Fig. S1

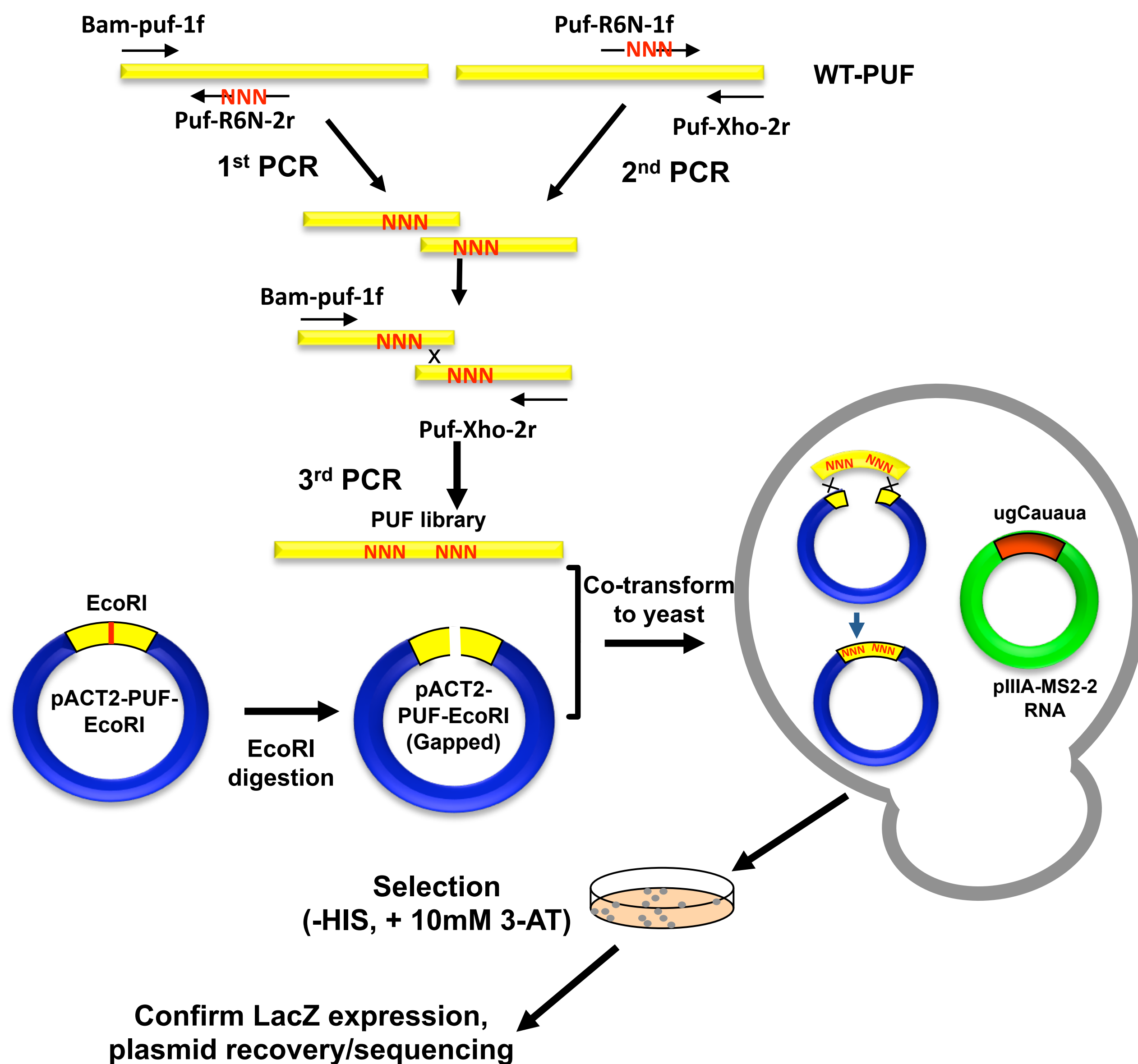

Fig. S1. Schematic diagram of the PUF library generation and screen. The PCR library of randomized PUF domain was generated by 3-step PCRs and the yeast library was generated by gap-repair. The potential C-binding PUF mutants was screened by yeast-3-hybrid. See the text for the details.

Fig. S2

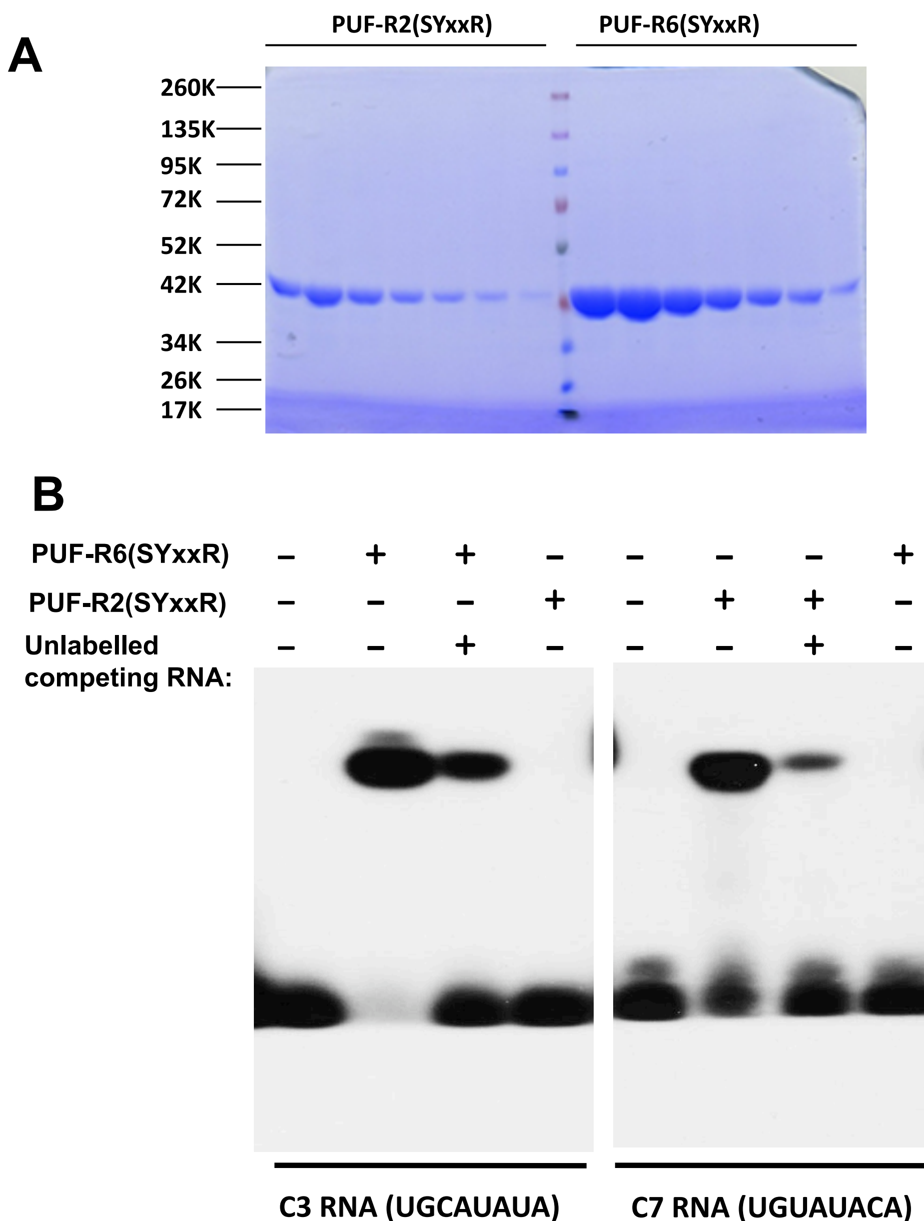

Fig. S2 PUF (SYxxR) proteins specifically bind their cognate RNAs with cytosine. (A) SDS gels showing the quality of purified PUF proteins, PUF-R6(SYxxR) and PUF-R2(SYxxR). The proteins were expressed in pTYB3 plasmid and purified with IMPACT kit (Intein Mediated Purification with an Affinity Chitin-binding Tag, NEB). **(B)** Analysis of RNA binding for PUF proteins with their target RNAs. PUF-R6(SYxxR) shifted the C3 RNA while PUF-R2(SYxxR) did not (lanes 2 and 4). Excess un-labeled C3 RNA competed to bind PUF-R6(SYxxR) and inhibited the shift of Biotin-labeled C3 RNA (lane3). Vise versa, PUF-R2(SYxxR) specifically bound and shifted the C7 RNA but not C3 RNA (lane 6 and 8). The binding of PUF-R2(SYxxR) with C7 was also inhibited by excess un-labeled C7 RNA (lane 7). Each binding reaction contained 20 fmol of Biotin-labeled C3 or C7 RNAs, 4 pmol of proteins, and 5pmol of unlabeled RNAs.

Fig. S3

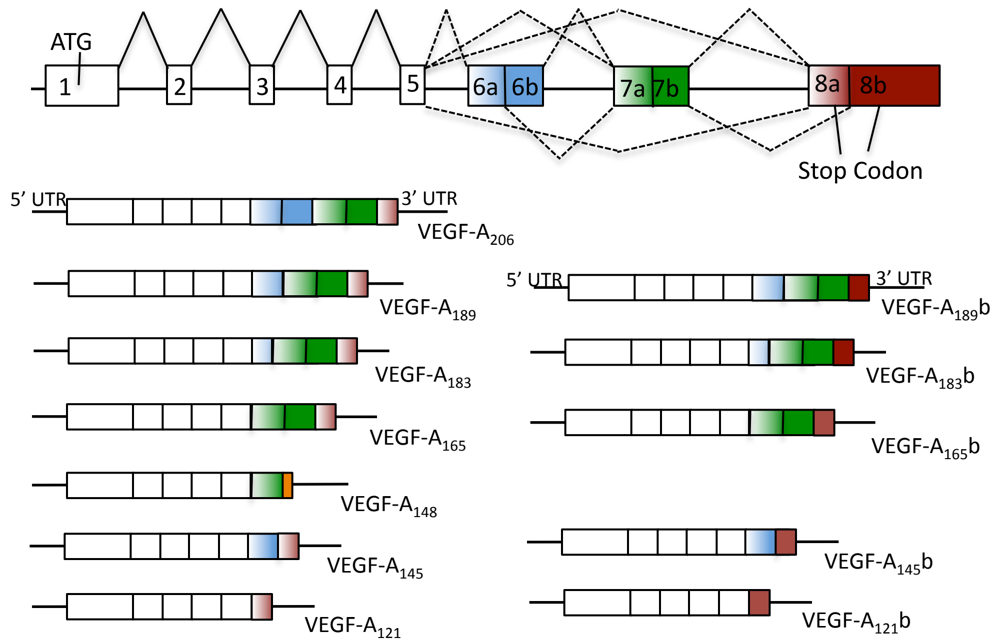

Fig. S3. Alternative splicing of VEGF-A gene VEGF-A gene
The exons and introns of VEGF-A are indicated with boxes and thin line (size not in scale). The exons 6, 7 and 8 of VEGF-A undergo extensive AS to produce multiple isoforms named as VEGF-A_{xxx}, with xxx denoting the amino acid number of mature protein. A new class of splicing isoforms, VEGF-A_{xxx}b, was identified to use a distal 3' ss of exon 8. The b isoforms (right panel) have anti-angiogenic activity, whereas the VEGF-A_{xxx} isoforms have pro-angiogenic activity.

Fig. S4

```
L. thermotolerans: XP_002554529(779-813)   -IVSNFWTIAQNRYGARAVRACLEAHDIITREQLLV
A. gossypii: NP_986362(735-769)            SIVSNFCTIVQNRYGARAVRACLEAHDIITQEQLLV
K. lactis: XP_454218.1(742-777)            NIVSNFWNIVQNRYGARAVRACLEAHDIITQEQLLL
C. glabrata: XP_449741.1 (705-740)         SILSNFWTIVQNRYGARAVRACLEAHDIVTMEQTLV
S. cerevisiae: NP_012624 (725-760)         SIIANFWVIVQNRYGARAVRACLEAHDIVTPEQSIV
S. cerevisiae: NP_015367(684-719)          SVLSHFWTIVQNRYGSRAVRACLEADSIITQCQLLT  Puf1
V. polyspora: XP_001644441(733-767)        NILSNFWVIVQSRYGARAVRACLEATDIVTKEQTLI
V. polyspora: XP_001645762(713-747)        NIVNNFWIIAQNRYGSRAVRACLEAHEIVTHEQTI-   Puf2
                                            :: :*  *.*.***:********* .*:*  * :
```

Fig. S4. Alignment of Puf2 homologs in different budding yeast strains.
The *S. cerevisiae* Puf2 has SRxxR sequence in the conserved base recognition positions in its repeat 4. This Puf2 is a yeast specific protein and the most closely related sequences are the putative Puf2 in *Vanderwaltozyma polyspora* and the Puf1/JSN1 proteins in different Saccharomycetes. The alignment of the repeat 4 of in Puf1/2 proteins showed that the putative C-binding code (SRxxR) was only found in the Puf2 of *V. polyspora* and *S. cerevisiae* while the other proteins have ARxxR in cognate positions (highlighted). Interestingly, both *V. polyspora* and *S. cerevisiae* have two paralogs (Puf1 and Puf2), whereas the other saccharomycetes have only one copy of the orthologous gene known as JSN1/Puf1. The C recognition code was only found in budding yeasts with two Puf2 paralogs, one of which has SRxxR sequence in the base recognition positions and the other has ARxxR sequence. In yeast with one copy (Puf 1), only ARxxR sequences were found in orthologous positions. It was known that thousands of duplicated gene pairs had been generated in several budding yeast genomes through whole genome duplication (WGD) followed by selective gene lose. Since both *V. polyspora* and *S. cerevisiae* are post-WGD strains, it is very likely that Puf1 and Puf2 were generated by WGD in *V. polyspora* and *S. cerevisiae*, and the additional mutations in repeat 4 of ancestor gene changed the codon from Ala into Ser

Table S1 Data collection and refinement statistics

	PUF-R6(SYxxR)
Data collection	
Space group	P212121
Cell dimensions	
a, *b*, *c* (Å)	35.8, 59.3, 344.7
α, β, γ (°)	90, 90, 90
Resolution (Å)	25-2.6 (2.64-2.60) *
R_{sym} or R_{merge}	12.7 (27.2)
$I / \sigma I$	9.0 (3.4)
Completeness (%)	89.6 (79.2)
Redundancy	3.9 (2.6)
Refinement	
Resolution (Å)	25-2.6
No. reflections	20661
R_{work} / R_{free}	21.0%/30.6%
No. atoms	
Protein	5484
Ligand/ion	422
Water	116
B-factors	
Protein	42.3
Ligand/ion	44.7
Water	38.5
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.092

*Values in parentheses are for highest-resolution shell.