Landes Bioscience

www.landesbioscience.com

Supplemental Material to:

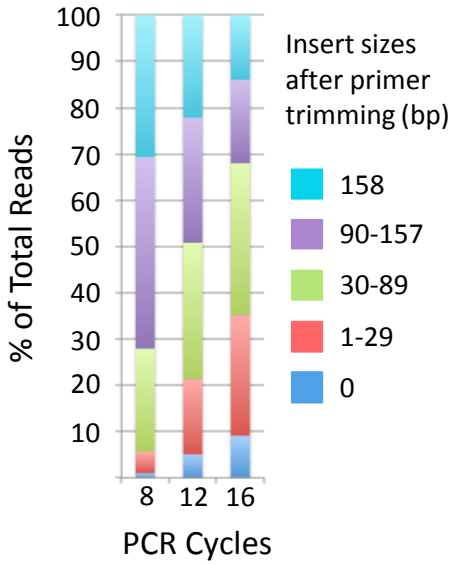
Stanley A. Langevin, Zachary W. Bent, Owen D. Solberg, Deanna J. Curtis, Pamela D. Lane, Kelly P. Williams, Joseph S. Schoeniger, Anupama Sinha, Todd W. Lane and Steven S. Branda

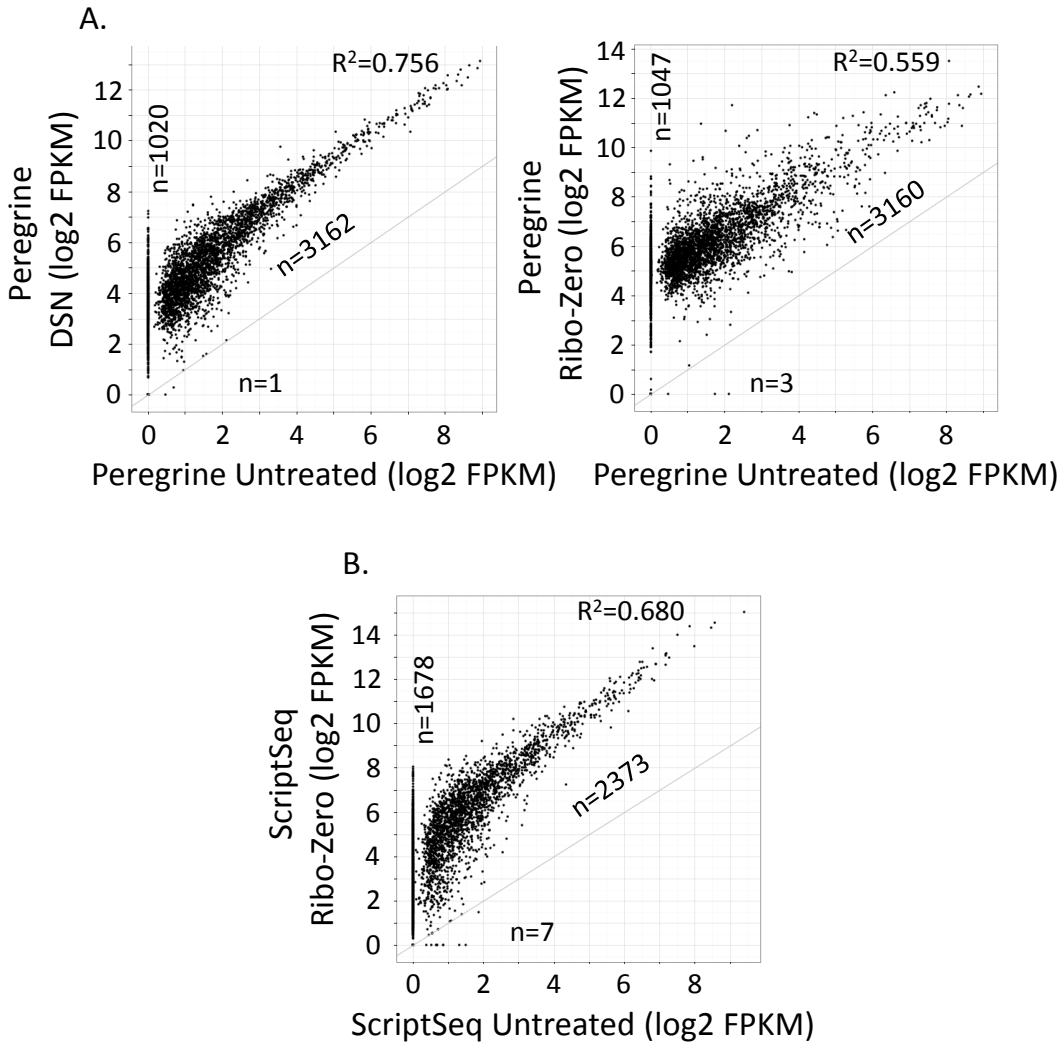
Peregrine: A rapid and unbiased method to produce strand-specific RNA-Seq libraries from small quantities of starting material

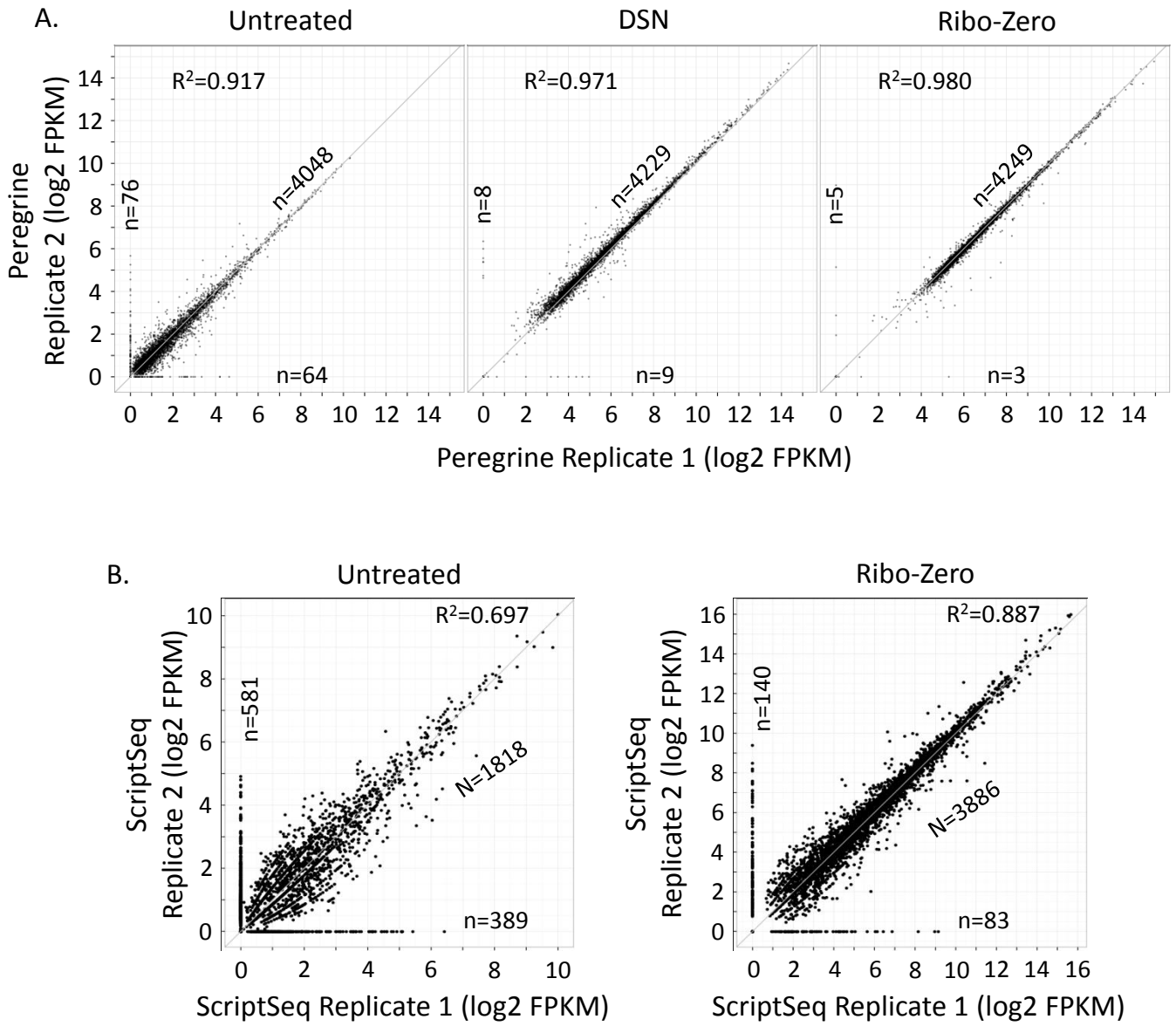
2013; 10(4)

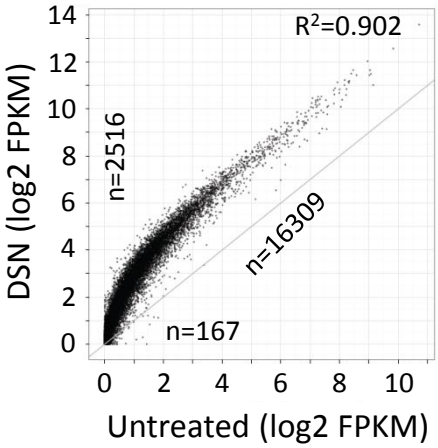
http://dx.doi.org/10.4161/rna.24284

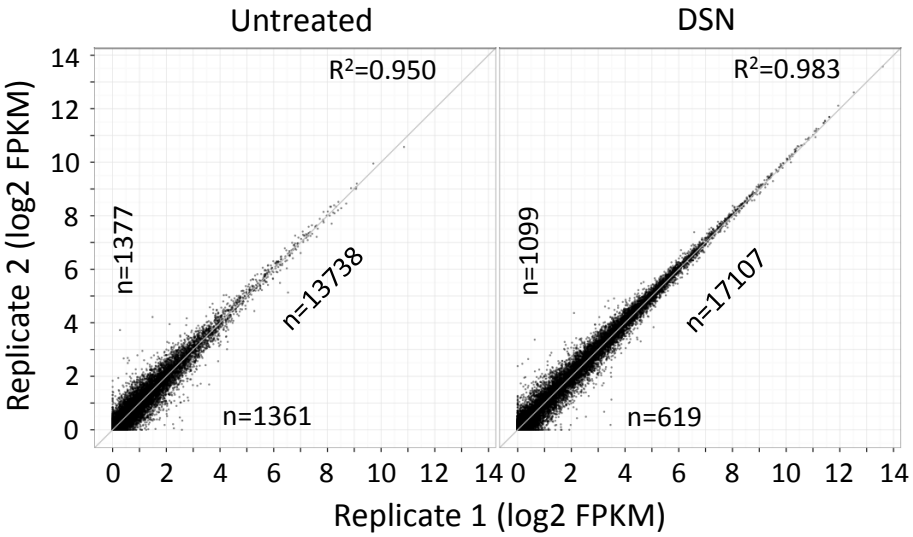
www.landesbioscience.com/journals/rnabiology/article/24284/


100
90
80
70
60
50
40
30
20
10
% of Total Reads
8
12
16
PCR Cycles
Insert sizes after primer trimming (bp)
158
90-157
30-89
1-29
0


A.
R²=0.756
n=1020
n=3162
n=1
Peregrine DSN (log2 FPKM)
Peregrine Untreated (log2 FPKM)
R²=0.559
n=1047
n=3160
n=3
Peregrine Ribo-Zero (log2 FPKM)
Peregrine Untreated (log2 FPKM)
B.
R²=0.680
n=1678
n=2373
n=7
ScriptSeq Ribo-Zero (log2 FPKM)
ScriptSeq Untreated (log2 FPKM)


A.
Untreated
DSN
Ribo-Zero
R²=0.917
R²=0.971
R²=0.980
n=4048
n=4229
n=4249
n=76
n=8
n=5
n=64
n=9
n=3
Peregrine Replicate 2 (log2 FPKM)
Peregrine Replicate 1 (log2 FPKM)
B.
Untreated
Ribo-Zero
R²=0.697
R²=0.887
n=581
N=1818
n=389
n=140
N=3886
n=83
ScriptSeq Replicate 2 (log2 FPKM)
ScriptSeq Replicate 1 (log2 FPKM)


R²=0.902
n=2516
n=16309
n=167
DSN (log2 FPKM)
Untreated (log2 FPKM)
0
2
4
6
8
10
12
14


Untreated
DSN
R²=0.950
R²=0.983
n=1377
n=13738
n=1361
n=1099
n=17107
n=619
Replicate 2 (log2 FPKM)
Replicate 1 (log2 FPKM)
0
2
4
6
8
10
12
14

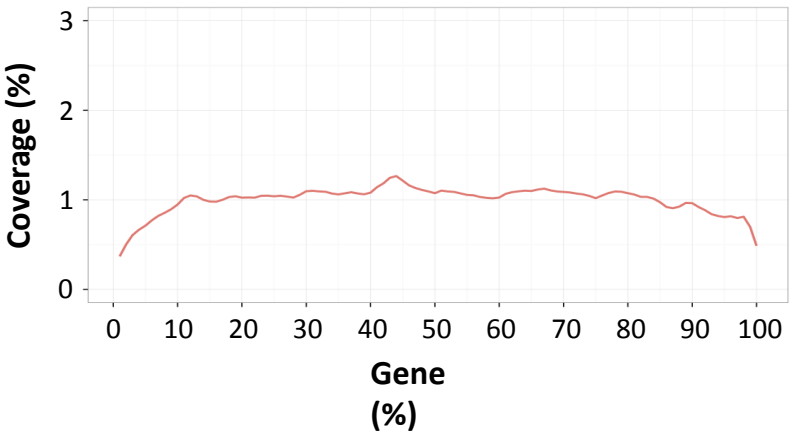

Coverage (%)
3
2
1
0
0
10
20
30
40
50
60
70
80
90
100
Gene
(%)

Figure S1: Effect of PCR cycling on insert size during library enrichment.

A single Peregrine-prepared *E. coli* K-12 ss cDNA library was subjected to our qPCR assay, and the optimal cycle number was determined to be 12 cycles. The ss cDNA library was then amplified and indexed using the indicated number of PCR cycles. Following sequencing, the sizes of all inserts were determined.

Figure S2: Effect of rRNA depletion treatments on relative abundances of *E. coli* transcripts in Peregrine and ScriptSeq libraries.

Libraries were prepared from *E. coli* K-12 total RNA, sequenced, and aligned to the reference *E. coli* genome. Read statistics for the library sub-samples analyzed are included in Tables S3 and S4. Transcript abundances (FPKM) are plotted as a function of library treatment. Points above the diagonal line represent transcripts that are enriched in treated (DSN or Ribo-Zero) libraries. Points on the axes represent transcripts with zero coverage in one of the two libraries. The number of data points in the diagonal cloud and on the axes is indicated. The coefficient of determination (R^2) value, as calculated for all transcripts represented in the libraries, is indicated as well. **A:** Libraries prepared using Peregrine in combination with DSN-mediated normalization (left panel) or Ribo-Zero treatment (right panel). The R^2 and Spearman rank correlation coefficient (ρ) values, as calculated for the 50% most abundant transcripts detected in the untreated library, are 0.964 and 0.937 (untreated *versus* DSN), and 0.802 and 0.740 (untreated *versus* RiboZero), respectively. **B:** Libraries prepared using ScriptSeq in combination with Ribo-Zero treatment. The R^2 and ρ values, as calculated for the 50% most abundant transcripts detected in the untreated library, are 0.802 and 0.799, respectively.

Figure S3: Technical replicates of Peregrine and ScriptSeq library preparation from *E. coli* K-12 RNA.

Libraries were prepared from *E. coli* K-12 RNA, sequenced, and aligned to the reference *E. coli* genome. Read statistics for the library sub-samples analyzed are included in Tables S3 and S4. Transcript abundances (FPKM) from two replicate libraries are plotted in each panel. Points on the axes represent transcripts with zero coverage in one of the two libraries. The number of data points in the diagonal cloud and on the axes is indicated. The coefficient of determination (R^2) value, as calculated for all transcripts represented in the libraries, is indicated as well. **A:** Libraries prepared using Peregrine alone (left panel) or in combination with DSN-mediated normalization (middle panel) or Ribo-Zero treatment (right panel). **B:** Libraries prepared using ScriptSeq alone (left panel) or in combination with Ribo-Zero treatment (right panel).

Figure S4: Effect of rRNA depletion treatment on relative abundances of human PBMC transcripts in Peregrine libraries.

Peregrine libraries were prepared from human PBMC total RNA, sequenced, and aligned to the reference human genome. Transcript abundances (FPKM) are plotted as a function of library treatment. Dots above the diagonal line represent transcripts that are enriched in treated (DSN-normalized) libraries. Points on the axes represent transcripts with zero coverage in one of the two libraries. The number of data points in the diagonal cloud and on the axes is indicated. The coefficient of determination (R^2) value, as calculated for all transcripts represented in the libraries, is indicated as well. Read statistics for the libraries analyzed are included in Tables 2 and S5.

Figure S5: Technical replicates of Peregrine library preparation from human PBMC RNA.

Libraries were prepared from human PBMC RNA using Peregrine alone (left panel) or in combination with DSN-mediated normalization (right panel). The libraries were sequenced, and reads aligned to the reference human genome. Transcript abundances (FPKM) from two replicate libraries are plotted in each panel. Read statistics for the libraries analyzed are included in Tables 2 and S5.

Figure S6: Uniformity of coverage along length of transcripts in Peregrine libraries prepared from human PBMC RNA.

Average coverage at each percentile of length for all genes. Coverage depth was tabulated at each nucleotide position within each gene of 1000-5000 bp. Each gene was length-normalized by percentile, and coverage was calculated as the total number of sequenced bases in each percentile divided by the total number of sequenced bases. The first percentile represents the 5' end and the 100th percentile represents the 3' end of the gene. Read statistics for the libraries analyzed are included in Tables 2 and S5.

Table 1: Strand specificity of Peregrine & ScriptSeq libraries prepared from *E. coli* RNA.

Prep & Treatment	rRNA	CDS	CDS w/known asRNA
Peregrine Untreated	99.73 ± 0.01	94.77 ± 0.04	86.97 ± 0.35
Peregrine DSN	99.86 ± 0.01	94.20 ± 0.23	87.17 ± 0.19
Peregrine Ribo-Zero	2.21 ± 0.09	87.15 ± 1.28	81.72 ± 1.51
ScriptSeq Untreated	99.53 ± 0.10	97.85 ± 0.19	94.04 ± 1.24
ScriptSeq Ribo-Zero	22.33 ± 4.27	97.87 ± 0.30	93.11 ± 0.92

Table 2: Strand specificity of Peregrine libraries prepared from human PBMC RNA.

Treatment	rRNA	CDS
Untreated	99.50 ± 0.28	96.37 ± 0.00
DSN	99.71 ± 0.15	97.03 ± 0.00

Table S1: SGS read statistics for Peregrine libraries prepared from *E. coli* RNA.

Treatment	Libraries	Raw Reads (millions)	Passed Qfilter (% of Raw)	Mapped to K12 (% of Qfilter)	Mapped to rRNA (% of Mapped)	Mapped to CDS (% of Mapped)	Coverage (% of CDS)
Untreated	2	26.7 ± 1.2	79.93 ± 0.13	99.41 ± 0.18	96.65 ± 0.06	2.34 ± 0.04	98.37 ± 0.00
DSN	4	21.9 ± 4.9	81.45 ± 1.20	95.89 ± 1.18	40.90 ± 2.88	49.87 ± 2.34	99.31 ± 0.00
Ribo-Zero	4	31.4 ± 2.9	62.52 ± 085	91.14 ± 0.93	26.17 ± 0.71	62.16 ± 0.78	99.30 ± 0.00

Table S2: SGS read statistics for ScriptSeq libraries prepared from *E. coli* RNA.

Treatment	Libraries	Raw Reads (millions)	Passed Qfilter (% of Raw)	Mapped to K12 (% of Qfilter)	Mapped to rRNA (% of Mapped)	Mapped to CDS (% of Mapped)	Coverage (% of CDS
Untreated	3	2.3 ± 0.2	96.72 ± 0.51	99.66 ± 0.04	98.28 ± 0.35	1.10 ± 0.11	59.93 ± 2.1
Ribo-Zero	3	1.6 ± 0.3	96.71 ± 0.30	96.21 ± 0.43	2.73 ± 0.46	72.55 ± 1.73	96.83 ± 0.9

Table S3: SGS read statistics for sub-sampled Peregrine & ScriptSeq libraries prepared from *E. coli* RNA.

Prep and Treatment	Qfiltered Reads (millions)	Mapped to K12 (% of Qfilter)	Mapped to rRNA (% of Mapped)	Mapped to CDS (% of Mapped)	Coverage (% of CDS)
Peregrine Ribo-Zero	2.0	91.03	27.38	61.20	99.25
ScriptSeq Ribo-Zero	2.0	96.46	2.57	70.03	97.34

Table S4: Strand specificity of sub-sampled Peregrine and ScriptSeq libraries prepared from *E. coli* RNA.

Prep and Treatment	rRNA	CDS	CDS w/known asRNA
Peregrine Untreated	99.94	95.16	87.06
Peregrine DSN	99.93	94.13	86.94
Peregrine Ribo-Zero	1.48	86.80	81.61
ScriptSeq Untreated	99.59	97.91	94.75
ScriptSeq Ribo-Zero	17.87	97.89	93.50

Table S5: SGS read statistics for Peregrine libraries prepared from human PBMC RNA.

Treatment	Libraries	Raw Reads (millions)	Passed Qfilter (% of Raw)	Mapped to hg19 (% of Qfilter)	Mapped to rRNA (% of Mapped)	Mapped to CDS (% of Mapped)	Coverage (% of CDS)
Untreated	2	21.9 ± 4.9	68.52 ± 10.15	97.29 ± 1.81	80.77 ± 1.58	3.98 ± 0.17	78.32 ± 0.00
DSN	4	26.7 ± 1.1	82.70 ± 1.29	93.17 ± 0.31	7.92 ± 3.33	25.56 ± 1.35	88.01 ± 0.01

Table S6: SGS read statistics for Peregrine libraries prepared from different amounts of human cell line A549 RNA.

Input	Libraries	Raw Reads (millions)	% Passing Qfilter	% Mapped to Human	% CDS Covered	% CDS	% rRNA
100 ng	2	29.7 ± 2.3	78.32 ± 1.93	88.73 ± 0.03	80.70 ± 0.01	33.45 ± 1.20	14.61 ± 2.64
50 ng	2	28.9 ± 4.0	80.43 ± 2.29	88.48 ± 0.35	80.71 ± 0.01	34.69 ± 0.98	8.99 ± 1.39
10 ng	2	32.7 ± 3.1	83.13 ± 0.89	88.18 ± 0.13	80.30 ± 0.00	36.76 ± 0.55	6.30 ± 0.66
1 ng	2	33.3 ± 1.0	16.24 ± 1.22	45.49 ± 1.49	63.12 ± 0.00	6.65 ± 0.55	25.32 ± 3.01