Protocol

NanoCAGE: A High-Resolution Technique to Discover and Interrogate Cell Transcriptomes

Md Salimullah, Sakai Mizuho, Charles Plessy,¹ and Piero Carninci

RIKEN Yokohama Institute, Omics Science Center, Yokohama City, Kanagawa, 230-0045, Japan

INTRODUCTION

Cap analysis gene expression (CAGE) is a method to identify the 5' ends of transcripts, allowing the discovery of new promoters and the quantification of gene activity. Combining promoter location and their expression levels, CAGE data are essential for annotation-agnostic studies of regulatory gene networks. However, CAGE requires large amounts of input RNA, which usually are not obtainable from highly refined samples such as tissue microdissections or subcellular fractions. The nanoCAGE method can capture the 5' ends of transcripts from as little as 10 ng of total RNA and takes advantage of the capacity of current sequencers to produce longer (50-100 bp) reads. The method prepares cap-selected cDNAs ready for direct sequencing of their 5' ends (optionally mated with the 3' end) that can provide information about downstream sequences. This protocol describes how to prepare nanoCAGE libraries from as little as 50 ng of total RNA within two working days. The libraries can be sequenced using an Illumina Gene Amplifier_{IIx} with a level of sensitivity 1000 times higher than CAGE.

RELATED INFORMATION

Instead of the cap-trapper method (Carninci et al. 1996, 1997), the nanoCAGE protocol uses a template-switching method based on the reverse transcription of the cap of the messenger RNA (mRNA) to enrich for 5' ends (Chenchik et al. 1998), as well as a semisuppressive polymerase chain reaction (PCR) approach (Plessy et al. 2010) to minimize short PCR artifacts (Fig. 1). The protocol for nanoCAGE library preparation described here has been comprehensively revised (Table 1) and simplified from the original (Plessy et al. 2010) and no longer uses the time-consuming enzymatic cleavage step that limited the informative length of the tags to ~25 bp.

Amounts of total RNA ranging from 10 to 1250 ng can be used for nanoCAGE library preparation. For best results, use at least 50 ng of total RNA; this protocol is optimized for that amount. For samples where ~50 µg of RNA is available, the original CAGE method (Kodzius et al. 2006) might be preferable, because it is free from PCR bias. For samples where a few hundreds of nanograms of RNA are available, the preparation of technical replicates is recommended for use as a backup, especially if the analysis is centered on the use of the expression levels.

Libraries prepared using different reverse transcriptases require different conditions for efficient template switching; the reverse transcription of the cap on which the switching depends could play a significant role in these differences (Plessy et al. 2010). Comparisons of the fraction of aligned reads corresponding to the ribosomal RNA (rRNA; <10%) and of redundancy (less than three; see Table 1) showed that libraries prepared with PrimeScript contain approximately twice as many reads aligned to promoter regions (as determined from RefSeq gene models) as those prepared using SuperScript III (Fig. 2).

¹Corresponding author (rgscerg@gsc.riken.jp).

Cite as: Cold Spring Harb Protoc; 2011; doi:10.1101/pdb.prot5559

www.cshprotocols.org

CAGE libraries can be used for different types of analyses (Carninci 2010), e.g., promoter discovery (particularly when the available transcript annotation for a tissue or a species is rudimentary), differential expression analysis (Valen et al. 2009), inference of transcription-factor binding sites (Vitezic et al. 2010), or gene networks (Suzuki et al. 2009). Libraries prepared with the nanoCAGE protocol can be paired-end sequenced as CAGEscan libraries (Plessy et al. 2010) that link transcript 5' ends to downstream regions assembled from the mated 3' ends.

FIGURE 1. Flowchart of nanoCAGE protocol. Template-switching oligonucleotide (TS) and reverse transcription (RT) primer are added to the first-strand cDNA synthesis reaction. The three guanosine ribonucleosides at the 3' end of the template-switching primer hybridize in a cap-dependent manner to cytidine deoxynucleosides added to the 3' end of the newly synthesized cDNA strand by the reverse transcriptase (Hirzmann et al. 1993). After hybridization, the reverse transcriptase extends the cDNA strand using the template-switching oligonucleotide as a template. Hence, the cDNA originating from a capped RNA will have a 3' end sequence reverse-complementary to the sequence of the template-switching oligonucleotide. These sequences are necessary to synthesize and amplify the second cDNA strand by semi-suppressive PCR, which minimizes amplification of shorter artifacts (e.g., primer dimers or aberrant cDNAs template-switched from an RT primer or reverse-transcribed from a TS oligo). End sequences required for sequencing using a Genome Analyzer₁₅ are introduced by “library PCR,” and the cDNAs are sequenced as single or paired ends. (For color figure, see doi: 10.1101/pdb.prot5559 online at www.cshprotocols.org.)

MATERIALS

CAUTIONS AND RECIPES: Please see the end of this protocol for appropriate handling of materials marked with <I>, and recipes for reagents marked with <R>.

Reagents

<I>Acetic acid
Agarose gel (1%)
Agencourt AMPure XP (Beckman Coulter A63881)
Agencourt RNAClean XP (Beckman Coulter A63987)
Betaine (5 M; e.g., Wako 023-10862)
DNA loading dye (6X; e.g., MassRuler; Fermentas R0621)
DNA marker: FastRuler DNA ladder, middle range (Fermentas SM1113)
dNTP mixture (2.5 mM; TaKaRa Bio 4030)
<I>DTT (Dithiothreitol; 0.1 M)
<R>EDTA (0.5 M, pH 8.0)
Ethanol (70%)
<R><I>Ethidium bromide (10 mg/mL)
Ex Taq Hot Start Version (TaKaRa Bio RR006A)

High Sensitivity DNA Reagent Kit (Agilent Technologies 5067-4626)

Oligonucleotide, DNA/RNA template-switching, desalted: 5'-TAGTCGAACTGAAGGTCTCCAGCA (rG)(rG)(rG)-3' (1 mM; Integrated DNA Technologies)

All of the bases are deoxynucleotides except for the guanosine ribonucleotides (rG). These are essential and cannot be replaced by deoxyriboguanosines. Although tests with pure deoxyribo-oligonucleotides for template switching could amplify some libraries, the tags obtained did not identify promoters. To multiplex libraries in single sequencing lanes, a sequence identifier ("barcode") can be inserted 5' of the riboguanosines.

Oligonucleotide, library PCR, forward, desalted (Invitrogen): 5'-AATGATACGGCGACCACCGAGA TCTACACTAGTCGAACTGAAGG-3'

Oligonucleotide, library PCR, reverse, desalted (Invitrogen): 5'-CAAGCAGAAGACGGCATAACGAGA TCGGTCTCGGCATTCTGCTGAACCGCTCTCCGATCT-3'

Oligonucleotide, reverse transcription primer, desalted: 5'-TAGTCGAACTGAAGGTCTCCGAACCG CTCTCCGATCTNNNNNN-3' (100 μM; Invitrogen)

The use of oligo-dT reverse transcription primers is not recommended, because they will preferentially amplify short transcripts that originate from 3' promoters (see Fig. 4B,C in Carninci et al. 2006).

Oligonucleotide, forward second-strand PCR, desalted (Invitrogen): 5'-TAGTCGAACTGAAGGTCT CCAGC-3'

Oligonucleotide, reverse second-strand PCR, desalted (Invitrogen): 5'-TGACGTCGTCTAGTCGAA CTGAAGGTCTCCGAACC-3'

Oligonucleotide, sequencing primer, forward, desalted (Invitrogen): 5'-TAGTCGAACTGAAGGTC TCCAGCA-3'

Oligonucleotide, sequencing primer, reverse, desalted (Invitrogen): 5'-CGGTCTCGGCATTCTG CTGAACCGCTCTCCGATCT-3'

Reverse transcriptase (200 U/μL; e.g., PrimeScript; TaKaRa Bio 2680A)

The use of PrimeScript for reverse transcription is strongly recommended: Tests of this protocol using different commercially available reverse transcriptases found that PrimeScript was the best for nanoCAGE library preparation (Fig. 2).

RNA samples, total (50 ng/μL)

Appropriate methods should be used to extract (e.g., Trizol LS) and purify (e.g., PureLink RNA Micro Scale Kit [Invitrogen 12183016]) total RNA for nanoCAGE library preparation. Nuclease-free water (or a suitable elution buffer provided with the RNA extraction kit) should be used to elute the RNA. Test-purified total RNA (>50 ng) using an Agilent Bioanalyzer or a NanoDrop spectrophotometer. We recommend total RNA samples with an RNA integrity number (RIN) value ≥7. Store total RNA samples at -80°C until used.

RNaseZap (Ambion 9780)

<R>Sorbitol/trehalose stock solution

SYBR Premix Ex Taq (TaKaRa Bio RR041A)

The product includes the SYBR Premix Ex Taq (Perfect Real Time) (2X) and the ROX Reference Dye II (50X).

FIGURE 2. Selection of reverse transcriptase for nanoCAGE library preparation. SuperScript III (SSIII) and PrimeScript (PS) give a higher yield of nanoCAGE cDNAs (data not shown). Libraries were evaluated based on the percentage of aligned reads corresponding to the rRNA and their redundancy (see Table 1 for details). SSIII worked best at 50°C in the presence of sorbitol and trehalose, whereas PS worked best at 40°C in the presence of betaine. Using conditions optimized for each enzyme, SSIII and PS were compared further after alignment of the reads to RefSeq gene models.

Tween 20 (Sigma P9416) (0.1% [v/v], prepared in nuclease-free water)
Water, nuclease-free (e.g., Invitrogen 10977-015)

Equipment

Bioanalyzer (e.g., Agilent 2100; Agilent Technologies)
Centrifugal concentrator (e.g., TOMY)
Centrifuge, benchtop, adapted for use with 96-well plates (e.g., Allegra X-12R; Beckman Coulter)
Electrophoresis unit (e.g., Mupid- α or equivalent)
Gel maker set
Genome Analyzer _{ix} (Illumina)
This protocol was designed for use with an Illumina platform. However, it can be adapted for other sequencing platforms by designing suitable primers.
Gloves
Ice-water bath
Imaging system (e.g., BioDoc-It; UVP)
Magnetic stand for bead separation (e.g., Dynal)
Micropipettors
PCR plates, 96-well (e.g., μ ltraAmp; Sorenson BioScience)
PCR seals, adhesive (e.g., 4ti-0500; 4titude)
PCR system, real-time (e.g., StepOnePlus; Applied Biosystems)
Pipette tips (e.g., Multifit; Sorenson BioScience)
Sealant applicator (e.g., PA1; 3M)
Spectrophotometer (e.g., NanoDrop 1000 [Thermo Fisher] or equivalent)
Thermocycler
Tubes, microcentrifuge, low-retention, 1.5-mL and 0.6-mL
Tubes, PCR, siliconized, 0.2-mL (e.g., FastGene)
<!--UV (ultraviolet) light source
Water purification system (e.g., Milli-Q Advantage A10; Millipore)

METHOD

Before investigating valuable samples, beginning investigators should practice making control libraries with some available total RNA to develop their abilities to check cDNA synthesis by real-time PCR, to visualize semisuppressive PCR smears on gels, and to interpret Bioanalyzer profiles of prepared libraries.

First-Strand cDNA Synthesis

This process produces first-strand cDNAs from total RNA by random priming; the 5' ends are captured through template switching. Perform all procedures in RNase-free conditions on a workbench dedicated only to RNA work.

1. Mix 8 μ L of sorbitol/trehalose stock solution, 1 μ L of 100 μ M reverse transcription primer, and 1 μ L of 1 mM template-switching DNA/RNA oligonucleotide.
Sorbitol/trehalose solutions are viscous; mix very carefully by pipetting at least 10 times.
2. Aliquot 1 μ L each of the primer solution from Step 1 into separate 0.2-mL siliconized PCR tubes for each sample and for a negative control.
3. Add 1 μ L of total RNA (i.e., 50 ng) into the sample tubes. Add 1 μ L of nuclease-free water into the control tube. Mix the solutions several times for complete homogenization.
If the total RNA is in a larger volume, reduce to 2 μ L by centrifugal evaporation at room temperature (to avoid RNA degradation). The presence of sorbitol/trehalose will slow evaporation as the concentration increases, reducing the risk of accidentally drying the RNA.
4. Incubate the solutions at 65°C for 10 min in a thermocycler.

Table 1. Comprehensive testing of different nanoCAGE parameters

Library name	Reverse transcriptase (RT)	PCR ext.time (sec) ^a	RT temp (°C)	RT additive ^b	RT time (min)	RT cooltemp (°C) ^c	Total tags ^d	Unique tags ^e	rRNA (%) ^f	Redundancy ^g	Semisuppr. PCRcycles ^h	Semisuppr. PCRconc. (ng/μL) ⁱ
J53-GA	PrimeScript	120	50	betaine	30	---	1,572,426	181,750	23.23	8.65	25	16.34
				S/T	30	---	2,322,991	359,674	29.35	6.46	24	15.49
				water	30	---	723,881	108,291	27.09	6.68	25	12.47
			60	betaine	30	---	1,098,557	40,599	5.13	27.06	32	13.08
				S/T	30	---	1,917,298	83,125	4.27	23.07	33	14.22
				water	30	---	2,150,491	68,244	5.74	31.51	33	11.73
		30	50	betaine	30	---	1,130,096	112,900	19.64	10.01	28	9.25
				S/T	30	---	2,359,863	202,012	26.46	11.68	28	8.17
				water	30	---	1,631,842	125,415	24.89	13.01	30	8.85
			60	betaine	30	---	665,230	48,412	5.09	13.74	32	9.28
				S/T	30	---	1,407,092	34,991	4.93	40.21	37	7.64
				water	30	---	607,120	20,445	8.39	29.7	37	9.35
J54-GA	SuperScript III	120	50	betaine	30	---	250,499	103,388	16.86	2.42	23	28.8
J54-GA	SuperScript III	120	50	S/T	30	---	1,224,346	215,729	9.77	5.68	23	26.01
				water	30	---	955,875	300,447	20.45	3.18	24	33.99
			60	betaine	30	---	927,392	119,713	12.76	7.75	26	20.3
				S/T	30	---	205,397	43,212	16.31	4.75	27	20.42
				water	30	---	1,159,443	109,583	10.19	10.58	27	17.89
			30	betaine	30	---	574,609	158,085	15.67	3.63	28	32.93
				S/T	30	---	859,521	244,124	11.17	3.52	27	14.27
				water	30	---	508,443	155,528	22.29	3.27	28	13.11
			60	betaine	30	---	924,509	68,005	3.12	13.59	34	13.46
				S/T	30	---	881,722	119,491	6.39	7.38	34	16.9
				water	30	---	1,842,304	99,336	3.12	18.55	33	21.13
J55-GA	SuperScript II	120	50	betaine	30	---	1,210,696	240,266	14.36	5.04	25	26.78
				S/T	30	---	2,263,452	443,357	12.67	5.11	25	29.55
				water	30	---	1,678,459	363,513	20.77	4.62	24	22.81
			60	betaine	30	---	1,617,617	126,028	5.52	12.84	30	16.12
				S/T	30	---	549,364	42,617	3.28	12.89	30	14.5
				water	30	---	431,085	33,658	7.8	12.81	30	14.96
		30	50	betaine	30	---	768,911	182,116	14.81	4.22	28	13.69
				S/T	30	---	811,168	238,256	18.97	3.4	28	15.01
				water	30	---	2,034,725	324,021	18.14	6.28	30	15.84
			60	betaine	30	---	1,474,825	87,977	3.38	16.76	36	10.76
				S/T	30	---	783,407	47,124	4.63	16.62	36	12.53
				water	30	---	1,066,288	57,003	4.34	18.71	36	8.27
J61-GA	SuperScript III	120	40	betaine	30	---	1,509,574	489,529	21.48	3.08	25	51
				S/T	30	---	1,241,620	311,979	23.19	3.98	25	29.08
				water	30	---	1,660,969	499,814	26.79	3.32	25	58.91
			45	betaine	30	---	2,314,533	588,948	31.62	3.93	25	36.68
				S/T	30	---	1,194,233	340,681	22.45	3.51	25	36.03
				water	30	---	1,366,263	332,132	17.32	4.11	25	28.62
		30	55	betaine	30	---	1,282,996	181,602	24.66	7.06	27	27.52
				S/T	30	---	896,055	227,808	23.12	3.93	24	17.3
				water	30	---	1,336,866	262,873	22.77	5.09	27	28.33
			40	betaine	30	---	940,078	301,506	22.53	3.12	28	22.36
				S/T	30	---	1,294,643	443,824	21.7	2.92	28	27.6
				water	30	---	1,843,239	473,334	27.17	3.89	28	32.47
		45	45	betaine	30	---	743,362	191,280	22.1	3.89	30	18.71
				S/T	30	---	951,527	257,689	22.14	3.69	30	22.99
				water	30	---	695,530	226,251	16.97	3.07	30	27.02

(continued)

Table 1. *Continued*

J62-GA	SuperScript II	120	55	betaine	30	---	973,648	159,747	30.95	6.09	32	17.87			
				S/T	30	---	1,201,971	208,873	32	5.75	30	19.96			
				water	30	---	1,410,042	165,453	32.18	8.52	30	17.52			
			40	betaine	30	---	94,717	67,342	15.42	1.41	28	51.91			
				S/T	30	---	143,954	91,610	8.71	1.57	26	75.86			
				water	30	---	177,528	117,916	15.15	1.51	25	42.51			
			45	betaine	30	---	190,529	99,655	12.64	1.91	25	37.96			
				S/T	30	---	48,803	30,511	3.6	1.6	22	31.54			
				water	30	---	188,310	121,086	23.04	1.56	25	53.77			
			55	betaine	30	---	157,084	72,413	2.71	2.17	27	11.16			
				S/T	30	---	15,840	10,994	2.5	1.44	27	16.92			
				water	30	---	78,471	47,392	1.38	1.66	27	15.22			
			30	40	betaine	30	---	273,149	188,191	10.14	1.45	25	23.89		
					S/T	30	---	249,603	186,982	9.45	1.33	25	18.95		
					water	30	---	318,088	245,244	13.19	1.3	25	28.14		
			45	betaine	30	---	124,887	74,796	7.21	1.67	25	16.8			
				S/T	30	---	94,417	69,055	13.34	1.37	25	19.86			
				water	30	---	34,632	24,388	3.91	1.42	25	18.91			
			J63-GA	PrimeScript	120	55	betaine	30	---	395,583	132,449	5.75	2.99	34	16.49
							S/T	30	---	137,211	65,220	16.68	2.1	30	17.46
							water	30	---	96,976	25,855	3.48	3.75	30	11.43
						40	betaine	30	---	1,120,560	516,165	8	2.17	22	27.87
							S/T	30	---	929,189	442,893	11.74	2.1	22	25.11
							water	30	---	1,305,601	601,415	10.01	2.17	22	28.33
45	betaine	30				---	1,368,533	580,968	17.64	2.36	22	18.73			
	S/T	30				---	649,456	318,800	15.27	2.04	22	20.76			
	water	30				---	739,001	326,532	18.53	2.26	22	19.22			
55	betaine	30				---	1,003,632	173,383	22.31	5.79	29	25.95			
	S/T	30				---	1,217,507	303,138	17.41	4.02	30	42.3			
	water	30				---	1,987,403	401,085	17.62	4.96	30	25.45			
30	40	betaine				30	---	873,112	428,106	16.26	2.04	27	22.09		
		S/T				30	---	930,354	443,523	17.4	2.1	27	20.09		
		water				30	---	848,335	426,379	18.12	1.99	27	30.32		
45	betaine	30				---	574,837	269,437	21.96	2.13	28	26.98			
	S/T	30				---	881,363	370,124	20.19	2.38	28	23.45			
	water	30				---	312,482	151,889	13.84	2.06	28	17.7			
55	betaine	30				---	1,170,345	200,170	12.99	5.85	38	42.86			
	S/T	30				---	586,387	127,215	34.81	4.61	30	14.94			
	water	30				---	965,347	119,385	9.25	8.09	40	37.31			
J64-GA	SuperScript II	120				50	betaine	30	30	917,523	355,315	26.38	2.58	25	38.37
									25	908,191	371,519	33.06	2.44	25	49.14
									20	787,615	320,633	27.46	2.46	25	30.73
					15	987,514	348,994	22.04	2.83	25	25.71				
					45	1,387,713	484,912	33.68	2.86	25	31.38				
					60	857,554	334,155	39.52	2.57	21	18.69				
			SuperScript III	120	50	betaine	30	30	722,783	263,787	28.9	2.74	25	55.85	
								25	834,995	286,939	24.78	2.91	25	45.05	
								20	1,396,843	391,729	43.93	3.57	25	46.16	
							15	1,400,473	347,451	37.41	4.03	27	51.23		
							45	1,102,061	320,173	48.44	3.44	24	45.75		
							60	776,672	266,856	39.35	2.91	24	55.63		
	PrimeScript	120	50	betaine	30	30	1,017,950	274,698	34.21	3.71	28	42.12			
						25	907,559	219,400	47.38	4.14	27	18.09			
						20	705,686	190,585	31.01	3.7	28	29			
					15	1,285,912	270,461	42.59	4.75	28	29.72				
					45	1,143,364	259,958	54.18	4.4	23	20.19				
					60	1,191,776	196,585	27.24	6.06	26	25.93				

(continued)

Table 1. *Continued*

Note: The table lists quality indicators for nanoCAGE libraries made with 50 ng of Hep G2 total RNA using different reverse transcriptases and reaction conditions. The libraries were sequenced on an Illumina Genome Analyzer_{IIx}. The rows in bold show conditions that gave the best results with SuperScript III and PrimeScript.

^aExtension time for the semisuppressive PCR and the PCR to add sequencing adapters. Because the analyzer does not efficiently sequence cDNAs >1 kb in mixtures, short incubation times were tested to reduce the size of the amplified cDNAs.

^bS/T: sorbitol/trehalose; water: nuclease-free water.

^cAfter the RT reaction, the cDNAs were incubated for 5 min at the indicated temperatures before cooling in an ice-water bath (see Steps 8-9). "----" indicates that the cool-down treatment was not applied to these combinations.

^dNumber of tags remaining after removal of artifacts, primer dimers, etc., from the raw sequences.

^eNumber of unique tags remaining after removal of artifacts, primer dimers, etc., from the raw sequences.

^fPercentage of total tags mapped to ribosomal RNA.

^gTotal number of tags divided by the number of unique tags.

^hNumber of PCR cycles required to amplify the purified cDNAs by semisuppressive PCR (see Step 17).

ⁱConcentration of semisuppressive PCR products after purification (see Step 24).

5. Prepare 8 µL of the reverse transcription reaction mixture per reaction:

2 µL of 5X PrimeScript buffer

2.5 µL of 2.5 mM dNTPs

1 µL of 0.1 M DTT

1.5 µL of 5 M betaine

1 µL of 200 U/µL PrimeScript

Reverse transcription reactions in the presence of betaine produce better nanoCAGE library outputs (Table 1).

6. Pause the PCR machine at 22°C. Snap-cool the samples (from Step 4) in an ice-water bath for 2 min.

RNAs must not refold, and cooling occurs faster in an ice-water bath than on ice.

7. Return the samples to 22°C in the thermocycler.

8. Add 8 µL of the reverse transcription reaction mixture (from Step 5) to the samples.

9. Incubate for 10 min at 22°C, 30 min at 40°C, and 15 min at 75°C in the thermocycler.

10. Immediately after finishing the reaction, snap-cool in an ice-water bath for 2 min.

Checking the reverse transcription is not recommended at this step, because this will reduce the amount of cDNAs for PCR amplification and therefore increase the number of cycles needed.

11. Purify the first-strand cDNAs:

This removes smaller artifacts and primer dimers from the cDNAs that can make it difficult to determine the number of PCR cycles required for semisuppressive PCR. Again, checking the purification at this step is not recommended.

i. Use an Agencourt RNAClean XP kit according to the manufacturer's instructions; mix the cDNA and beads with pipettes.

ii. Wash with 100 µL of 70% ethanol.

Before eluting, remove as much ethanol as possible. Do not let the beads dry, because this can reduce recovery.

iii. Elute with 40 µL of nuclease-free water.

Quantitative Real-Time PCR

Quantitative real-time PCR is performed on a small scale to determine the number of cycles required for large-scale synthesis. This is necessary to keep the number of PCR cycles as low as possible (to reduce bias) and to suppress the synthesis of smaller artifacts.

12. For each sample, prepare 8.5 μL of real-time PCR mixture:
 - 5 μL of 2X SYBR Premix Ex Taq
 - 0.1 μL of 10 μM forward second-strand PCR primer
 - 0.1 μL of 10 μM reverse second-strand PCR primer
 - 0.2 μL of 50X ROX Reference Dye II
 - 3.1 μL of nuclease-free water

Perform real-time PCR in triplicate for the samples and control.
13. Prepare one ultraAmp PCR plate. Add 8.5 μL of real-time PCR mixture per well.
14. Add 1.5 μL of purified cDNA or negative control (from Step 11.iii) to each well.
15. Seal the plate with adhesive PCR seal. Ensure a tight seal with a sealing applicator.
Centrifuge briefly if any solution has splashed above the mixture.
16. Perform quantitative real-time PCR using a StepOnePlus system according to the manufacturer's instructions, with the following conditions:
 - i. 95°C for 1 min
 - ii. 40 cycles of 95°C for 15 sec, 65°C for 10 sec, 68°C for 2 min
 - iii. Hold at 4°C

Use "Comparative Ct ($\Delta\Delta\text{Ct}$)" for quantification.
17. Determine the optimum number of PCR cycles required for second-strand cDNA synthesis:
 - i. Determine the "Cycle threshold: (CT)" values for each sample and control reaction after real-time PCR (Fig. 3).
See Troubleshooting.
 - ii. Calculate the average CT values for each sample and for the negative control.
The average CT value for purified cDNA usually is ~16-20, depending on the source of RNA.
 - iii. Add 4 to the average CT value.
The optimum cycle number should be lower than 25 for the samples and very low or no signal for the negative control.

FIGURE 3. Determination of the optimum number of PCR cycles for synthesis of second-strand cDNAs by semisuppressive PCR. A cycle threshold (CT) value is determined by quantitative real-time PCR as the number of cycles required for the fluorescent signal to cross the threshold, i.e., to exceed background levels. The optimum number of cycles for large-scale synthesis is the average CT value plus four.

Second-Strand cDNA Synthesis and Amplification by Semisuppressive PCR

This procedure synthesizes the nanoCAGE cDNA libraries.

18. Prepare 85 μ L of semisuppressive PCR mixture per reaction:

10 μ L of 10X *Premix Ex Taq* Hot Start buffer
8 μ L of dNTPs (2.5 mM each)
1 μ L of 10 μ M forward second-strand PCR primer
1 μ L of 10 μ M reverse second-strand PCR primer
0.5 μ L of 5 U/ μ L *Premix Ex Taq* Hot Start
64.5 μ L of nuclease-free water

Amplify purified samples in duplicate; a single 100- μ L sample is sufficient for the purified control.

The competition between primer annealing and self-annealing is the key mechanism of PCR suppression (Chenichik et al. 1998). The concentration of the primers is thus lower than usual to strengthen the suppressive effect.

19. Add 85 μ L of semisuppressive PCR mixture to each PCR tube.
20. Add 15 μ L of purified cDNA or negative control (from Step 11.iii) to each PCR tube.
Store the remaining purified cDNAs at -20°C.
21. Perform semisuppressive PCR using the following conditions:
- i. 95°C for 1 min
 - ii. *n* (optimum PCR cycle number) cycles of 95°C for 15 sec, 65°C for 10 sec, 68°C for 2 min
 - iii. Hold at 4°C

*The optimum cycle number (*n*, as determined in Step 17.iii) is usually ~20-24 cycles, depending on the source of RNA.*

PCR tubes can be left overnight in the PCR machine.

Purification of Amplified cDNAs

Perform purification and PCR at different locations and use different equipment to avoid cross-contamination.

22. Purify the amplified cDNA libraries:
- i. For each sample, pool the PCR solutions from the duplicate tubes.
Do not purify the negative control. It will be required for confirmation that the primer dimers and short PCR artifacts were eliminated from the amplified cDNAs (see Step 23.i).
 - ii. Use an Agencourt AMPure XP kit according to the manufacturer's instructions; mix the cDNA and beads with pipettes.
 - iii. Wash with 600 μ L of 70% ethanol.
Before eluting, remove as much ethanol as possible. Do not let the beads dry, because this can reduce recovery.
 - iv. Elute with 30 μ L of nuclease-free water.
23. Confirm PCR amplification and purification by agarose gel electrophoresis:
- i. Mix 2 μ L of the purified nanoCAGE cDNAs with 3 μ L of nuclease-free water and 1 μ L of 6X loading dye. Mix 5 μ L of nonpurified control (see Step 22.i.) with 1 μ L of 6X loading dye.
 - ii. Load the mixtures on a 1% agarose gel. Load 4 μ L of a DNA marker.
 - iii. Electrophorese until the markers separate in the range of from 100 bp to 2 kb (e.g., for agarose minigels, 15-30 min).

iv. Visualize the gel under UV light (see, e.g., Fig. 4).

Primer dimers or artifacts should not be present in the purified cDNAs. Confirm their elimination by comparing the purified cDNAs with the nonpurified negative control.

See Troubleshooting.

24. Determine the concentration of the purified cDNAs in triplicate using a NanoDrop spectrophotometer according to the manufacturer's instructions. Calculate the average value.

The concentration of the purified semisuppressive cDNA libraries should be >20 ng/μL to avoid artificial redundancy (see Discussion).

See Troubleshooting.

25. Dilute the samples with nuclease-free water to 10 ng/μL.

Addition of Sequencing Adapters by PCR

This procedure adds adapter sequences to the nanoCAGE libraries for binding to the genome analyzer's flow cell and for amplification by bridge PCR.

26. Prepare 92 μL of PCR mixture per reaction:

10 μL of 10X *Premix Ex Taq* Hot Start buffer

8 μL of dNTPs (2.5 mM each)

2 μL of 10 μM forward library PCR primer

2 μL of 10 μM reverse library PCR primer

0.5 μL of 5 U/μL *Premix Ex Taq* Hot Start

69.5 μL of nuclease-free water

Prepare reactions in triplicate for each sample. Also prepare a control reaction without added template.

The concentration of PCR primers is lower than usual because of the small number of cycles, which results in a large leftover.

27. Add 92 μL PCR reaction mixture to each tube.

28. Add 8 μL (i.e., 80 ng) of purified DNA (from Step 25) to each sample tube. Add 8 μL nuclease-free water to the control.

29. Perform PCR using the following conditions:

i. 95°C for 1 min

ii. 1 cycle of 95°C for 15 sec, 55°C for 10 sec, 68°C for 2 min

iii. 6 cycles of 95°C for 15 sec, 65°C for 10 sec, 68°C for 2 min

iv. Hold at 4°C

30. Purify the library PCR products:

i. For each sample, pool the PCR solutions from the triplicate tubes.

Do not purify the control.

ii. Use an Agencourt AMPure XP kit according to the manufacturer's instructions; mix the cDNA and beads with pipettes.

FIGURE 4. Confirmation of semisuppressive PCR amplification and purification. Electrophoresis was performed on a 1% agarose gel at 100 V for 15 min. The purified nanoCAGE library (L) is visible as a smear with small molecular weight artifacts. The nonpurified negative control (NC) is used to visualize the artifacts before removal. (M) DNA ladder, middle range.

iii. Wash with 900 μ L of 70% ethanol.

Before eluting, remove as much ethanol as possible. Do not let the beads dry, because this can reduce recovery.

iv. Elute with 30 μ L of 0.1% Tween 20 in nuclease-free water.

31. Determine the concentration of the purified libraries in triplicate using a NanoDrop spectrophotometer according to the manufacturer's instructions. Calculate the average value.

The concentration of the purified library PCR products is usually >20 ng/ μ L (typically, ~20-30 ng/ μ L). Store the prepared nanoCAGE libraries at -20°C until sequenced.

Analysis of nanoCAGE Libraries

It is important to check the removal of PCR primer dimers/artifacts, as well as to know the profile of the amplified libraries (e.g., size distribution, molar concentration) to calculate the amount of molecules to apply in the sequencing reaction.

32. Dilute 1 μ L of each purified library to 5 ng/ μ L with 0.1% Tween 20.

33. Analyze 1 μ L each of the diluted samples using a High Sensitivity DNA Reagent Kit according to the manufacturer's instructions.

Analyze the libraries in triplicate.

See Troubleshooting.

34. Determine the nanomolar concentration of the library from the Bioanalyzer data (Fig. 5):

This protocol assumes that only cDNAs in the 200-700-bp range are efficiently sequenced in the Gene Analyzer_{IIx} sequencer. There should be no peak shorter than 150 bp visible on the Bioanalyzer data.

- Open the library DNA analysis file (.xad).
- Double-click on the electropherogram data of the sample to be analyzed.
- Click on the "Region table" icon under the page. Scroll into the table.
- Right-click the mouse button. Select "Modify region."
- Input 200 and 700 in the boxes "From[bp]" and "To[bp]", respectively.
- Click the "OK" button to determine the molarity.

FIGURE 5. Bioanalyzer profile of a typical nanoCAGE library (5 ng/ μ L). This protocol assumes that only cDNAs in the 200-700-bp range are sequenced efficiently. Longer templates are likely to be outcompeted by shorter ones for binding to the flow cell; they also could be difficult to amplify using the bridge PCR.

- vii. Calculate the average concentration of the triplicate assays.
 - viii. Multiply this concentration with the dilution factor (from Step 32) to obtain the original library concentration.
35. Sequence the nanoCAGE libraries at a final concentration of 15 pM using an Illumina Genome Analyzer_{ix} platform for single or paired-end reads.
- One lane on this sequencer generally yields more than 15 million reads. Although not tested, a higher number of readings is possible with other systems (e.g., HiSeq2000 [Illumina]).*

TROUBLESHOOTING

Problem: CT value is lower than usual (i.e., 16-20).

[Step 17.i]

Solution: Perform a small-scale PCR using the extra purified cDNA solution (from Step 20) as described in Steps 18-21; scale the reaction volumes down to one-fifth (i.e., a final volume of 20 μ L). Take aliquots after 20, 22, and 24 cycles. Confirm amplification and detect optimal cycle number by running aliquots of the products on a 1% agarose gel. The number of cycles to be used for the large-scale reaction is defined by the cycle where a signal becomes visible on the gel for a sample, while the negative control still has no detectable products.

Problem: CT value is higher than usual (e.g., >20).

[Step 17.i]

Solution: The most probable causes are partial degradation of RNA, impurities or inhibitory substances in the RNA, reverse transcriptase inactivation, or a high concentration ($\sim$ 50 ng/ μ L) of the smaller (<200-bp) fraction. Consider the following:

1. If possible, recheck the RNA quality with a Bioanalyzer.
2. Check the reverse transcriptase using a control template.
3. Check for any possible source of contamination during the reverse transcription preparation.
4. Perform the reverse transcription reaction within the shortest time possible.

Problem: Primer dimers or short artifacts could not be removed from the library.

[Steps 23.iv and 33]

Solution: This can occur with nanoCAGE libraries begun with <50 ng of total RNA. Repurification using the AMPure XP Kit should solve this problem.

Problem: The concentration of the purified semisuppressive PCR library is <20 ng/ μ L.

[Step 24]

Solution: Although the library could be sequenced, it might be highly redundant (see Discussion). Therefore, if a backup sample is available, remaking the library is recommended. Alternatively, the library can be used for promoter discovery (see Discussion).

DISCUSSION

The transcripts detected by nanoCAGE are strand specific. Once prepared, the libraries can be sequenced for both 5'-end single and paired-end reads. Single-read sequencing can pinpoint the gene's product (i.e., the mRNA molecule) from the transcription start site, whereas paired-end sequencing can link start sites to downstream sequences, providing insight into the architecture of transcripts and thus into their possible functions (Plessy et al. 2010). Currently, the ability to fully scan the 3' end of long transcripts is limited by the length of the paired-end reads, although further developments in sequencing technology could overcome this limitation.

NanoCAGE libraries can be prepared from <50 ng of total RNA, although with a risk of molecular bottlenecks and the need for a higher number of PCR cycles (see below). NanoCAGE libraries have been successfully prepared from nonpolyadenylated or polysomal RNA, RNA extracted from histological sections that were stained by immunohistochemistry, and samples microdissected with laser capture. In general, nanoCAGE will also work on degraded RNA at the progressive expense of its

accuracy in detecting promoters. NanoCAGE even detects gene expression using decapped and fragmented samples, albeit similarly to RNA-seq protocols (Plessy et al. 2010). It might require a higher number of PCR cycles (>30) to prepare nanoCAGE libraries from total RNA samples that contain a large fraction of small RNAs (<150 bp), e.g., when their concentration is within an order of magnitude of the oligonucleotides used in the reverse transcription.

Depending on the source of total RNA, the fraction of sequencing reads that align to rRNA can be higher than usual (>8%). Because rRNA is not capped, a high number of reads matching rRNA suggests that the template-switching reaction was not highly specific, and that the detection of the promoters thus will not be accurate at a base-pair level. Some cell lines contain more rRNA than others, and some RNA samples contain a larger fraction of rRNAs (polysomal RNA, nonpolyadenylated RNA, etc.); nanoCAGE libraries prepared from these samples will always contain more reads matching the ribosomal DNA. Conversely, for many organisms, a full-length sequence of the ribosomal DNA-repeated unit is not available, and the percentage of reads matching rRNA will not be an accurate measurement of the efficiency of template switching.

The hallmark of CAGE (Kodzius et al. 2006; Valen et al. 2009; Carninci 2010) is the addition of extra guanosines at the 5' end of the tags. However, in nanoCAGE these guanosines become part of the 5' linker, and most of them are removed as linker sequence. Nevertheless, in most cases mismatches at the 5' end of the tags are expected to be guanosines. Results contrary to this could indicate a problem in the library, e.g., degradation of the starting material resulting in a much lower fraction of capped 5' ends.

A high redundancy usually indicates a molecular bottleneck, where only a small fraction of the original mRNA molecules contribute reads in the libraries. Libraries where the concentration of purified cDNAs after semisuppressive PCR is <20 ng/μL often have a redundancy higher than 5 (Table 1). This could result from preferential amplification of a subset of templates or by initial rarity of PCR templates, e.g., because of low efficiency in reverse transcription or template switching. Libraries with a high redundancy, nevertheless, can still be used for promoter discovery. Keep in mind, however, that rare transcripts might not appear, and that the shape of the transcription start site could become artificially sharpened, even with broad promoters (Carninci et al. 2006).

The nanoCAGE method has the ability to grasp the complexity of the promoter landscape from fewer cells without the need to rely on predetermined gene models such as RefSeq. NanoCAGE libraries prepared using this protocol can be exploited for new applications in drug screening, biopsy analysis, and whole-transcriptome association studies.

ACKNOWLEDGMENTS

This work was funded by a Grant-in-Aid for Scientific Research (A) 20241047 to P.C., a grant of the 7th Framework of the European Union commission to P.C. (Dopaminet), U.S. National Human Genome Research Institute grant U54 HG004557 to P.C., and Research Grant for RIKEN Omics Science Center from the Japanese Ministry of Education, Culture, Sports, Science and Technology to Y. Hayashizaki. We acknowledge RIKEN GeNAS for sequencing the libraries J53-GA to J55-GA and J61-GA to J65-GA using the Genome Analyzer_{IIx}, as well as for subsequent data processing.

REFERENCES

- Carninci P. 2010. Tagging transcription start sites with CAGE. In *Cap-analysis gene expression (CAGE): The science of decoding gene transcription* (ed. P. Carninci), pp. 7–20. Pan Stanford, Singapore.
- Carninci P, Kvam C, Kitamura A, Ohsumi T, Okazaki Y, Itoh M, Kamiya M, Shibata K, Sasaki N, Izawa M, et al. 1996. High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics* **37**: 327–336.
- Carninci P, Westover A, Nishiyama Y, Ohsumi T, Itoh M, Nagaoka S, Sasaki N, Okazaki Y, Muramatsu M, Schneider C, et al. 1997. High efficiency selection of full-length cDNA by improved biotinylated cap trapper. *DNA Res* **4**: 61–66.
- Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CAM, Taylor MS, Engström PG, Frith MC, et al. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38**: 626–635.
- Chenck A, Zhu YY, Diatchenko L, Li R, Hill J, Siebert PD. 1998. Generation and use of high quality cDNA from small amounts of total RNA by SMART™ PCR. In *Gene cloning and analysis by RT-PCR* (ed. P. Siebert, J. Larrick), pp. 305–319. Eaton, Natick, MA.
- Hirzmann J, Luo D, Hahnen J, Hobom G. 1993. Determination of messenger RNA 5'-ends by reverse transcription of the cap structure. *Nucleic Acids Res* **21**: 3597–3598.
- Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, Tagami M, Sasaki D, Imamura K, Kai C, Harbers M, et al. 2006. CAGE: Cap analysis of gene expression. *Nat Methods* **3**: 211–222.
- Plessy C, Bertin N, Takahashi H, Simone R, Salimullah M, Lassmann T, Vitezic M, Severin J, Olivarius S, Lazarevic D, et al. 2010. Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEScan. *Nat Methods* **7**: 528–534.
- Suzuki H, Forrest AR, van Nimwegen E, Daub CO, Balwierz PJ, Irvine KM, Lassmann T, Ravasi T, Hasegawa Y, de Hoon MJ, et al. 2009. The transcriptional network that controls growth arrest and dif-

ferentiation in a human myeloid leukemia cell line. *Nat Genet* **41**: 553–562.
Valen E, Sandelin A, Winther O, Krogh A. 2009. Discovery of regulatory elements is improved by a discriminatory approach. *PLoS Comput Biol* **5**: e1000562. doi: 10.1371/journal.pcbi.1000562.

Vitezic M, Lassmann T, Forrest ARR, Suzuki M, Tomaru Y, Kawai J, Carninci P, Suzuki H, Hayashizaki Y, Daub CO. 2010. Building promoter aware transcriptional regulatory networks using siRNA perturbation and deepCAGE. *Nucleic Acids Res* doi: 10.1093/nar/gkq729.

RECIPES

[NOTE: Recipes for reagents marked with the <R> symbol not listed below can be found online at <http://www.cshprotocols.org/recipes>.]

EDTA

<R>EDTA (ethylenediaminetetraacetic acid)
<R>NaOH

To prepare EDTA at 0.5 M (pH 8.0): Add 186.1 g of disodium EDTA•2H₂O to 800 mL of H₂O. Stir vigorously on a magnetic stirrer. Adjust the pH to 8.0 with NaOH (~20 g of NaOH pellets). Adjust the volume to 1 L with H₂O. Dispense into aliquots and sterilize by autoclaving. The disodium salt of EDTA will not go into solution until the pH of the solution is adjusted to ~8.0 by the addition of NaOH.

Ethidium bromide

Add 1 g of ethidium bromide to 100 mL of H₂O. Stir on a magnetic stirrer for several hours to ensure that the dye has dissolved. Wrap the container in aluminum foil or transfer the 10 mg/mL solution to a dark bottle and store at room temperature.

Sorbitol/trehalose stock solution

MATERIALS

Reagents

Prepare all reagents with nuclease-free H₂O.

Chelex 100 resin (Sigma C7901)

D-Sorbitol (Wako 198-03755) (0.89 g/mL [4.9 M])

Autoclave for 30 min at 121°C.

D-(+)-Trehalose dihydrate (Fluka/Sigma-Aldrich 90208) (0.727 g/mL)

Autoclave for 30 min at 121°C.

Equipment

Centrifuge, benchtop (e.g., Allegra X-12R; Beckman Coulter)

Centrifuge tubes (15-mL)

Pipettes, disposable

Vortex mixer

METHOD

Perform all procedures under RNase-free conditions.

1. Add 5 mL of the saturated trehalose solution to a tube containing 10 mL of 4.9 M sorbitol solution.
2. Using a disposable pipette, add <1 cm of Chelex 100 resin to the tube. Vortex well.
3. Incubate the solution for 3 h at room temperature (25°C) without further agitation.
4. Centrifuge at 2000 rpm for 10 min.
5. Transfer the supernatant to a fresh tube. Store for up to 6 mo at –20°C.

The final solution is 3.3 M sorbitol/0.66 M trehalose.

Solutions deionized with Chelex 100 can turn yellow even after centrifugation, but this does not affect their use in nanoCAGE library production.

CAUTIONS

[**NOTE:** For reagents marked with the <!--> symbol not listed below, please consult the manufacturer's Material Safety Data Sheet for further information.]

Acetic acid (concentrated) must be handled with great care. It may be harmful by inhalation, ingestion, or skin absorption. Wear appropriate gloves and goggles. Use in a chemical fume hood.

DTT (Dithiothreitol) is a strong reducing agent that emits a foul odor. It may be harmful by inhalation, ingestion, or skin absorption. When working with the solid form or highly concentrated stocks, wear appropriate gloves and safety glasses and use in a chemical fume hood.

EDTA (Ethylenediaminetetraacetic acid) may be harmful by inhalation, ingestion, or skin absorption. Wear appropriate gloves and safety glasses. Severe overexposure can result in death.

Ethidium bromide is a powerful mutagen and is toxic. Consult the local institutional safety officer for specific handling and disposal procedures. Avoid breathing the dust. Wear appropriate gloves when working with solutions that contain this dye.

NaOH (Sodium hydroxide) and solutions containing NaOH are highly toxic and caustic and should be handled with great care. Wear appropriate gloves and a face mask. All concentrated bases should be handled in a similar manner.

UV light and/or UV radiation is dangerous and can damage the retina of the eyes. Never look at an unshielded UV light source with naked eyes. Examples of UV light sources that are common in the laboratory include handheld lamps and transilluminators. View only through a filter or safety glasses that absorb harmful wavelengths. UV radiation is also mutagenic and carcinogenic. To minimize exposure, make sure that the UV light source is adequately shielded. Wear protective appropriate gloves when holding materials under the UV light source.

Cold Spring Harbor Protocols

NanoCAGE: A High-Resolution Technique to Discover and Interrogate Cell Transcriptomes

Md Salimullah, Sakai Mizuho, Charles Plessy and Piero Carninci

Cold Spring Harb Protoc; doi: 10.1101/pdb.prot5559

Email Alerting Service

Receive free email alerts when new articles cite this article - [click here](#).

Subject Categories

Browse articles on similar topics from *Cold Spring Harbor Protocols*.

[Analysis of Gene Expression](#) (197 articles)
[Analysis of Gene Expression, general](#) (129 articles)
[Bioinformatics/Genomics, general](#) (192 articles)
[cDNA](#) (62 articles)
[cDNA Libraries](#) (26 articles)
[DNA Sequencing](#) (96 articles)
[Expression Libraries](#) (46 articles)
[Gene Prediction](#) (12 articles)
[Libraries](#) (147 articles)
[Libraries, general](#) (113 articles)
[Molecular Biology, general](#) (1293 articles)
[mRNA](#) (132 articles)
[Polymerase Chain Reaction \(PCR\)](#) (139 articles)
[Polymerase Chain Reaction \(PCR\), general](#) (184 articles)
[RNA](#) (317 articles)
[RNA, general](#) (269 articles)