+)	*spd_1441*, *spd_1442*	*spd_sr80*	1458804 – 1458975 (+)	sRNA-seq
riboswitch-21	1491275-1491462 (-)	*divIVA*, *ileS*	ND	ND	ND
riboswitch-22[a]	1619505-1619652 (-)	*spd_1605*, *spd_1604*	*spd_sr88*	1619052 – 1619299 (-)	sRNA-seq & dRNA-seq
riboswitch-23	1642795-1642969 (+)	*spd_1626*, *xpt*	ND	ND	ND
riboswitch-24[a]	2006750-2006885 (-)	*cbpD*, *spd_2027*	*spd_sr114*	2006263 – 2006428 (-)	sRNA-seq & dRNA-seq

[a] sRNAs validated for expression in this study.

[b] sRNAs validated for expression by Tsui et al 2010 (1).

[c] sRNAs validated for expression by Slager et al 2018 (2).

[d] sRNAs predicted to be present in D39V by Slager et al 2018 (2).

ND = not detected.

Table S2: List of sRNAs predicted in TIGR4, the coordinates of which are absent in D39W

Mann et al (2012)			**Acebo et al (2012)**		
sRNA Name	**Start**	**Stop**	**sRNA Name**	**Start**	**Stop**
			trn0027	79336	79386
F1[a]	91593	91664	**trn0036**	91593	91651
F56	158993	159090			
			srn081	273681	273799
			trn0256	436064	436160
			srn135		
F12	444624	444663		438151	438276
			trn0290	497393	497444
			srn151	501260	501364
F57	553923	553982			
F24	610528	610660			
R13	623327	623372			
			trn0355	673130	673199
F29	743792	743828			
F62	995726	995787			
			srn266	1035461	1035582
F40	1063101	1063151			
F41[a]	1071112	1071214	**srn277**	1071102	1071212
F46	1454220	1454277			
			trn0757	1515505	1515588
R16	1227703	1227744			
R12[a]	1731041	1731440	**trn0830**	1731046	1731098
			srn448	1901982	1902084
R11	2002164	2002195			

[a] sRNAs identified to be common in TIGR4 by Mann et al., 2012 (3) and Acebo et al., 2012 (4).

Table S3: Classification of *S. pneumoniae* D39W sRNAs into different functional categories

sRNA name	Start	End	Upstream Gene	Downstream Gene	Predicted/Putative Function
Spd_sr46	781128	781485	*spd_0768*	*spd_0770*	tmRNA[a], defective in lung infection[b], attenuated in invasive pneumonia model[b]
Spd_sr85	1597869	1598142	*spd_1580*	*spd_1581*	6S RNA[a]
Spd_sr48	862699	862827	*spd_0846*	*infC*	L20 leader[a]
Spd_sr55	999541	999605	*spd_0988*	*rplU*	L21 leader[a]
Spd_sr54	1215844	1215967	*spd_1189*	*rplJ*	L10 leader[a]
Spd_sr43	643872	644040	*lctO*	*spd_0622*	TPP-riboswitch[a]
Spd_sr44	646754	646926	*thiE-1*	*spd_0625*	TPP-riboswitch[a], defective replication in bloodstream[b]
Spd_sr114	2006263	2006428	*cbpD*	*spd_2027*	TPP-riboswitch[a]
Spd_sr22	173827	173979	*ruvA*	*ribE*	FMN-riboswitch[a]
Spd_sr28	376371	376530	*spd_0371*	*spd_0372*	Glycine-riboswitch[a]
Spd_sr29	381067	381205	*spd_0376*	*serS*	T-box element[a]
Spd_sr32	496899	497104	*spd_0490*	*spd_0491*	T-box element[a]
Spd_sr34	512889	513087	*bglA-2*	*pheS*	T-box element[a], defective in lung infection and nasopharynx colonization[b], attenuated in invasive pneumonia model[b]
Spd_sr12	967941	968177	*ppC*	*spd_0954*	T-box element[a]
Spd_sr70	1249557	1249736	*potD*	*spd_1217*	T-box element[a]
Spd_sr74	1326066	1326252	*spd_1308*	*spd_1307*	T-box element[a]
Spd_sr80	1458804	1458975	*spd_1441*	*spd_1442*	T-box element[a]
Spd_sr88	1619052	1619299	*spd_1605*	*spd_1604*	T-box element[a], defective in lung infection[b]
Spd_sr62	1168377	1168512	*nth*	*pyrR*	PyrR-regulated RNA
Spd_sr64	1174647	1174790	*gidB*	*uraA*	PyrR-regulated RNA[c], defective in lung infection[b], attenuated in invasive pneumonia model[b]
Spd_sr49	825802	826129	*spd_0808*	*cad*	defective in lung infection[b]
Spd_sr17	912572	912715	*spd_0899*	*asd*	defective in nasopharynx colonization[b]
Spd_sr111	1973304	1973394	*spd_1996*	*adcA*	Toxin module III[c], defective in nasopharynx colonization[b]
CcnE	212278	212425	*spd_0221*	*spd_0222*	Stationary phase autolysis[d], competence development[e], attenuated in invasive pneumonia model[b]

Spd_sr23	231823	232091	*spd_0240*	*ruvB*	Toxin module I[c]
Spd_sr24	231853	232035	*ruvB*	*spd_0240*	Antitoxin sRNA I[c]
Spd_sr109	1972859	1973060	*spd_1996*	*adcA*	Toxin module II[c]
Spd_sr110	1973000	1973113	*adcA*	*spd_1996*	Antitoxin sRNA II[c]
Spd_sr112	1973342	1973456	*adcA*	*spd_1996*	Antitoxin sRNA III[c]

[a] Function predicted from Rfam database.

[b] D39 sRNAs correspond to the sRNAs mutants in TIGR4 strain, which exhibited virulence defect in different host niches as determined by (3).

[c] D39 sRNAs in this functional category identified to be common between *S. pneumoniae* D39W (this study) and D39V (2). These specific functional categories are assigned to respective sRNAs by Slager et al., 2018 (2) in D39V.

[d] sRNA function first determined in *S. pneumoniae* R6 (5).

[e] sRNA function elucidated in D39 by Schnorpfeil et al 2013 (6) and Tsui et al 2010 (1).

TABLE S4: List of *S. pneumoniae* D39 genes with antisense expression[a]

SPD #	Gene	Sense	Antisense	Antisense/ Sense
spd_0022	pseudo gene	202	26121	129
spd_0034	IS1167, transposase, truncation	648	54856	85
spd_0048	IS1167, transposase, truncated	5233	56419	11
spd_0085	hypothetical protein	52	68430	1316
spd_0100	hypothetical protein	578	103989	180
spd_0137	IS1167, transposase, truncated	4512	36854	8
spd_0142	conserved hypothetical protein;	836	4311	5
spd_0223	iron(III) ABC transporter, permease protein	**1696**	**8697**	**5**
spd_0224	iron(III) ABC transporter, permease protein	**1009**	**5124**	**5**
spd_0225	ABC transporter, ATP-binding protein	**1161**	**7847**	**7**
spd_0413	hypothetical protein	**67**	**34000**	**507**
spd_0415	transposase, IS116/IS110/IS902 family , degenerate	**1313**	**315472**	**240**
spd_0416	transposase, IS116-IS110-IS902 family, degenerate	**542**	**106534**	**197**
spd_0450	type I restriction-modification system, S subunit, putative	20155	160108	8
spd_0466	conserved hypothetical protein	1262	16139	13
spd_0470	***blp*C**	**278**	**19445**	**70**
spd_0471	***blp*B**	**2082**	**361566**	**174**
spd_0472	***blp*A**	**10387**	**288906**	**28**
spd_0540	amino acid ABC transporter, amino acid-binding protein, putative	10502	138547	13
spd_0552	hypothetical protein	673	10445	16
spd_0607	transposase, IS30 family, degenerate	186	30722	165
spd_0615	ABC transporter, substrate-binding protein, truncated	**3365**	**26180**	**8**
spd_0616	amino acid ABC transporter, ATP-binding protein	**3051**	**22951**	**8**
spd_0618	amino acid ABC transporter, permease protein	**991**	**8219**	**8**
spd_0638	transposase family protein	**650**	**10972**	**17**
spd_0639	conserved hypothetical protein	**285**	**2399**	**8**
spd_0682	hypothetical protein	1928	15842	8
spd_0715	MutT/nudix family protein	613	4792	8
spd_0758	IS1167 transposase, truncation	1960	801301	409
spd_0770	tRNA-Ser	168	845	5
spd_0791	conserved hypothetical protein	1064	27564	26

spd_0794	IS1381 transposase OrfA, truncated	819	19512	24
spd_0810	pilin gene inverting-related protein, authentic frameshift	1067	244038	229
spd_0831	conserved domain protein	1021	7845	8
spd_0841	*ald* (pseudo gene)	12126	124948	10
spd_0895	*hem*H	4218	66266	16
spd_0938	Tn5252, relaxase; truncated relative to TIGR4	1647	8488	5
spd_0966	IS1167, transposase, truncated	6053	108659	18
spd_0975	DNA repair protein *radC* homolog	6602	77356	12
spd_0992	conserved hypothetical protein	101	1968	19
spd_0993	conserved hypothetical protein	54	1013	19
spd_1059	hypothetical protein	44	784	18
spd_1189	conserved hypothetical protein	256	1472836	5753
spd_1213	membrane protein, putative	**9134**	**127449**	**14**
spd_1214	ABC transporter, ATP-binding protein	**4191**	**40066**	**10**
spd_1261	conserved hypothetical protein	1781	9001	5
spd_1266	membrane protein, putative	4490	22971	5
spd_1275	transcriptional regulator, GntR family	96	11619	121
spd_1430	*fer*	16179	79670	5
spd_1452	membrane protein, putative; truncated relative to TIGR4	**5519**	**61420**	**11**
spd_1453	pseudo gene	**2444**	**38771**	**16**
spd_1454	conserved hypothetical protein	**12049**	**118485**	**10**
spd_1455	conserved hypothetical protein	**155**	**2199**	**14**
spd_1456	conserved hypothetical protein	**204**	**3306**	**16**
spd_1500	ABC transporter, permease protein	44	754	17
spd_1502	ABC transporter, substrate-binding protein	280	1541	6
spd_1578	conserved hypothetical protein	2047	16927	8
spd_1579	conserved hypothetical protein	1676	28532	17
spd_1593	type IV prepilin peptidase, putative	1547	12407	8
spd_1618	conserved hypothetical protein	220	453455	2061
spd_1628	***xpt***	**2175**	**18732**	**9**
spd_1629	***pbu*X**	**5962**	**61196**	**10**
spd_1666	IS1167 transposase, truncation	10284	124603	12
spd_1681	IS1167, transposase	606	35981	59
spd_1770	conserved hypothetical protein	156	33068	212
spd_1795	hypothetical protein	1029	7597	7
spd_1901	transposase, putative	323	59946	186
spd_1906	IS1381, transposase OrfB	4511	185846	41
spd_1939	hypothetical protein	539	91168	169

spd_1955	hypothetical protein	409	52774	129
spd_1967	IS1381, transposase OrfB	473	124528	263
spd_1980	IS3-Spn1, hypothetical protein, truncated	419	2448	6
spd_1983	hypothetical protein	122	36220	297
spd_2001	hypothetical protein	87	3345	38
spd_2008	group II intron, maturase, degenerate	346	6284	18
spd_2038	conserved hypothetical protein	293	6298	21

[a] Bolded rows correspond to genes in an operon. The numbers in the sense and antisense column represent the RPKM (Reads Per Kilobase of transcript per Million mapped reads) values for the (+) and the (-) strand and the ‘antisense/ sense’ column represents the ratio between the respective RPKM values of two strands for each gene.

Table S5: List of oligonucleotide probes used for Northern blot analysis

sRNA name	Oligonucleotide probe sequence in 5′-3′ direction
Spd_sr5[a]	CAAATAAATCATTGATAGTAACACTTACGACATCC
Spd_sr6[a]	GGAAATCATCGCCAAACTGGACTACGACGCCCCATC
Spd_sr8	CAAACCGCGTCAACGTCGCCTTGCCGTATAGATGTTACTG
Spd_sr13	CAGTGTTTTGAGCAGCCCGCAGCTAGTTTCCTAGTTTG
Spd_sr16[a]	GAAATATTTTAACGGTCATAACGGACTATTCCAGA
Spd_sr19	CATTCCATTTTGTGTCTAGATTGACCTTGATTCGTTAC
Spd_sr20	GCCCCATATGACCTATAATGAAAAGCGTCTAACCAACTC
Spd_sr22	CATCCAGACTATACTGTCGGCTTTGGAATTTCACC
Spd_sr23	CTTCGTATAGTCGATAAATGGCTAAACAAGGACAA
Spd_sr24	CCATCATGGGCTATCTCACCTCCTTTCGTAGGCGG
Spd_sr31	TAAAACCAAACCAGCTCCTAGTAAGGCTAGAAGGG
Spd_sr32	AACCAGTTTTGACACATTTGTGGACTTCTCAGCG
Spd_sr35	GGGATAACAAAGAGAATGACCGAAAGGGCAGAGAG
Spd_sr36-1[a,b]	GCCGTCAAGCAACTCAAAGGAAAATAGGAAATCGA
Spd_sr36-2[b]	CGGGCGATTAGCTAAATGCTTTACTAACTCTCTCG
Spd_sr42	GTTACGATAGTCTTGGTAAAATAGAATTGCCCAATAAACC
Spd_sr43	CCCTTCGCCAGTCTTAACTGTATCAGGTTCAATGGG
Spd_sr45	CTTCCATCAGCTAGCATTACCTAGATTGAGTCAGAGGG
Spd_sr47	GAAAACACCTCTGTGTTATACTTGTTGTTCAACCACAAAC
Spd_sr49	ATCGTTTCACAAGATGACGTGTGCTTTCACCACAC
Spd_sr54	GATACGTCTTGTCCTCGGCAGGATATTTATGAGTCG
Spd_sr55	CCCATAAGATCTATAATGAAAAGCGACCAAACAACTC
Spd_sr57	GTGCCTGATTTTCTCTCCTAATTGTTATTAGTAGC
Spd_sr58[a]	GGATTTCCCCGATTTTTTGTGAAGACTCGCCTAGC
Spd_sr62	GACTTCACCCTTCTACTTTATCGCGCTCCTTGCCTGCCTC
Spd_sr66	TTTCCCAACTAAAAGCACTCCAGTTACCGCAACGG
Spd_sr67	CTAAATCGCATTCCATCAACTTCATCTGTTTCGTC
Spd_sr70	CTCTGGACTAAGACAAGTGAAAATCAATTCTCAAC
Spd_sr71	CTACATGTGATGGCAATCAAGATATCAAGAATCATCCTAC
Spd_sr73	AAGAGTAAACTCAGCTAGTCCAACTAACTGAGTT
Spd_sr74	CTACAGAAAGCGCCAGCCCTTTATTTTGCCCTACT
Spd_sr78	CACTATCACAGGTTTCTGCCAGCTACTCCCTTGAG
Spd_sr80	GAGGGACACCTCTGATCGGCTCTAACGTGGCCACC
Spd_sr81	GGCAAATACATAAAAAGCTAACTGAACATTCTCGTATCC
Spd_sr82	CATTATCACACCTTTCTAAGGTGGTTTTTTTATCCCGT
Spd_sr83	TTACTTTAATCGTTACTGTCATATGAGAGTCCTCG
Spd_sr84	AATATGGGAGCCTTTACCGCCTCTAGTCTTTATCC
Spd_sr85	AGACGGCCTACATGCTGTTAGAAGACTTTTGTTTC
Spd_sr88	CCAACCACCTATTCACAATCACCACAGGCTCCC
Spd_sr89	CTGATTGATCGTTCATTATGATTCCTTGTACTGAACGAAC
Spd_sr96	GATTGCTCCAGTAATAAAACCATTGGGAATGAAGG
Spd_sr99	CATCAAGTGCTCCTGACGCAACATACCATAGTCCG
Spd_sr100[a]	TTCTCTGTGTGTAGTGTACTTGCCACAATGCTTAC
Spd_sr106	ATGTGAGATAAAACGTGCTGTCGGTAAAAGCAAGC
Spd_sr108	GAAAATCAAAGTGCAAACTAGGAAGCTAGCCGCAG
Spd_sr109	CTATGATAAGGCTCATTAGTTTCACCTCCTCTCAC

Spd_sr111	CACCTCCTCTCACGAACCCATAGGAACGTAATCGG
Spd_sr114	GTGCGGTCGAAGTTTAACACTTCCTCTCAGACTGT
Spd_sr116	ACTTCAACTCGCAAAACGACCCAAGGGCAACTGCTT
CcnC	TTTGTTACAACAAGTTAGGAGGTCTTCTTGTAACC
CcnB	CTCCACCTGATTGGGTGGAGTTAAGGGAGATTATTATG
CcnE	TAAAAGCCACCCATACAGGCGACTTTTGAAGGAG
CcnD	GGGAGATTATTATGAAAAAGTTTTAGGAGTTTAAGTTAAGG
F14	CAACTACTGTCCAAATCATACCAATTATGAACGG
F23	CGGTTATCTTCTACCTATCTATTTTTTCAACTCCT

[a] Probes were designed within regions that were slightly upstream or downstream of the coordinates predicted for these sRNAs.

[b] The presence of Spd_sr36 can be validated with both the probes -Spd_sr36-1 and Spd_sr36-2.

SUPPLEMENTAL FIGURES

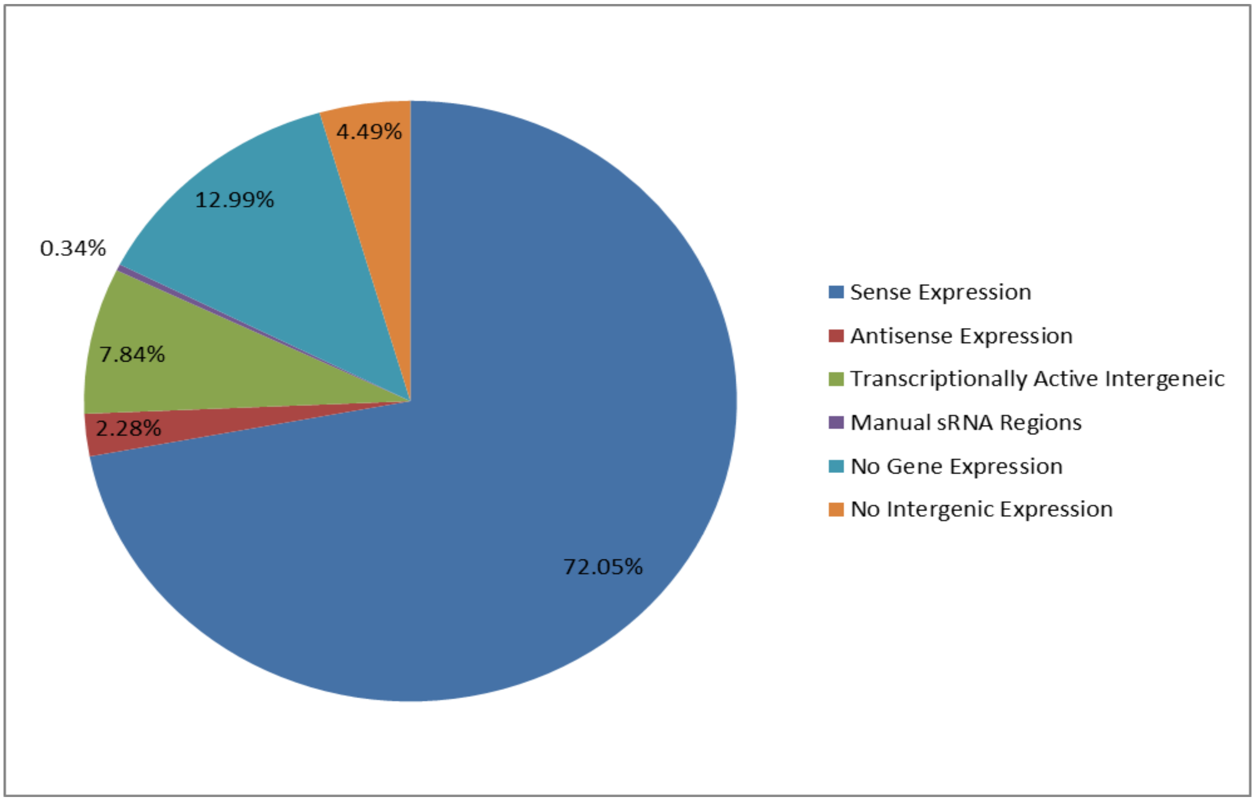

Fig. S1. Transcriptome map of *S. pneumoniae* D39. Comprehensive map showing the distribution of expressed and unexpressed genes and intergenic regions, expressed sRNAs, and antisense transcripts of the *S. pneumoniae* D39W genome based on the minimum coverage threshold of eight read per base as determined for the mRNA-seq analysis.

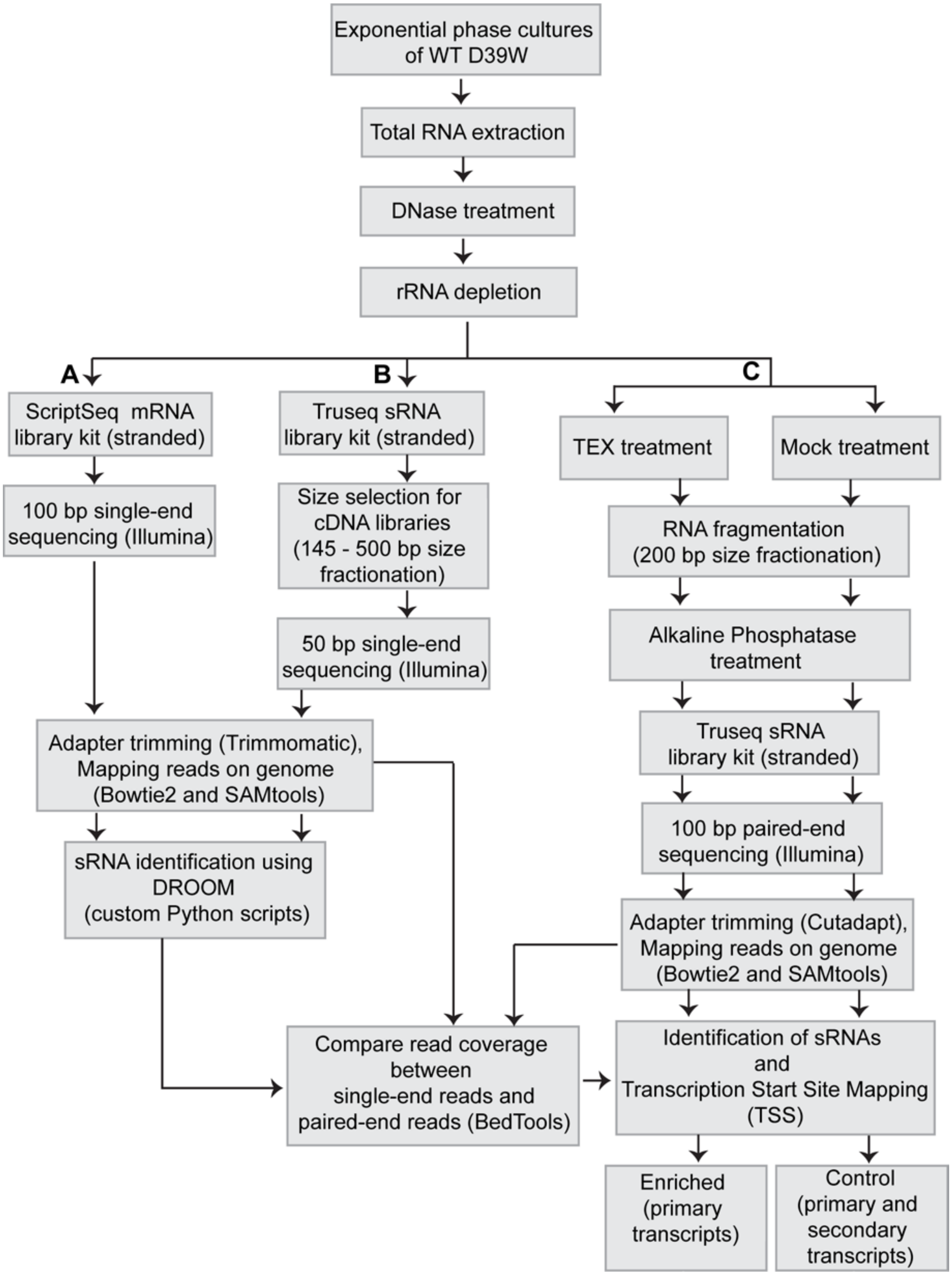

Fig. S2. Experimental design and data analysis pipeline used to identify sRNAs in D39W.

Combined data from mRNA-seq (A) and sRNA-seq (B) analysis were used to identify sRNAs in

D39W. Transcription start sites of the identified sRNAs were mapped using differential RNA-seq analysis (C). Paired-end sequencing data from dRNA-seq were used to independently identify sRNAs in D39W. Analysis was performed with sequencing reads (single-end and paired-end) that mapped to the D39W reference genome. Details of the experimental design and analysis are described in Materials and Methods.

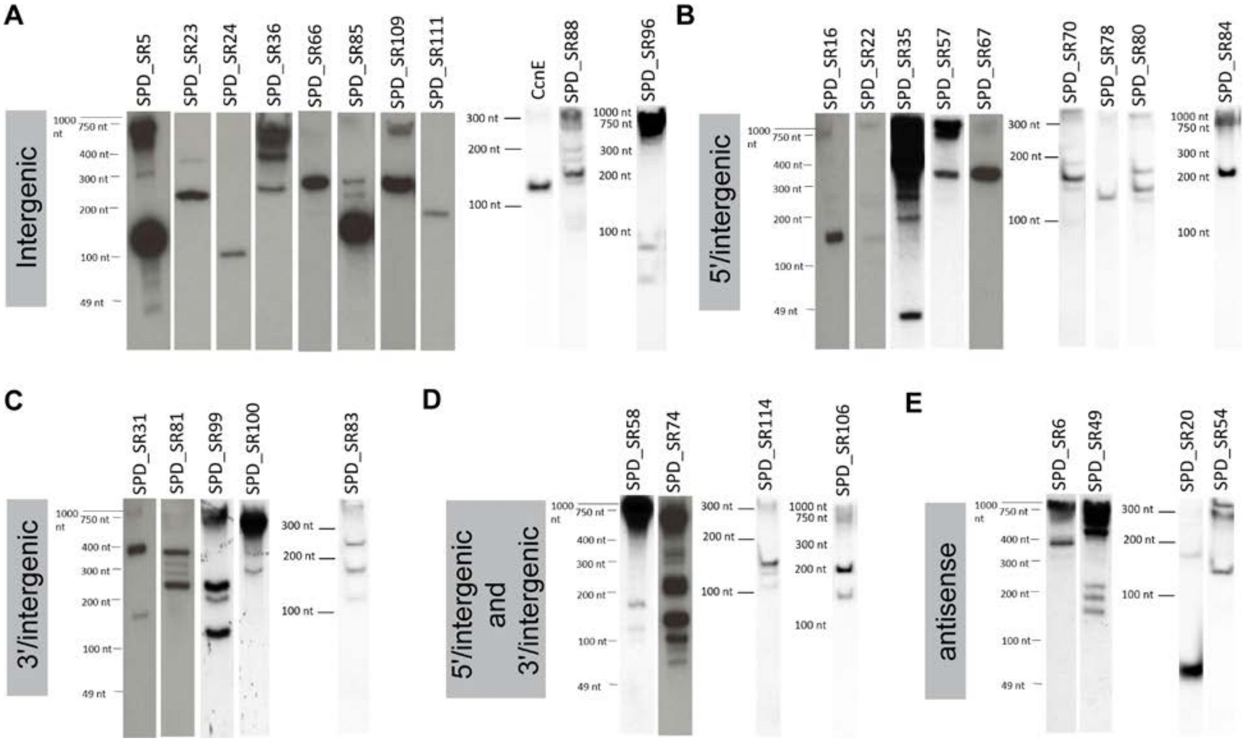

Fig. S3. Northern blots corresponding to sRNAs identified in sRNA-seq. Northern blot validations corresponding to the sRNAs identified as intergenic, 5'/intergenic, 3'/intergenic, 5'/intergenic and 3'/intergenic or antisense by sRNA-seq analysis. Probes used for northern blotting are listed in Table S5.

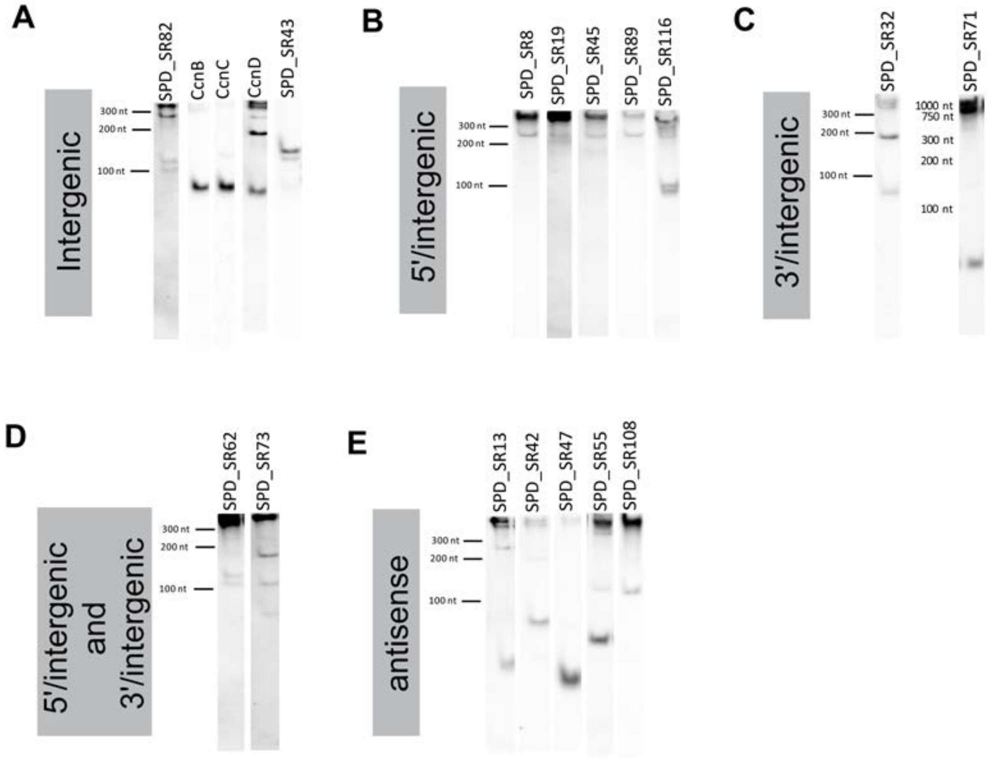

Fig. S4. Northern blots corresponding to sRNAs identified in dRNA-seq. Northern blot validations corresponding to the sRNAs identified as intergenic, 5'/intergenic, 3'/intergenic, 5'/intergenic and 3'/intergenic or antisense by dRNA-seq analysis. Probes used for northern blotting are listed in Table S5.

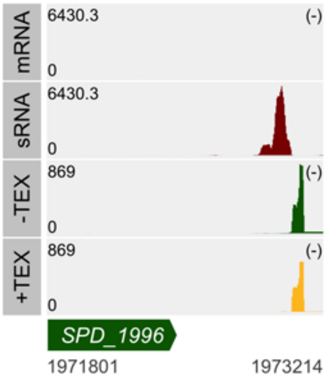

Fig. S5. Coverage graph of Spd_sr110, identified as an sRNA in D39W. Spd_sr110 was predicted to be present in D39V (2), but was not identified as a sRNA by RNA-seq analysis in D39V. Spd_sr110 identified as a sRNA by dRNA-seq analysis in D39W represented by the peaks in –TEX (in green) and + TEX (in yellow) tracks. Track labels corresponding to read coverage maps are described in the legend of Figure 2. The peak in the sRNA track corresponds to Spd_sr109 sRNA identified by sRNA-seq analysis on the (+) strand. Spd_sr109 and Spd_sr110 are predicted to constitute a toxin-antitoxin module in D39.

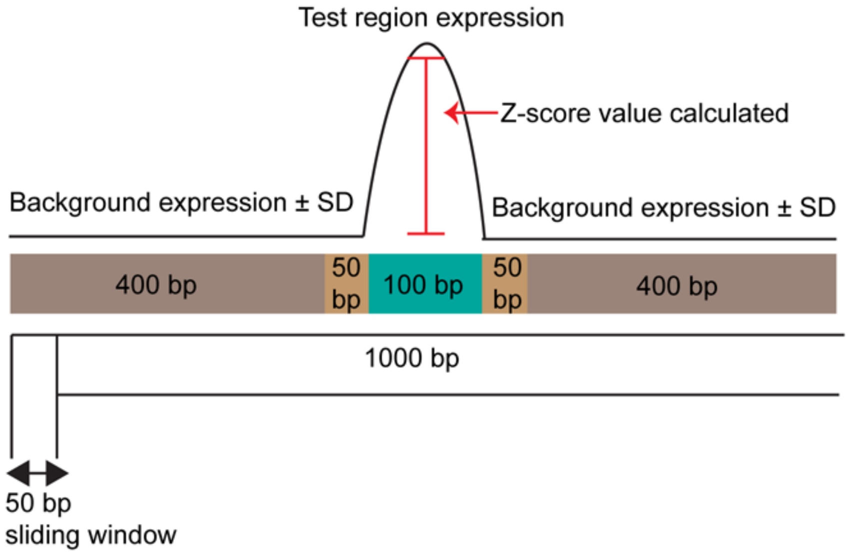

Fig. S6. Schematic showing strategy used to identify sRNAs by RNA-seq. A window of 1,000 base pairs in length was used. Predicted sRNAs identified as peaks in expression (100 bp test interval at the center in blue) relative to expression in neighboring portions of the genome (400 bp total to each side; flanking intervals in brown). The test area was flanked on either side by 50 bp of unused spacer sequence (in orange) to eliminate edge effects. The z-score of the test interval was determined relative to the mean and standard deviation (SD) observed in the flanking intervals. Analysis was performed genome-wide and z-scores calculated at 50 bp sliding intervals across the genome.

SUPPLEMENTAL REFERNCES

1. **Tsui HCT, Mukherjee D, Ray VA, Sham LT, Feig AL, Winkler ME.** 2010. Identification and characterization of noncoding small RNAs in *Streptococcus pneumoniae* serotype 2 strain D39. J Bacteriol doi:10.1128/JB.01204-09.
2. **Slager J, Aprianto R, Veening J-W.** 2018. Deep genome annotation of the opportunistic human pathogen *Streptococcus pneumoniae* D39. Nucleic Acids Res **46:**9971-9989.
3. **Mann B, van Opijnen T, Wang J, Obert C, Wang Y-D, Carter R, McGoldrick DJ, Ridout G, Camilli A, Tuomanen EI, Rosch JW.** 2012. Control of virulence by small RNAs in *Streptococcus pneumoniae*. PLoS pathog **8:**e1002788-e1002788.
4. **Acebo P, Martin-Galiano AJ, Navarro S, Zaballos A, Amblar M.** 2012. Identification of 88 regulatory small RNAs in the TIGR4 strain of the human pathogen *Streptococcus pneumoniae*. RNA **18:**530-546.
5. **Halfmann A, Kovács M, Hakenbeck R, Brückner R.** 2007. Identification of the genes

directly controlled by the response regulator CiaR in *Streptococcus pneumoniae*: five out of 15 promoters drive expression of small non-coding RNAs. Mol Microbiol **66:**110-126.

6. **Schnorpfeil A, Kranz M, Kovács M, Kirsch C, Gartmann J, Brunner I, Bittmann S, Brückner R.** 2013. Target evaluation of the non-coding csRNAs reveals a link of the two-component regulatory system CiaRH to competence control in *Streptococcus pneumoniae* R6. Mol Microbiol **89:**334-349.