Supplemental Methods

The Preparation of 454 Samples

Transposition

1. Transform strains of *S. pombe* with pHL2673 and purify single colonies.
2. Test the transposition efficiency and perm the new transformants.
3. Streak patches on EMM +B1 -uracil plates. Twelve patches/plate; include INfs (YHL1691) and PRfs (YHL1689) on each plate. Incubate at 32°C 2 days. Using more than 20 plates is recommended.
4. Replica print onto EMM-B1 -uracil plates. Incubate plates at 32°C four days.
5. Replica print onto EMM+5-FOA +B1 plates. Incubate at 32°C two days.
6. Replica print onto YES +5-FOA+G418 plates. Incubate at 32°C for 40 hours.
7. Harvest the patches and purify genomic DNA (protocol shown below).

MseI digestion

1. Set up six duplicate digestions with 1~2 μg (up to 10 μl) of genomic DNA and MseI (or other Restriction Enzyme) in 100 μl volumes. Add 30 units MseI (New England Biolabs) to each reaction to generate a final volume of 100 μl. Incubate at 37°C for 16 hours.
2. Combine the 6 samples and add 3 ml buffer PB (Qiagen PCR purification kit). Perform two sequential loadings of 0.6 ml of the combined sample onto each of three PCR purification columns. Use the protocol of the Qiagen PCR purification kit and elute each column with 50 μl buffer EB. Combine the product.

Linker ligation

Ligation buffer 5x.....4 μl
T4 DNA ligase (Invitrogen).....0.5 μl (0.50 U)
Purified MseI digest product.....12.5 μl
Annealed linker oligonucleotides (10 μM).....3.5 μl

Perform ten duplicate ligation reactions.
Incubate the samples at 25°C for one hour, then incubate at 65°C for 10

min. If using another ligase, the temperature and time should be optimized. If using a restriction enzyme other than MseI in the previous step, the ratio of linker to genomic DNA should be optimized.

SpeI digestion

Add 3 µl NEB buffer 2 (or other buffer that the present NEB catalog recommending for SpeI) to the 20 µl ligation reactions, add 5 µl 10x BSA, 10 units of SpeI (New England Biolabs), and distilled water to 50 µl. Incubate at 37°C for 2 hours, then add 10 more units of SpeI, and incubate at 37°C for 2-3 hours. Combine the samples.

PCR

Use the SpeI cut product above (no purification) as template in PCR reaction. Use Titanium Taq polymerase from Clontech.

Buffer 10x.....200 µl
dNTP 50x.....40 µl
LTR primer 10 µM.....40 µl
Linker primer 10 µM.....40 µl
Template.....320µl
Taq (Titanium).....40 µl
dd H_2O.....1320 µl
Total volume.....2000 µl

Combine all ingredients above to produce 2.0 ml of reaction mixture and distribute 20 µl of the mix into each of 96 wells of a microtiter PCR plate.

PCR program:

1. 94°C 2 min
2. 94°C 15 s
3. 65°C 30 s
4. 72°C 45 s
5. goto step 2 for a total of six cycles.
6. 94°C 15 s
7. 60°C 30 s
8. 72°C 45 s
9. goto step 6 for a total of 24 cycles.
10. 72°C 10 min
11. 4°C until sample is retrieved.

PCR product purification

1. Combine all PCR reactions.
2. Divide mixture into six equal amounts and concentrate each on a Qiagen PCR purification column. Elute each column with 70 µl buffer EB.
3. Add loading dye (bromphenyl blue only), and run sample on a 10 cm TBE gel made of 2% agarose. Run the gel at 70 volts until the dye moves seven cm. Cut out the DNA of size 130-500 bp, weigh the gel. Purify with Qiagen gel extraction kit, one column/400 mg gel, elute with buffer EB in a total volume of 500 µl.
4. Extract with phenol one time, extract with phenol:chloroform:IAA one time, and extract with chloroform one time.
5. Add 5 volumes of buffer PB (Qiagen PCR purification kit), load onto one Qia PCR purification column. Wash with 500 µl PB, wash with 750 µl buffer PE (Qiagen PCR purification kit) twice, wash with 750 µl 75% ethanol.
6. Elute with 50-100 µl TE.
7. Measure the concentration with the PiCo-Green kit (Invitrogen) and a fluorimeter.

Purification of genomic DNA from *S. pombe*

1. Resuspend cells in 40 ml water.
2. Pellet 200 OD units cells by spinning at 3000 rpm (Beckman JS4.2 rotor, RCF=1,855) for 5 min. Resuspend the pellet in 2.5 ml of SP1 containing 15 mg zymolyase 100T (Seikagaku Kogyo).
3. Incubate at 37°C for 2 hours or until cells are dark when viewed with a phase contrast microscope.
4. Pellet spheroblasted cells by spinning at 3000 rpm (Beckman JS4.2 rotor, RCF=1,855) for 5 min.
5. Resuspend cells in 7.5 ml 5xTE. Add 0.75 ml 10% SDS and incubate at room temperature for 60 min. Then incubate at 65°C for 5 min.
6. Transfer to 12 ml Falcon snap cap tube. Add 2.5 ml 5M KOAc and incubate at room temperature for 30min.
7. Centrifuge sample in a 50 ml conical tube at 5000 rpm (Sorval GSA rotor with adaptors) for 15 min to remove debris. Pipette supernatant into a new tube.
8. Add 10 ml ice-cold isopropanol to supernatant, incubate on ice for 5 min.

9. Centrifuge at 8000 rpm (Sorval GSA rotor) for 10 min to pellet the DNA.
10. Resuspend pellet in 3 ml 5xTE. Add RNase A to a final concentration of 100 µg/ml.
11. Incubate at 37°C for 1 hour.
12. Add 0.1 ml 10% SDS and Protease K to final concentration of 50 µg/ml. Incubate 1 hour at 50°C.
13. Extract with phenol for several times until the interface between water and organic solvent become clear. Extract with phenol:choloroform:IAA until interface is clear. Extract with chloroform one time.
14. EtOH precipitate DNA by adding 1/10 volume 5 M NaCl and 2.5 volumes 100% ethanol. Centrifuge at 8000 rpm (Sorval GSA rotor) for 10 min. Wash the pellet using 2 ml 70% ethanol. Air dry the pellet. Resuspend DNA in 100 µl 1xTE.
15. Run 10 µl of the DNA product on a 1% TBE gel with lambda-Hind III marker to determine concentration. Typically we get a total of 100 µg total yield.

SP1 solution (with NaOH adjust to pH 5.6, autoclaved)

sorbitol..... 1.2 M
citric acid.....50 mM
Na_2HPO_4.....50 mM
EDTA .....40 mM

Supplemental Figure Legends

Figure S1. The ligation mediated PCR of Tf1 insertions. (A) The steps for the preparation of samples for 454 sequencing are shown. (B) Shown in red is the unique sequence tag in the U5 region of the upstream LTR and the flanking sequence. The blue sequence is the end of LTR; in italics are the nearest MseI sites upstream or downstream of the mutation tag; underlined is the SpeI site. (C) The linker oligonucleotides are shown. "P" is a 5' phosphate group. "N" is the primary amino group that blocks extension. (D) One of the top strand primers used in PCR amplification. The black sequence is the tag B for 454 sequencing; the blue part is the barcode; the nucleotides in red are complementary to the unique sequence tag in the U5 region. The top strand primers used in the different experiments differed only in the barcode site. (E) The bottom strand primer used in PCR amplification is shown. In black is tag A for 454 sequencing; the nucleotides in red are complementary to the end of the linker.

Figure S2. The number of insertions at promoters compared to their transcription activity. Each of the genes in tandem orientation is plotted using its transcription activity as an X coordinate and the number of

upstream insertions as a Y coordinate. The transcription data was based on Affymetrix expression-chip hybridization signals (Wilhelm, B. T. et al, 2008, *Nature*. **453**: 1239-1243). The integration data was from the Hap_Mse_2 experiment. No correlation was observed between transcription and integration (the correlation coefficient R = 0.08.

Figure S3. The average number of insertions within repeat sequences of *S. pombe*. (A-C) The number of inserts within one kb intervals of the centromeres 1, 2 and 3 are shown. In blue are the number of inserts that mapped to a unique sequence in the genome. In green are the total number of insertions match more than one region in the genome. In magenta are the total number of insertions in the one kb interval that match more than one region divided by the number of regions present in the genome. This is an average integration level for repeated sequences. (D) The insertions in the subtelomeric repeats of the left end of chromosome 1 are shown. (E) The inserts in the three annotated rDNA repeats of the left end of chromosome 3 are shown. (F) The inserts in Tf2-3 of chromosome 1 are shown.

Supplemental Table S1

Tf1 integrations of the 4 experiments in all intergenic regions in S. pombe genome (see the separate Excel file)

Supplemental Table S2

Convergent intergenic regions that contain integrations in all 4 experiments

Intergenic regions			Integrations			
Upstream gene	Downstream gene	Containing non-coding transcription*	Hap_Mse_1	Hap_Mse_2	Dip_Mse	Dip_Hpy
SPAC977.17	SPAPJ695.01c	√	3	4	1	2
fta5 SPAC1F8.06	SPAC1F8.07c		3	6	5	9
SPAC1F8.08	SPAC11D3.01c	√	33	45	37	40
SPAC227.15	SPAC227.16c		3	3	2	2
SPAC1A6.11	plb1 SPAC1A6.04c		5	13	6	6
SPAC23H4.05c	SPAC23H4.04		5	11	8	7
SPAC1002.18	SPAC1002.21		9	11	7	9
cdc42 SPAC110.03	pss1 SPAC110.04c		3	4	4	1
alg3 SPAC7D4.06c	SPAC7D4.05		17	21	17	14
rps6 SPAPB1E7.12	mns1 SPAPB1E7.13c		20	26	17	21
SPAC683.03	SPAC683.02c		6	6	4	13
fio1 SPAC1F7.08	SPAC1F7.09c		3	7	1	1
SPAC23D3.12	SPAC23D3.13c	√	11	20	19	20
SPAC26F1.12c	SPAC26F1.11	√	38	59	46	59
SPAPJ691.03	Tf2-7 SPAC13D1.01c		2	5	2	6
SPAC4D7.11	fet5 SPAC4D7.12c		1	3	2	7
SPAC922.04	SPAC922.05c	√	28	47	29	25
SPAC922.06	SPAC922.07c		5	3	1	3
SPAC869.10c	SPAC869.09		5	10	8	16
SPBC800.11	SPBC800.12c		1	3	2	2
SPBC428.11	SPBC428.12c		3	7	5	6
rlp1 SPBC1685.11	SPBC1685.12c	√	11	15	17	12
SPBC354.10	SPBC354.11c	√	4	12	8	13
tea4 SPBC1706.01	wtf2 SPBC1706.02c	√	4	3	5	3
SPBC530.02	bag102 SPBC530.03c		2	6	10	9
sds23 SPBC646.13	orc5 SPBC646.14c		1	1	2	1
SPBC3D6.16	wsc1 SPBC3D6.14c	√	5	9	6	13
ran1 pat1 SPBC19C2.05	SPBC19C2.06c mug124		30	33	25	25
scr1 SPBC1D7.02c	SPBC1D7.01		2	1	1	1
matmc SPBC23G7.09	SPBC23G7.10c	√	22	3	3	5
ape1 SPBC1921.05	pvg3 SPBC1921.06c		4	5	2	2
SPBC12C2.05c	SPBC12C2.04		2	3	4	1

SPBC2A9.13	sec22 SPBC2A9.08c		2	1	2	1
hhf3 h4.3 SPBC1105.12	SPBC1105.13c	√	2	1	2	3
grx4 SPBC26H8.06	SPBC26H8.13c		13	8	5	10
ste11 SPBC32C12.02	ppk25 SPBC32C12.03c		1	1	3	3
SPBC56F2.07c	SPBC56F2.06 mug147		1	3	4	4
SPBC8E4.05c	SPBC8E4.04		2	3	1	2
rhp7 SPCC330.02	SPCC330.19c		2	3	2	1
dga1 SPCC1235.15	wtf3 SPCC548.02c		7	8	6	9
ssl1 SPCC1682.07	mcp2 SPCC1682.08c		3	6	3	1
sec231 SPCC31H12.07	ccr4 SPCC31H12.08c		2	3	4	1
SPCC1393.08	SPCC1393.09c		1	2	3	1
SPCC1795.04c	gms1 SPCC1795.03		5	7	2	2
wtf12 SPCC622.21	SPCC622.01c		1	5	2	3
SPCC622.07	hta1 SPCC622.08c		11	14	9	23
rpl39 SPCC663.04	cia1 SPCC663.05c	√	10	16	5	16
SPCC417.04	SPCC417.05c cfh2		25	27	20	14
SPCC191.08	gst1 SPCC191.09c		3	6	6	9
SPCC1450.15	SPCC1450.16c		5	8	8	6
SPCC576.02	tpx1 SPCC576.03c		8	11	7	8

***** Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C. J., Rogers, J. and Bahler, J. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature.* **453**: 1239-1243.

Supplemental Table S3

ORFs with sites that have inserts in all four experiments.

Gene	Number of independent positions in all four experiments	Comment
SPAC5H10.01	1	Site is in non-conserved N-terminus
SPAC24B11.14 SPAC806.10	12	ORF annotation is dubious
SPAC630.15 mug177	1	Sequence orphan
SPAC56F8.05c mug64	1	Site in intron of conserved gene
SPAC22A12.10	1	Site in C-terminus conserved in S. japonicus
SPAC23H4.05c	1	ORF annotation is dubious
SPAC1002.20	2	Sequence orphan
SPAC17H9.18c	3	ORF annotation is dubious
amt3 SPAC2E1P3.02c	1	Site in N-terminus is highly conserved
SPAC6G9.01c	5	Sites in 400 nt intron of conserved gene

SPAC8C9.20	3	ORF annotation is dubious
aah2 SPAC23D3.14c	2	Sites in 117 nt intron of conserved gene
SPAC9E9.01	1	ORF annotation is dubious
SPAC22F8.05	1	Site in conserved coding sequence
SPAC22E12.15	2	ORF annotation is dubious
wtf1 SPAC2E12.05	5	Psuedogene of repreat element
SPAC19D5.10c	7	ORF annotation is dubious
SPAC869.05c	3	Sites in non-conserved N-terminus
SPBC1683.12	3	Sites in conserved coding sequence
SPBC947.03c	4	Sites in conserved coding sequence
SPBC83.19c	3	Sequence orphan
SPBC18H10.21c SPBC9B6.01c	1	ORF annotation is dubious
rpa2 SPBP23A10.07	3	Sites in N-terminus that has been reannotated as 5' UTR
SPBC29A3.21	5	Sequence orphan
skb1 SPBC16H5.11c rmt5	1	Site is in non-conserved N-terminus
SPBC19G7.02	1	Site in conserved coding sequence
SPBC2A9.13	1	Sequence orphan
dbp3 SPBC17D1.06	2	Sites in non-conserved N-terminus
SPBC1105.13c	1	Sequence orphan
SPBC13G1.15c	1	Sequence orphan
SPBC215.10	1	Site in conserved coding sequence
top3 SPBC16G5.12c	2	Sites in conserved C-terminus
spn3 SPBC16A3.01	1	site in non-conserved N-terminus
SPBC1289.15 SPBC8E4.07c	1	site in non-conserved N-terminus
SPBPB2B2.11	1	Site in conserved coding sequence
SPCC24B10.18	1	Site in intron of conserved gene
SPCC663.14c	1	Site in non-conserved C-terminus
SPCC417.09c	1	Site in conserved coding sequence
SPCC285.04	1	Site in conserved coding sequence

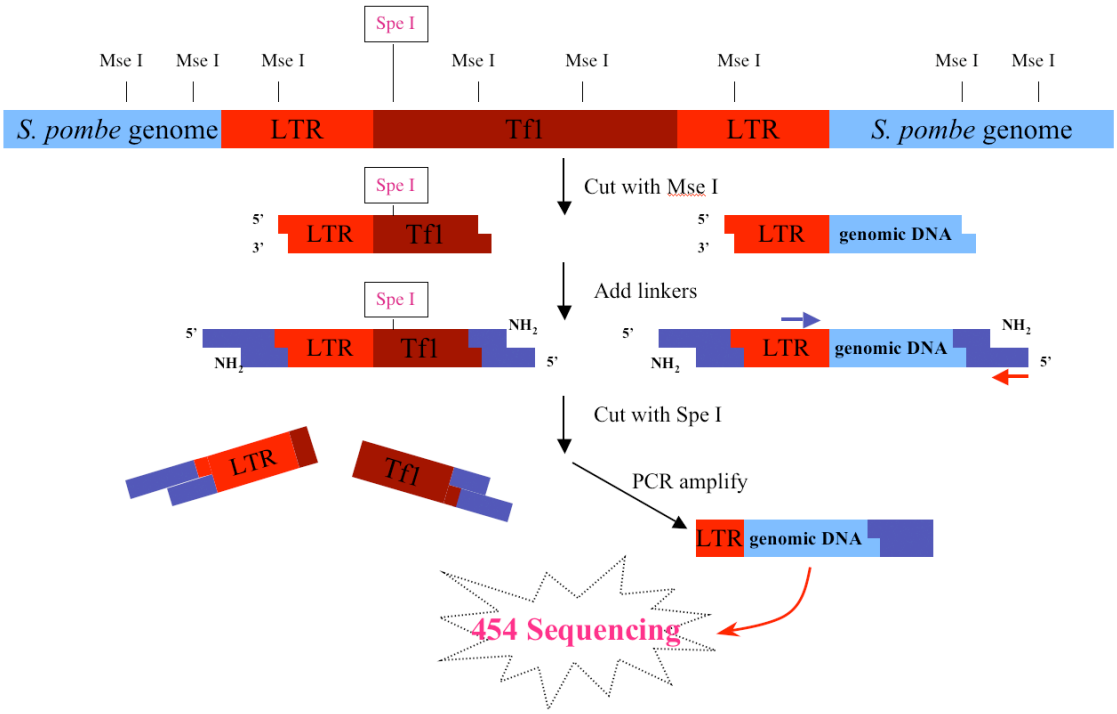
Supplemental Table S4

Strains

Strain	Genotype or description	Reference or source
YHL912	*h^- ura4-294 leu1-32*	{Levin, 1995 #1575}
YHL1605	*h^+ ura4-D18 ade6-M210 leu1-32::nmt1-lacZ*	{Singleton, 2002 #3618}
YHL5661	Stable diploid, *ura4-D18/ura4-D18, ade6-m210/ade6-m216, leu1-32::nmt1-lacZ/ leu1-32::nmt1-lacZ*	{Singleton, 2002 #3618}
YHL9426	YHL912/pHL2673	This study
YHL9530	YHL5661/pHL2673	This study
YHL9537	YHL1605/pHL2673	This study

Supplemental Figure S1

A



B

*TTAA*TACAATACCTATACTCAGTTGTTACTCATACAACCTGTGTATTGTAATAT
AATAGATCGCAAGGAAAACTCACCGCAGTTGATGCATAGGAATTTAGTTTAT
GGTAACTGCGTAGCTAA**CA**ATAACTGAACTCTTGTGATCTACAAATAATTCTT
TTCGAGAAAAAGGAATTATTGACTAGTATTTTCTCTCTGGAAA*TTAA*

C

HL1870: GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC

HL1871: P-TAGTCCCTTAAGCGGAG-N

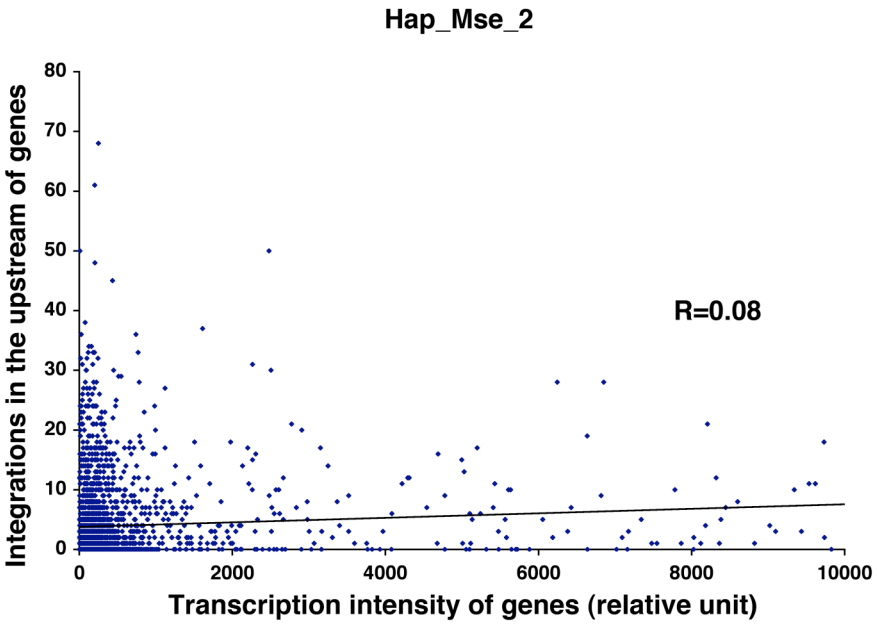
D

GCCTTGCCAGCCCGCTCAG**ACGT**GATGCATAGGAATTTAGTTTATGG

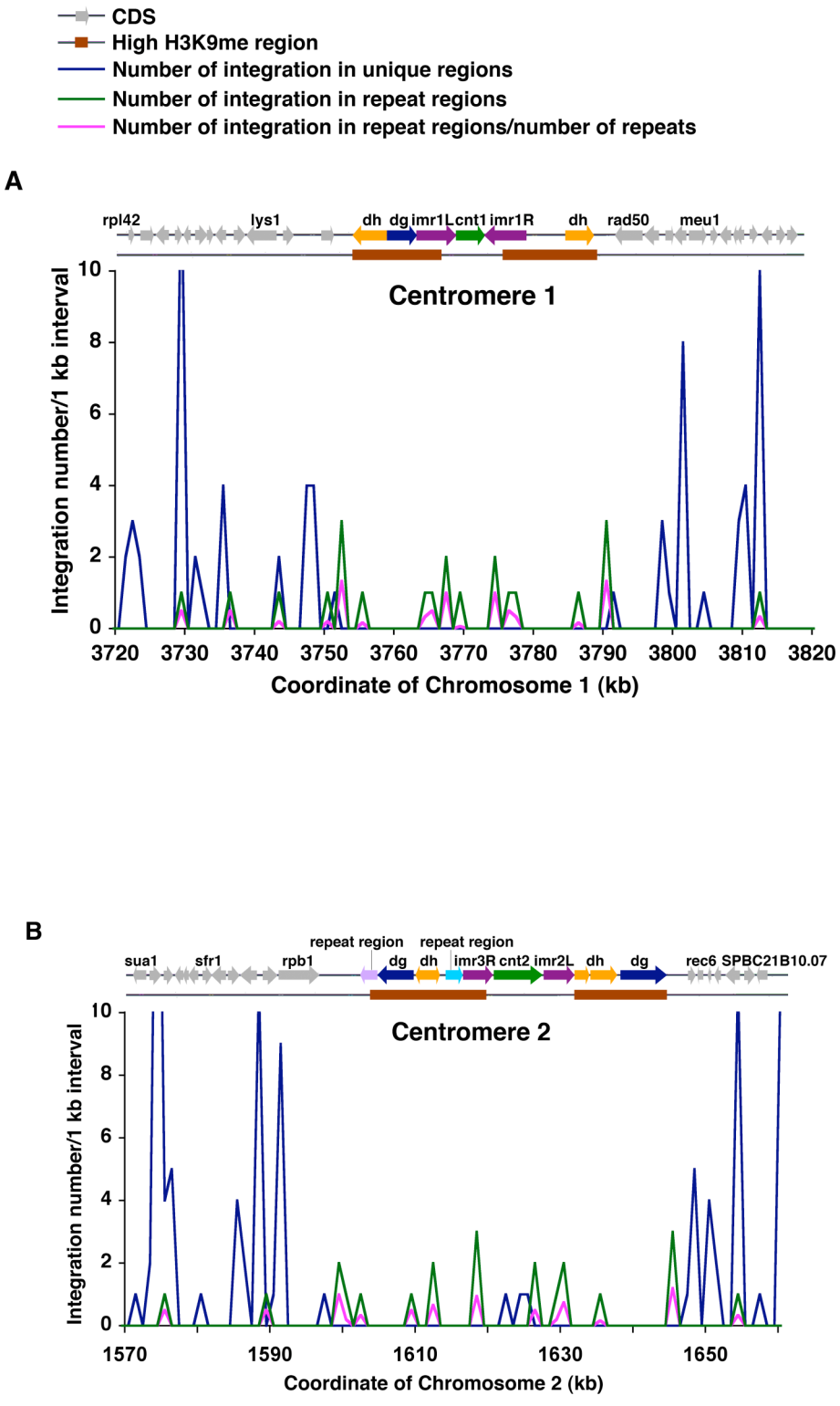
E

HL1954: GCCTCCCTCGCGCCATCAGGTAATACGACTCACTATAGGGC

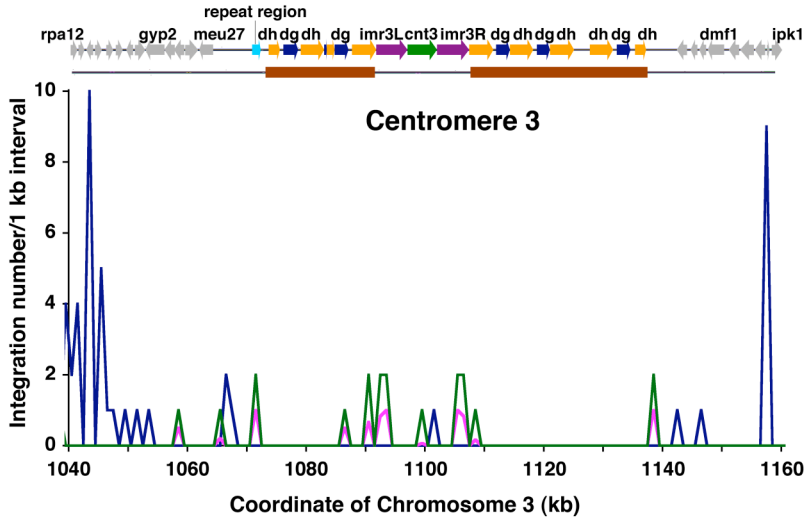
Supplemental Figure S2



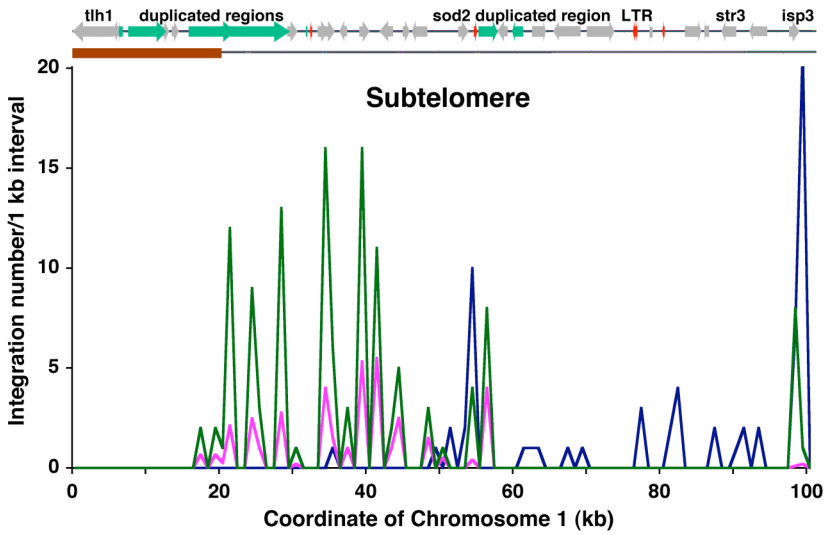
Supplemental Figure S3



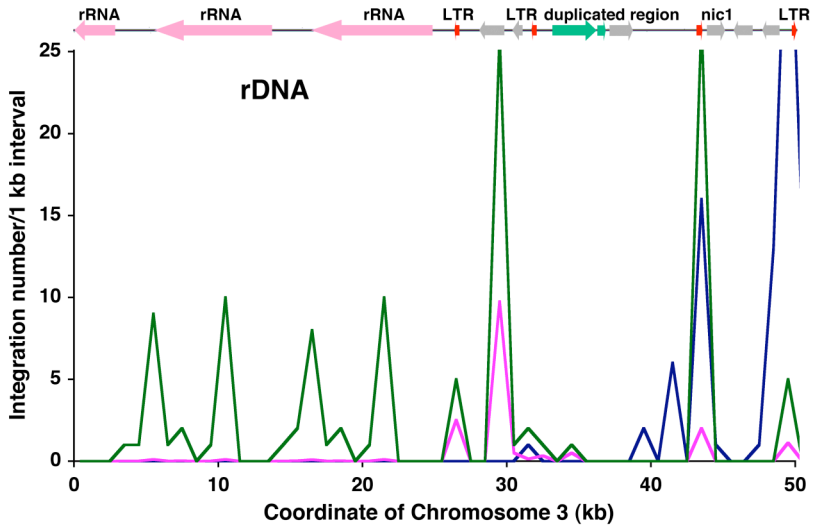
C


repeat region
rpa12
gyp2
meu27
dh dg dh
dg
imr3L
cnt3
imr3R
dg dh dg dh
dh dg dh
dmf1
ipk1
Centromere 3
Integration number/1 kb interval
0
2
4
6
8
10
1040
1060
1080
1100
1120
1140
1160
Coordinate of Chromosome 3 (kb)

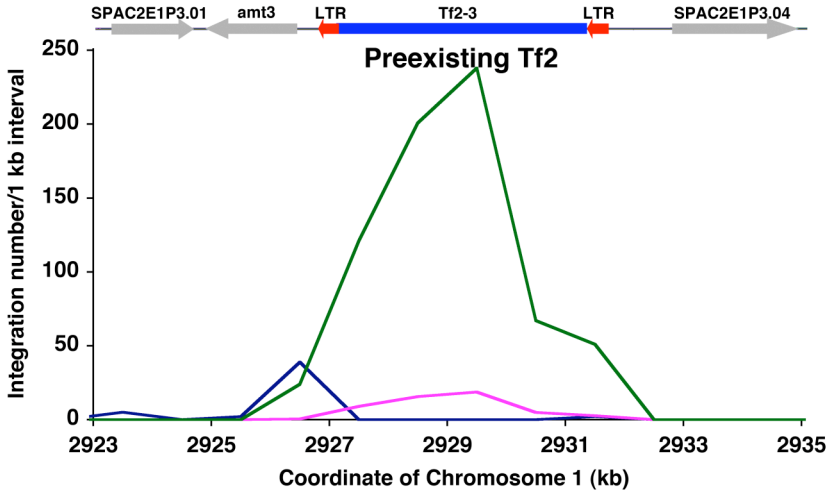
D


tlh1
duplicated regions
sod2 duplicated region
LTR
str3
isp3
Subtelomere
Integration number/1 kb interval
0
5
10
15
20
0
20
40
60
80
100
Coordinate of Chromosome 1 (kb)

E


rRNA
rRNA
rRNA
LTR
LTR
duplicated region
nic1
LTR
rDNA
Integration number/1 kb interval
0
5
10
15
20
25
0
10
20
30
40
50
Coordinate of Chromosome 3 (kb)

F


SPAC2E1P3.01
amt3
LTR
Tf2-3
LTR
SPAC2E1P3.04
Preexisting Tf2
Integration number/1 kb interval
0
50
100
150
200
250
2923
2925
2927
2929
2931
2933
2935
Coordinate of Chromosome 1 (kb)