Supporting Information

Kosugi et al. 10.1073/pnas.0900604106

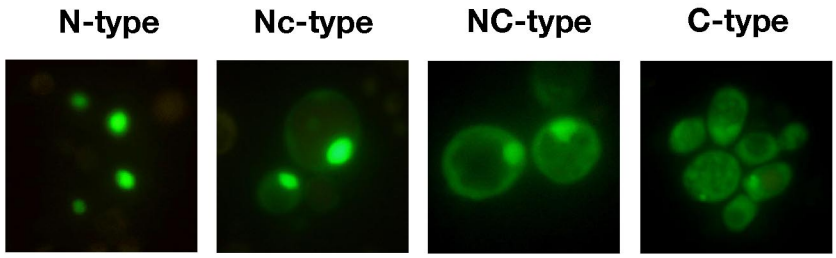
SI Text

Plasmid Constructions. To create the NLS profiles, double-stranded oligonucleotides encoding an NLS were inserted into the XbaI and BamHI sites of pTUE-GFP, a universal GFP reporter plasmid (1). For the localization analysis of yeast proteins fused to GFP, PCR-amplified yeast ORFs were inserted into the XbaI and BamHI sites or BamHI site of pGAD-GFP, a yeast-specific GFP expression plasmid that was generated by the replacement of the HindIII fragment of pGAD424 (Clontech) with the fragment containing the GFP-MCS region from pCMV-GFP2 (2). To alleviate the toxicity caused by the overexpression of several proteins, the ORFs were cloned into the low-copy plasmid pAUA-GFP or the inducible expression plasmid pYES-GFP. The former is derived from pAUR112 (Takara) and contains the GFP-MCS region under the control of the *adh1* promoter, whereas pYES-GFP contains the GFP-MCS fragment cloned downstream of the GAL1 promoter of pYES-Trp-2 (Invitrogen).

Yeast Manipulation and Measurement of Nuclear Import Activities. Yeast manipulations, including culture and transformation, were performed as described previously (2), and the strain SFY526 (Clontech) or YNN141 (*MAT*a *his3–532 trp1–289 ura3–1 ura3–2 ade2 leu2*::*HIS3*), a derivative of YNN140 (National Institute of Technology and Evaluation, Biological Resources Center), was used in most experiments. The *cdc28-as1* strain was produced by introducing a point mutation into YNN141 to replace Cdc28 Phe-88 with Gly (3). To produce the *msn5*Δ strain, an *msn5* deletion mutation was created by the transformation of YNN141 with a PCR fragment containing the *ADE2* marker and the 5′- and 3′-terminal sequences of the *MSN5* gene. A strain with a LMB-sensitive allele of Crm1 (T539C) was produced as described previously (4). Yeast was transformed with a GFP reporter plasmid or cotransformed with the pTUE-GFP3 and pGBK-tTA plasmids, and cultured at 30 °C with synthetic defined (SD) medium lacking the appropriate amino acids. GUS-GFP-NLS or GFP fusion proteins expressed in the transformed cells were observed with an epifluorescence microscope, model BX51 (Olympus), with an excitation filter specific for 460–490 nm. To determine the activity levels of various NLSs, the relative levels of the nuclear import activities were ranked as 1 of the 10 levels based on the localization phenotype of the GFP reporter, as described (2).

1. Kosugi S, et al. (2009) Six classes of nuclear localization signals specific to different binding grooves of importin α. *J Biol Chem* 284:478–485.
2. Kosugi S, et al. (2008) Design of peptide inhibitors for the importin α/β nuclear import pathway by activity-based profiling. *Chem Biol* 15:940–949.
3. Bishop AC, et al. (2000) A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature* 407:395–401.
4. Kosugi S, Hasebe M, Tomita M, Yanagawa H (2008) Nuclear export signal consensus sequences defined using a localization-based yeast selection system. *Traffic* 9:2053–2062.
5. Ubersax JA, et al. (2003) Targets of the cyclin-dependent kinase Cdk1. *Nature* 425:859–864.
6. Archambault V, et al. (2004) Targeted proteomic study of the cyclin-Cdk module. *Mol Cell* 14:699–711.

A



B

Scores	10	9	8	7	6	5	4	3	2	1
Ratio of four cell types	N	N>>Nc	N>Nc	Nc	Nc>NC	Nc<NC	NC	NC>C	NC<C	C

Fig. S1. Score representation of relative levels of NLS Activity. (*A*) Four different phenotypes for GFP localization. The extents of nuclear localization of the GUS-GFP reporter fusion protein in budding yeast were categorized into 4 types: (*i*) N, exclusively nuclear; (*ii*) Nc, partially nuclear; (*iii*) NC, localized into both the nucleus and the cytoplasm in a similar extent; and (*iv*) C, cytoplasmic. (*B*) Scores for relative NLS activity. Even a single species of NLS often conferred different GFP localization phenotypes in transformed cells. Thus, NLS activity was ranked in 1 of the 10 levels based on both the GFP localization phenotypes indicated in *A* and the proportion of cells with the specific GFP phenotypes among whole GFP-positive cells. The scoring was standardized as follows: score 10 (>99% of cells with phenotype N), 9 (99–90% of N), 8 (89–70% of N), 7 (69–40% of N, other Nc), 6 (Nc > *N* > NC), 5 (*N* < Nc < NC), 4 (>70% of NC, Nc > C), 3 (NC > C), 2 (NC < C), and 1 (>95% of C). In some cases, a relative difference in the intensity of GFP fluorescence between the nucleus and the cytoplasm was used to determine the final score. Several scores <1 and >10 were estimated based on the activities determined by using a different template with a contrasting level of basal activity. This figure is from ref. 2.

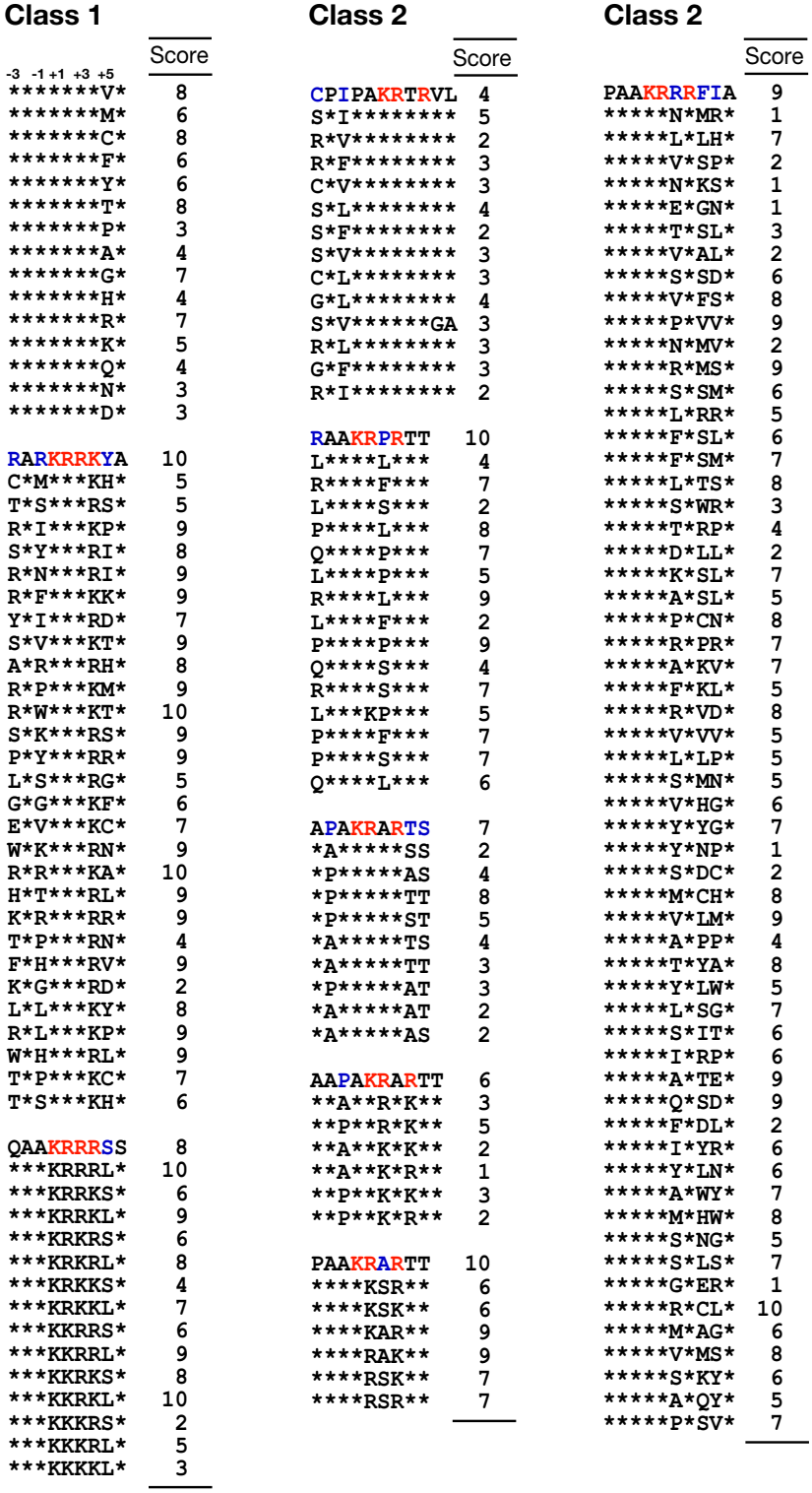


Fig. S2. Class 1 and class 2 NLSs simultaneously mutated at multiple sites. Two to 5 residues within the indicated classes of NLS were simultaneously replaced with other residues at the indicated positions, most of which were restricted to the flanking regions. The activity scores of the NLS mutants were measured in yeast. The core basic residues of the template sequences are marked in red. Asterisks represent unchanged residues.

A

class 2 NLS	NLS score
ssLaaKRLRStgs	2.5
ssLaaKRLRTtgs	3.5
ssLaaKRPRStgs	3.5
ssLaaKRPRTtgs	5.0
ssRaaKRLRStgs	7.0
ssRaaKRLRTtgs	8.5
ssRaaKRPRStgs	8.0
ssRaaKRPRTtgs	9.5

B

Standard NLS (score: 4.0)

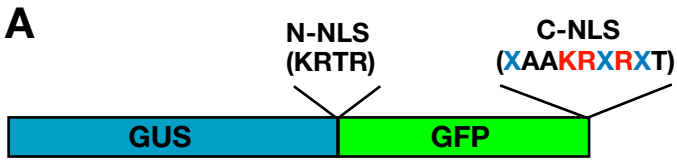
	−5	−3		+1	+3	+5							+11
	A	A	A	A	A	K	R	A	R	T	T	G	S	R	S	L
S – St:	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Sample NLS

	A	A	R	A	A	K	R	P	R	S	T	G	S	R	S	L
S – St:	0	0	4	0	0	0	0	1	0	-1.5	0	0	0	0	0	0

Total score = $\sum$ (S – St) + standard NLS score = 3.5 + 4 = **7.5**

Fig. S3. Calculation of NLS score based on the additivity principle. (*A*) Each residue within an NLS contributes additively to its entire activity. Class 2 NLS variants, whose residues were simultaneously replaced at 3 positions, were assayed in yeast, and the determined scores are indicated in the right column. Activating and repressing residues are marked in red and blue, respectively. Note that the effects of the replaced residues on the NLS scores were roughly independent and additive. (*B*) NLS scores can be calculated by the addition of the scores representing the contribution of each residue. The sequence indicated at the top is a class 2 standard NLS, which is the sequence used for the generation of the class 2 NLS profile used as the template (Fig. 1*A*). Because the total score (Ts) of this standard NLS is 4.0, the sum of the scores (S − St) obtained by subtracting the standard score (St) from the score (S) for each residue of the standard NLS is 0. A sample NLS indicated below differs in 3 residues, at positions −3, +3, and +5, from the standard NLS. Using the values for the class 2 profile in Fig. 1, we can obtain the S − St scores by subtracting St from S for each residue of the sample NLS. Finally, by adding St to the sum (3.5) of the S − St for each residue, we obtain the total score (Ts = 7.5) of the sample NLS, which is nearly identical to the measured score (8.0).



B

C-NLS	Actitvity score	
N-NLS	–	+
XAAKRXRXT		
L****S*S*	2	2.5
L****S*W*	1.5	2.5
L****S*T*	2	3.5
L****F*W*	2	-
L****F*T*	2	3.5
L****L*S*	2.5	3.5
L****L*W*	1.5	3
L****L*T*	3.5	5
L****P*S*	3.5	4.5
L****P*W*	5	6.5
L****P*T*	5	6
Q****S*S*	2	3.5
Q****S*W*	2	3.5
Q****S*T*	3.5	5.5
Q****F*S*	4.5	6.5
Q****F*W*	2.5	-
Q****F*T*	4	6
Q****L*S*	3.5	5.5
Q****L*W*	3	5
Q****L*T*	5.5	7
Q****P*S*	4.5	5.5
Q****P*W*	6.5	-
Q****P*T*	7	8
R****S*S*	6	6.5
R****S*W*	5.5	6.5
R****S*T*	7	8
R****F*S*	7	8
R****F*W*	5	6.5
R****F*T*	6.5	7.5
R****L*S*	7	-
R****L*W*	6	7.5
R****L*T*	8.5	9
R****P*S*	8	8
R****P*W*	9	9
R****P*T*	9.5	10
P****S*S*	5	6.5
P****S*W*	4.5	6
P****S*T*	7	8
P****F*S*	7	8
P****F*W*	5.5	7
P****F*T*	7	8
P****L*S*	7	8
P****L*T*	8	8.5
P****P*S*	7.5	-
P****P*T*	8.5	9

C

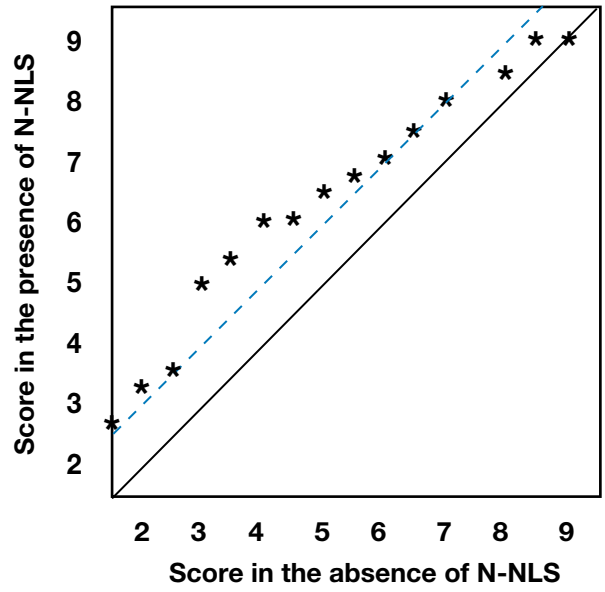


Fig. S4. Fine adjustment for a linear correlation between activity score levels. (*A*) Diagram of GUS-GFP reporter protein used to examine the score linearity. NLS mutants containing mixed residues at 3 positions (C-NLS) were fused at the C terminus of 2 different GUS-GFP reporters: one contained a weak NLS (N-NLS; PAAKRTRAP) at the N terminus of GFP and the other did not. (*B*) Scores for the C-NLS activities in the presence or absence of N-NLS. The activity scores for the indicated C-NLSs on the GUS-GFP reporter with (+) or without (−) N-NLS were measured in yeast. Asterisks represent unchanged residues. The data include those indicated in Fig. S3. (*C*) Plots of NLS scores determined in 2 different reporter contexts. The scores determined with the reporter lacking the N-NLS were plotted on the *x* axis against the scores determined with the reporter containing the N-NLS on the *y* axis. The average scores on the *y* axis against the scores on the *x* axis are represented with asterisks, and the linearity expected between these scores is indicated with a broken blue line.

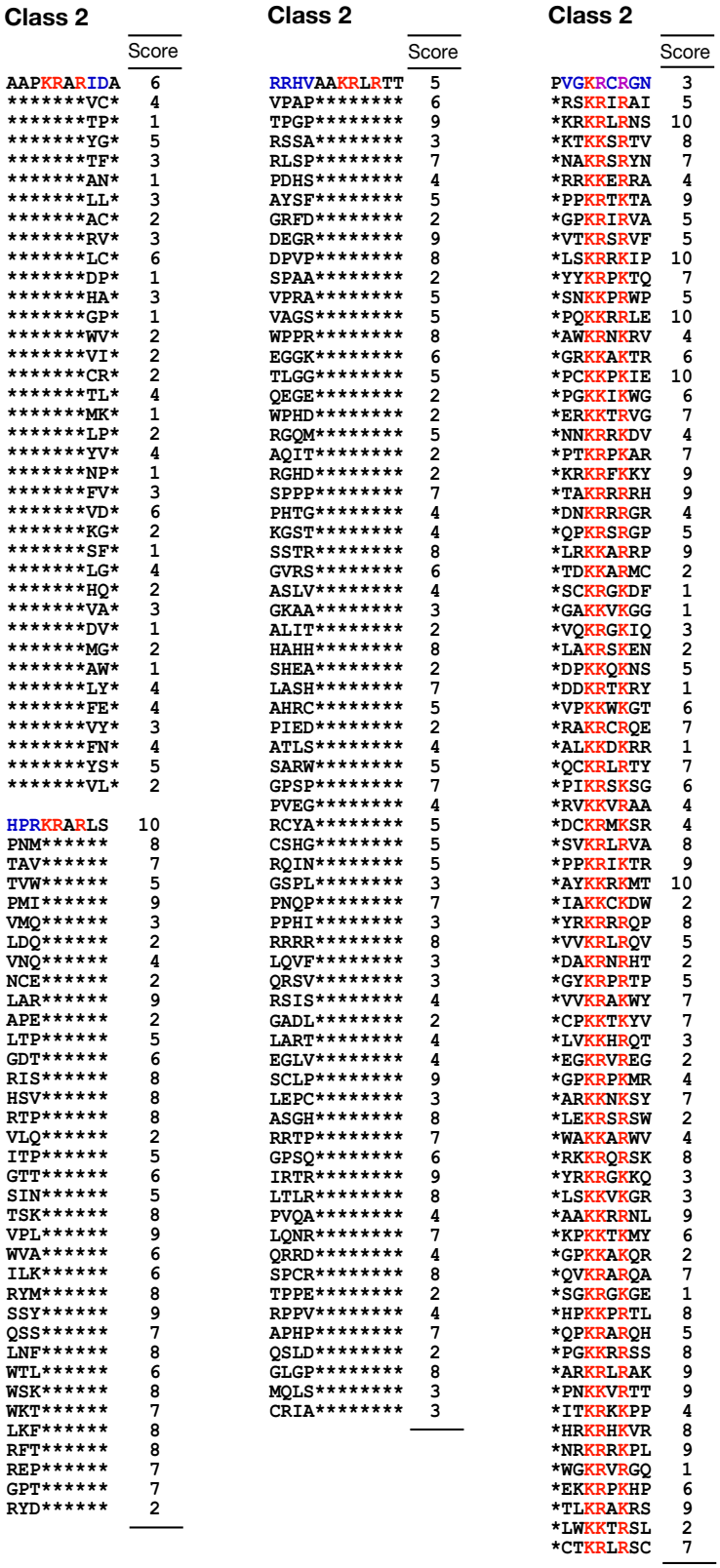


Fig. S5. Class 2 NLSs simultaneously mutated at multiple sites. The analysis was conducted as in Fig. S2. Asterisks represent unchanged residues.

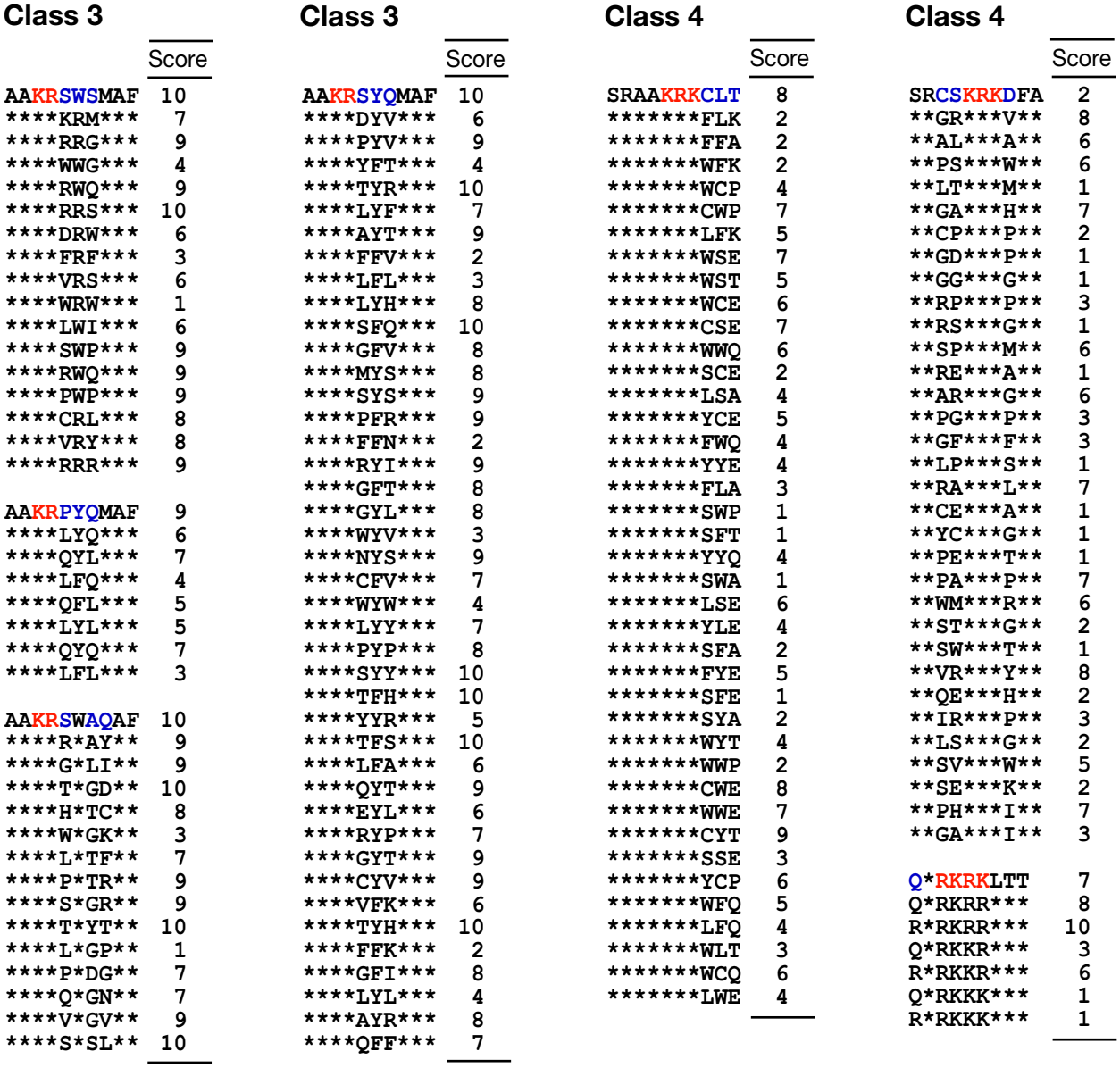


Fig. S6. Class 3 and class 4 NLSs simultaneously mutated at multiple sites. The analysis was conducted as in Fig. S2. Asterisks represent unchanged residues.

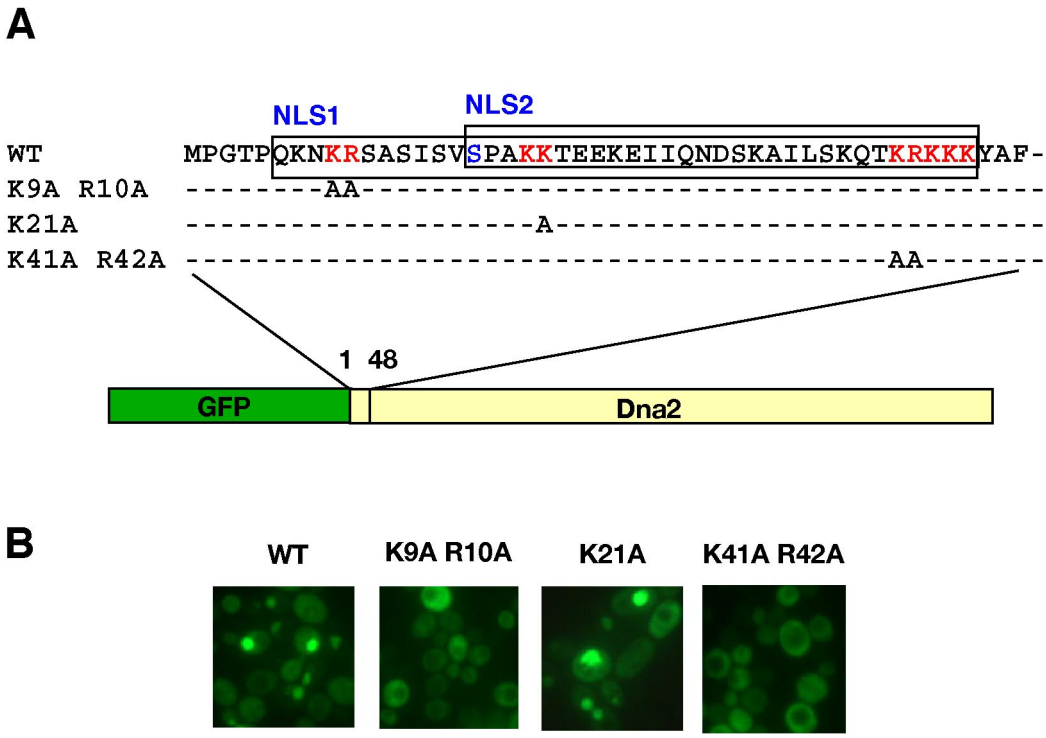


Fig. S7. A functional bipartite NLS of Dna2 contains a potential CDK site in its central linker region. (*A*) Schematic representation of the 2 predicted NLSs present in the N-terminal region of Dna2. Dna2 contains 2 putative bipartite NLSs at the N terminus (amino acids 1–48), where NLS1 and NLS2 contain a CDK1 consensus site in their linker and N-terminal flanking region, respectively. Although these NLSs have longer linker sequences than that of a standard linker with 10–12 aa, bipartite NLSs with long linkers have been found to be functional in structurally disordered regions, such as protein terminal regions. The putative core basic residues of NLS1 and NLS2 were separately replaced with alanine. The core basic stretches are highlighted in red, and a putative CDK1 phosphorylation site is marked in blue. (*B*) Subcellular localization of GFP-Dna2 fusion proteins mutated at NLS1 or NLS2. The subcellular localization of the alanine-replaced Dna2 mutants fused with GFP was observed in exponentially growing yeast cells.

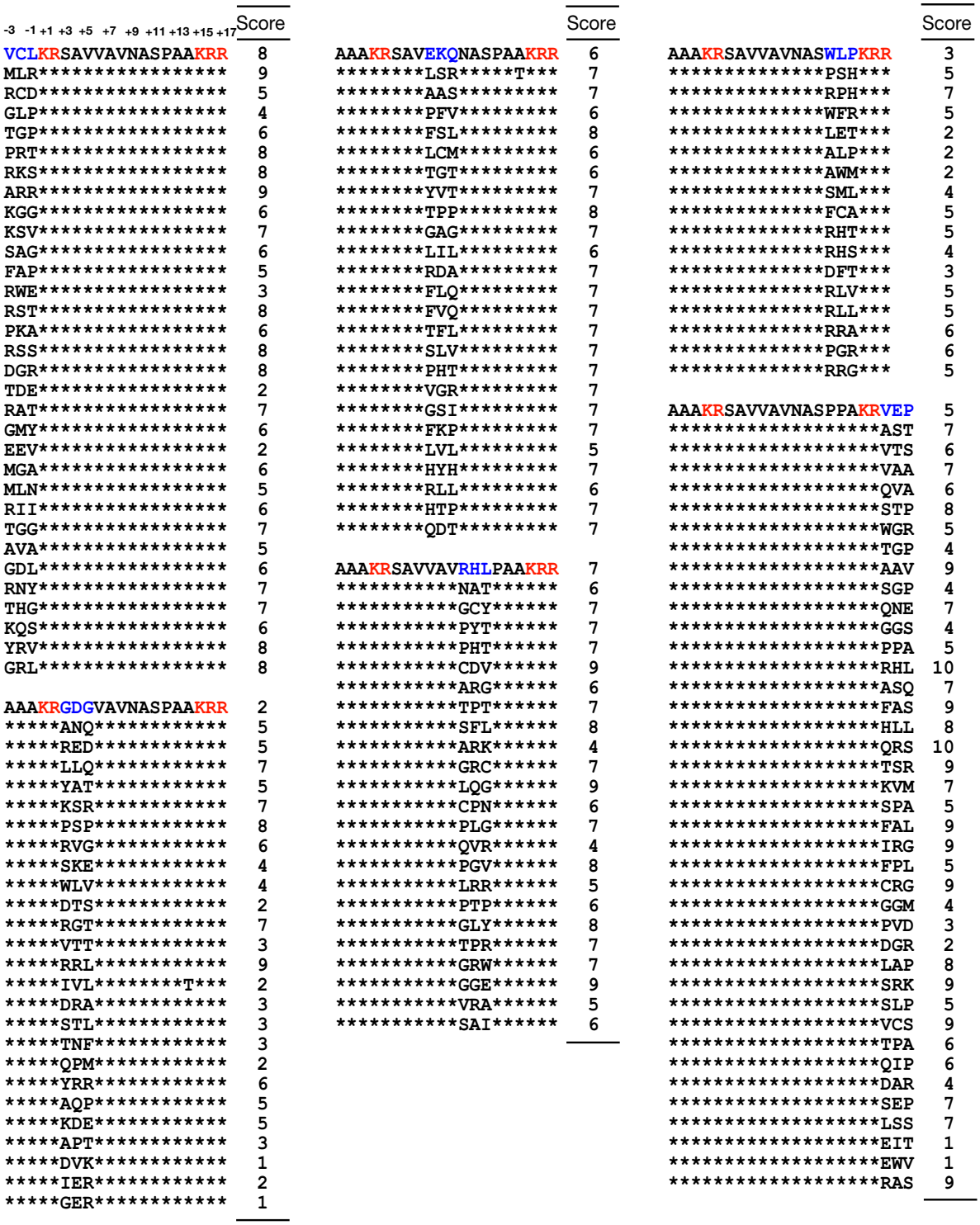


Fig. S8. Bipartite NLSs simultaneously mutated at multiple sites. Three consecutive residues within the indicated bipartite NLS templates were simultaneously replaced with other residues at the indicated flanking or linker positions. The activity scores of the NLS mutants were measured in yeast. Asterisks represent unchanged residues.

Table S1. Nuclear import activities of NLSs predicted in yeast nonnuclear proteins

Name	Annotation	Loc.	Predicted NLS	Calc.	Meas.
Fus1	Membrane protein localized to the shmoo tip	Vac	396 PLKKRKKRRQS	9	10
Fbp26	Fructose-2,6-bisphosphatase	Cyt	108 STRKRRKWLKDIC	10.5	10
Phm7	Protein of unknown function	Vac, CP	274 YVPHKKRPKHR	9	8
Tom71	Mitochondrial outer membrane protein	Mt	83 KKKNKRKRNNK	9	7
Gcn1	Positive regulator of the Gcn2p kinase	Cyt	166 QHRKRIRYCIF	10.5	7
Qri5	Mitochondrial protein of unknown function	Mt	90 KHKLRKRRKREK	12	9
Pho8	Repressible alkaline phosphatase	Vac	22 RPKKRRISKRSKI	11	9
Btn2	Protein that modulates arginine uptake	Cyt	34 YYPECKRRKAIK	8	9
Pex28	Peroxisomal integral membrane protein	–	560 PMFKRRRLI	12	10
Cct6	Subunit of the chaperonin Cct ring complex	Cyt	318 RRAKRRNMER	8	9
Gup2	Probable membrane protein	–	194 HSLKRKRLIAAF	8.5	6
Och1	Mannosyltransferase of cis-Golgi	Glg	313 YRHKRRHDET	10	10
Bni1	Formin, nucleates actin filaments formation	Cyt, Bud	1351 YPRPHKKLKQLHW	9.5	7
Pex31	Peroxisomal integral membrane protein	–	425 RVRKRKVSFS	9	7
End3	EH domain protein involved in endocytosis	PC	99 LIKERKKRKQI	9.5	9
Met7	Folylpolyglutamate synthetase	Cyt	416 RLSKRKKILLF	9.5	6
Ybt1	Member of the ABC transporter family	VM	334 HKVKRKRIFSLNLF	9	9
Pkc1	Serine/threonine kinase	Cyt, Bud	812 RAAKRRKVSLDNF	9.5	10
Gpb2	β subunit of the heterotrimeric G protein	Cyt, Mt	777 PPRKRKVDTL	13	8
Adp1	Permease of the ABC transporter family	ER	648 PQGKRRRIR	9	9
Gem1	Outer mitochondrial membrane GTPase	–	427 KPRKMRRRSGKLY	8	8
Jem1	DnaJ-like chaperone	ER	387 PDVKRAKLAAPFC	8.5	9
Isa1	Mitochondrial matrix protein	Mt	132 PPKKRKRKLR	13	9
Ena1	P-type ATPase sodium pump	CP	746 HRRKKFCTM	8.5	10
Mnn5	α-1,2-mannosyltransferase	PC	4 RLKKRKIL	9.5	10
Sln1	Histidine kinase osmosensor	CP	1193 LSKPIKRPKLK	8	7
YLR345W	Similar to 6-phosphofructo-2-kinase	Cyt	366 GPRKRTHDTAL	11	9
Sla1	Protein that binds to cytoskeletal proteins	Cyt, PC	263 GRKKKKAKLS	11.5	10
Rds1	Zinc cluster protein	–	19 KKIKRKCDKLR	9	8
Mnn10	Subunit of a mannosyltransferase complex	Glg, PC	104 VPPYSKRSRWSFW	9	3
			104 VPPY*A*KRSRWSFW	9	6
			104 VPPYSKR*A*RWSFW	11	7

The translated database of *S. cerevisiae* ORFs was searched for monopartite NLSs by using cNLS Mapper. The 30 NLSs presented were randomly selected from 80 predicted NLSs with calculated scores of $\leq$8 in apparently nonnuclear proteins, and they were assayed for NLS activity with the pTUE-GFP3 vector in yeast. The calculated (Calc.) and measured (Meas.) scores for the predicted NLSs are indicated in the last columns. In the Mnn10 NLS, the putative phosphorylation sites and their alanine-substituted sites are italicized. The subcellular localizations of the indicated proteins (Loc.), from the yeast localization database, are abbreviated as follows: cytoplasm (Cyt), cell periphery (CP), Golgi (Glg), mitochondria (Mt), punctate composite (PC), vacuole (Vac), vacuolar membrane (VM), and invisible or ambiguous localization (–).

Table S2. Nuclear import activities of bipartite NLSs in yeast nuclear proteins that contain a traditional consensus pattern but show a low score of calculation with cNLS Mapper

Name	Annotation	Putative NLS	Calc.	Meas
Smc3	Subunit of cohesin complex	993 NVNKRAFENFKKFNERRKD	< 1	2
Snt309	Protein involved in mRNA splicing	29 ANIKREPVHPEIRGILAKRKG	< 1	1
Tfa2	TFIID subunit	30 VGQKKTNDTVITIDGNTRKRT	1	1
Hda1	Subunit of histone deacetylase complex	4 VMVKKEVLENPDHDLKRKL	< 1	2
Mot1	Regulator interacting with TBP	192 NANKKSARMLAMARRKKKMS	1	2
Abf1	DNA binding transcription factor	417 QIARRITTYKARFVLKKKK	< 1	3
Prp43	RNA helicase involved in mRNA splicing	646 DNIRKALASGFFMQVAKKRS	2.5	1
Rap1	DNA binding transcription factor	801 EQIRKLVKKHGTGRMEMRKRF	< 1	1
Orc5	Subunit of origin recognition complex	349 IFSRKTRIIQGRAAYGRRKKK	2.5	2
Rox1	DNA binding transcription factor	74 EHERKYPEYKYKPVRKSKKKQ	< 1	1

The calculated (Calc.) and measured (Meas.) scores of the indicated bipartite NLSs are indicated in the last columns.

Table S3. List of predicted CDK1-regulated nucleocytoplasmic shuttling proteins

Name	Annotation	Localization	CDK1 target	Predicted NLS	Score
		Search for CDK1 site-containing monopartite NLSs			
Swi5*	Transcription factor that activates M/G1 transcription	N, NC, C	Yes	645 R**S**PR*K*R*G*RPR	6
Swi6*	Component of G1-specific transcription factors, MBF and SBF	N, NC, C	Yes	158 LG**S**PL*KK*L*K*ID	9
Mcm3*	Subunit of MCM2–7 prereplicative complex	Not tested	Yes	765 S**P**K*KR*Q*R*VR	6
Whi5*	Inhibitor of G1-specific transcription factor SBF	N, NC, C	Yes	4 RT**P***KR**S***RTSDE 47 **S**PVRLKNG 88 **S**PVN*KRR*VGIT	6.5 6
Acm1*	Inhibitor of anaphase promoting complex	N, NC, C	Yes	30 R**S**PS*KRR*SQID	4
Yen1†	Unknown function	N, NC, C	Yes	679 **S**PI*KK**S***RTT	4.5
Psy4†	Protein phosphatase involved in the repair of cross-linked DNA	N, NC, C	No	320 **T**PR*KR**K***PTDL 345 LT**T**P*KK*Y*K*HT	8
Msa1†	Activator of G1-specific transcription factors, MBF and SBF	N, SP, C	Yes	53 P**S**PN*KRR*LSID 84 **T**PTKKSSTN	8
Pds1†	Securin, anaphase inhibitor	N, NC, C	Yes	71 **S**PT*KR**L**H*TH	6
Rif1	Protein that controls telomere length	N	No	55 SP**T**P*KRRK*LA	13
Sfg1	Transcription factor that represses early G1	Ambiguous	Yes	28 **T**PS*KKC**K***YSSGF 45 **T**PS*KR**F***RLYQAK	6 6.5
Ela1	Elongin A, F-box protein	N, NP	No	269 **T**PV*KKRR*SESP	8
Nbp1	Spindle pole body component	SP	Yes	71 RK**T**P*KRR*LTKI	8.5
Eaf7	Subunit of the NuA4 histone acetyltransferase complex	N	No	397 **S**PKR*KRRK*AG	15
Sfi1	Protein required for spindle pole body duplication	NC, SP	Yes	799 DS**S**P*KRR*KDF	4
Vnx1	Monovalent cation/H + antiporter	VM	Yes	119 P**S****S**P*K*RM*HS*S	4.5
Nur1	Unknown function	PC	Yes	473 R**S**P*KKKK*NYH	5
Orc1	Subunit of the origin recognition complex	N	Yes	304 **S**PR*KR**G***RKI	5
Sum1	Transcriptional repressor for middle sporulation-specific gene	N	–	737 I**S**P*KKRR*TED	6
Bye1	Negative regulator of transcription elongation	N	–	176 E**S**PR*KR**K**R*SPD	9
Mcm21	Protein involved in minichromosome maintenance	NC	–	139 **S**PS*KR**K**R*KLK	8
Ies2	Subunit of the INO80 chromatin remodeling complex	N, NC	–	82 **S**PS*KKRH*LHTR	10
Heh2	Inner nuclear membrane protein	NP	–	123 **S**PK*KKRK*KRSS	9
Pct1	Phosphocholine cytidylyltransferase	N, No	–	59 **T**PR*KRRR*LTKE	15
		Search for CDK1 site-containing bipartite NLSs			
Cdh1*	Activator of anaphase promoting complex	N, NC, C	Yes	24 *KR*VSKRPISSSSSASLLS**S**PS*RRSR*	5.5
Dna2†	Nuclease/helicase, processing Okazaki fragment and DSBs	N, NC, C	Yes	9 *KR*SASISV**S**PAKKTEEKEIIQNDSK AILSKQT*KRKKK*	4.5
Srs2	DNA helicase involved in DNA double strand break repair	N, SP	Yes	879 **S**PE*KR*YAPETTSFHSPT*KKK*VY	6.5
Msh6	Protein required for mismatch repair	N	Yes	26 *KK*MKQSSLLSFFSKQVPSG**T** PS*KK*VQ	8
Kar3	Nuclear kinesin, microtubule motor protein	N, SP	Yes	6 **T**PT*K*GRSTQHLSTPSPKNDILAMN GH*KRR*NT	4.5
Far1	CDK inhibitor	N, NC	Yes	11 *KK*IHTPPSGDRDAER**S**PP*KK*FLR	5
Yox1	Homeodomain-containing transcriptional repressor	Ambiguous	Yes	349 *KR*PVSNPSG**S**PKR*KRK*FG	8.5
Kar4	Transcription factor that regulates the pheromone response pathway	NC	Yes	301 **S**P*KR*YKEEIANLGSNIPLKNEIELL *RPR*SP	5
Sir4	Protein involved in assembly of silent chromatin domains at telomeres	N	Yes	514 **T**PS*KR*PQLGEIPNPM*KK*H*K*PN	9.5
Spc110	Spindle pole body component	SP	Yes	36 **S**PT*K*VPNANNGDENEGPV*KK* *RQRR*S	8

The first 24 and the subsequent 10 proteins were predicted by the searches for CDK1 site-containing monopartite NLSs and CDK1 site-containing bipartite NLSs, respectively. To reduce the toxicity of some overexpressed proteins in yeast, the low-copy plasmid pAUA-GFP was used to express Whi5, Cdh1, Rif1, Sfg1, Sum1, Mcm21, Ies2, and Heh2, whereas the galactose-inducible plasmid pYES-GFP was used to express Pds1, Sfi1, and Srs2. The subcellular localizations of the indicated proteins (Localization) are abbreviated as follows: nucleus (N), cytoplasm (C), both nucleus and cytoplasm (NC), spindle pole (SP), punctate composite (PC), vacuolar membrane (VM), and nucleolus (No). Proteins included in the 235 reported CDK1 substrates (5, 6) are indicated as potential CDK1 targets, whereas proteins included in reported non-CDK1 substrates (496 proteins other than the top 200 substrates, from ref. 5) and proteins included in neither set are indicated as "no" and "–," respectively. In the next columns, potential CDK1 phosphorylation sites within and around the predicted NLSs are indicated by bold type, and the core basic residues of the predicted NLSs are indicated by italic type. The calculated scores (Score) of the predicted NLSs are indicated in the last columns.
*Previously reported proteins that exhibit nucleocytoplasmic shuttling in a cell cycle- and CDK1-dependent manner..
†Previously unidentified proteins that exhibit nucleocytoplasmic shuttling in a cell cycle- and CDK1-dependent manner.