SUPPORTING INFORMATION

Modulating Pathway Performance by Perturbing Local Genetic Context

Carmen Lopez[1], Yuxin Zhao[2, 3], Rick Masonbrink[4], and Zengyi Shao[1, 3, 4, 5 *]

[1] Interdepartmental Microbiology Program, Iowa State University

[2] The Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, College of Life Sciences, Nankai University, Tianjin, China

[3] Department of Chemical and Biological Engineering, Iowa State University

[4] Office of Biotechnology, Iowa State University

[5] NSF Engineering Research Center for Biorenewable Chemicals, Iowa State University

[6] Ames Laboratory

[*]Correspondence to: Dr. Zengyi Shao,
4140 Biorenewables Research Laboratory,
Iowa State University, Ames, IA 50011, USA
Phone: 515-294-1132
Email: zyshao@iastate.edu

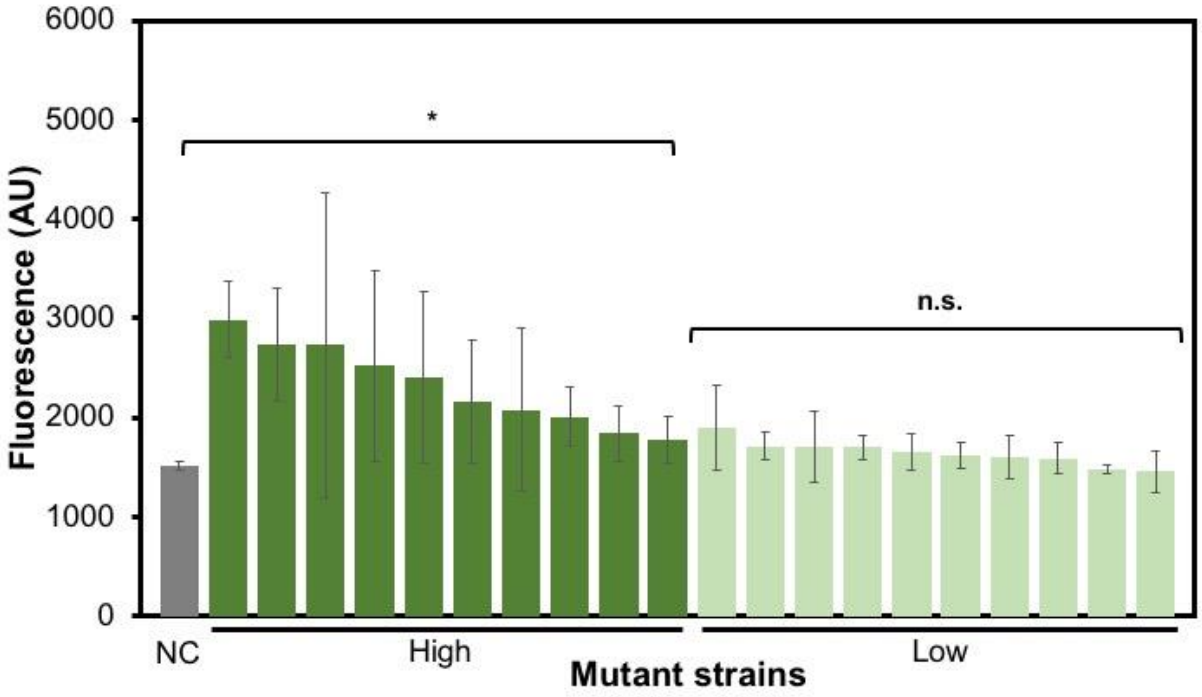

Figure S1. Fluorescence profile for the perturbed variants after plasmid re-transformation. Plasmid was isolated from the variants selected based on the fluorescence on the petri dishes and re-transformed to evaluate the consistency. Fluorescence was measured by flow cytometry after 44-h culture. Significant difference (*) was observed between the high-expression group and the negative control (NC) (two-sample t-test; $p<0.05$), but not between the low-expression group and NC. Error bars represent the standard deviation of three biological replicates. The mark 'n.s.' indicates 'not significant' as determined by the statistical test.

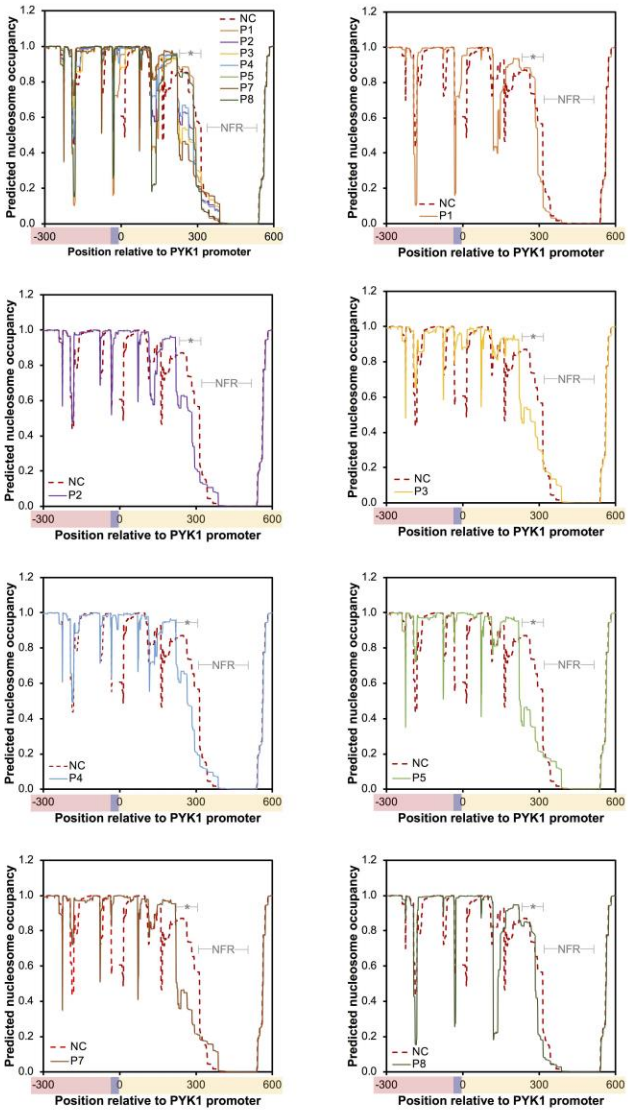

Figure S2. Predicted nucleosome occupancy for upstream region to *ghl-1*. Nucleosome occupancy as predicted by NuPop, a Hidden Markov Model[1,2]. First plot shows strains carrying perturbations compared to negative control; remaining plots show individual profiles for clarity. Genetic context is represented in the axis of abscissas: yellow shade indicates promoter region, blue rectangle indicates 30-bp perturbation sequence and red shade indicates any upstream region. NFR stands for nucleosome free region and asterisk (*) shows outstanding changes in occupancy. Error bars represent error propagation.

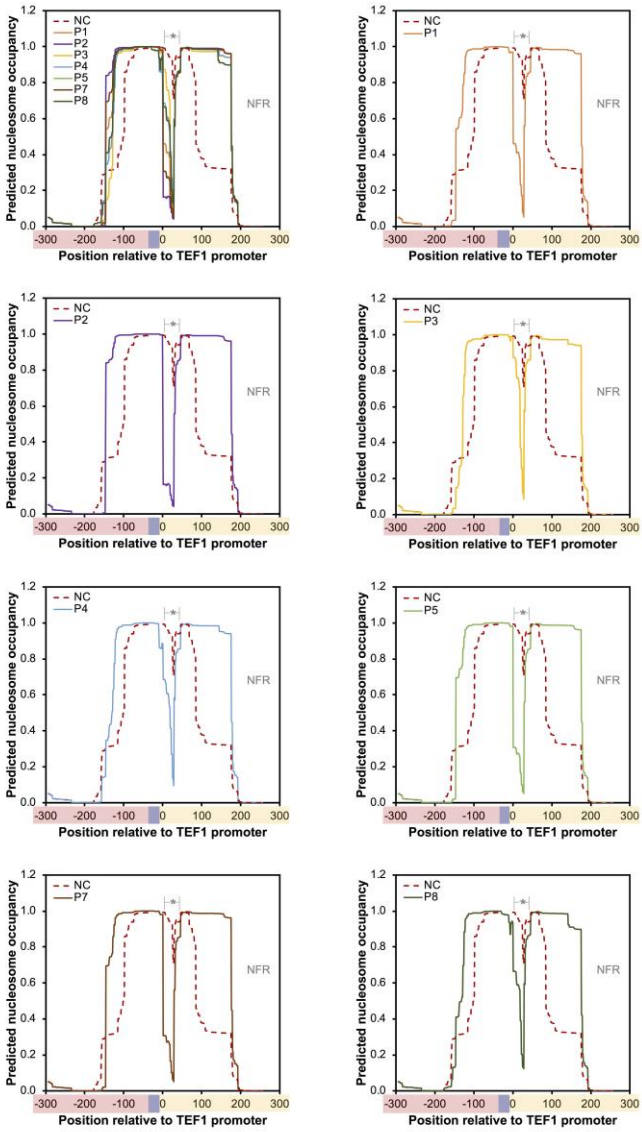

Figure S3. Predicted nucleosome occupancy for upstream region to *cdt-1*. Nucleosome occupancy as predicted by NuPop, a Hidden Markov Model[1,2]. First plot shows strains carrying perturbations compared to negative control; remaining plots show individual profiles for clarity. Genetic context is represented in the axis of abscissas: yellow shade indicates promoter region, blue rectangle indicates 30-bp perturbation sequence and red shade indicates any upstream region. NFR stands for nucleosome free region and asterisk (*) shows outstanding changes in occupancy. Error bars represent error propagation

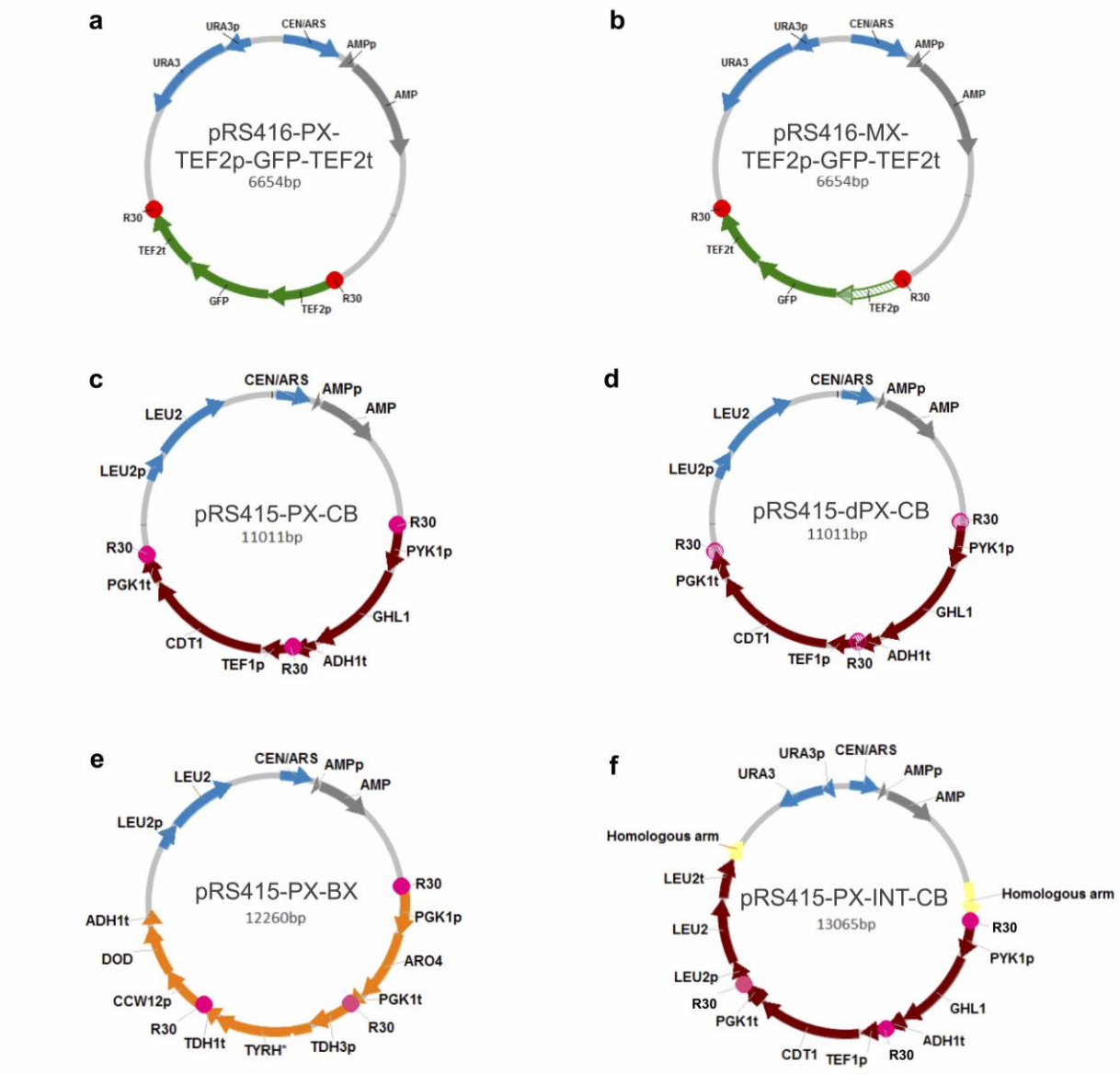

Figure S4. Maps for the main plasmids constructed in this work. a) Plasmid with perturbations (red circles) flanking *gfp* expression cassette. b) Plasmid with mutagenized TEF2 promoter (striped arrow). c) Plasmid bearing cellobiose pathway with intact perturbations (pink circles) or d) disrupted perturbations (striped circles). e) Plasmid bearing betaxanthin pathway with characterized perturbations (pink circles). f) Helper plasmid from which a perturbed cellobiose pathway is digested and integrated to genome.

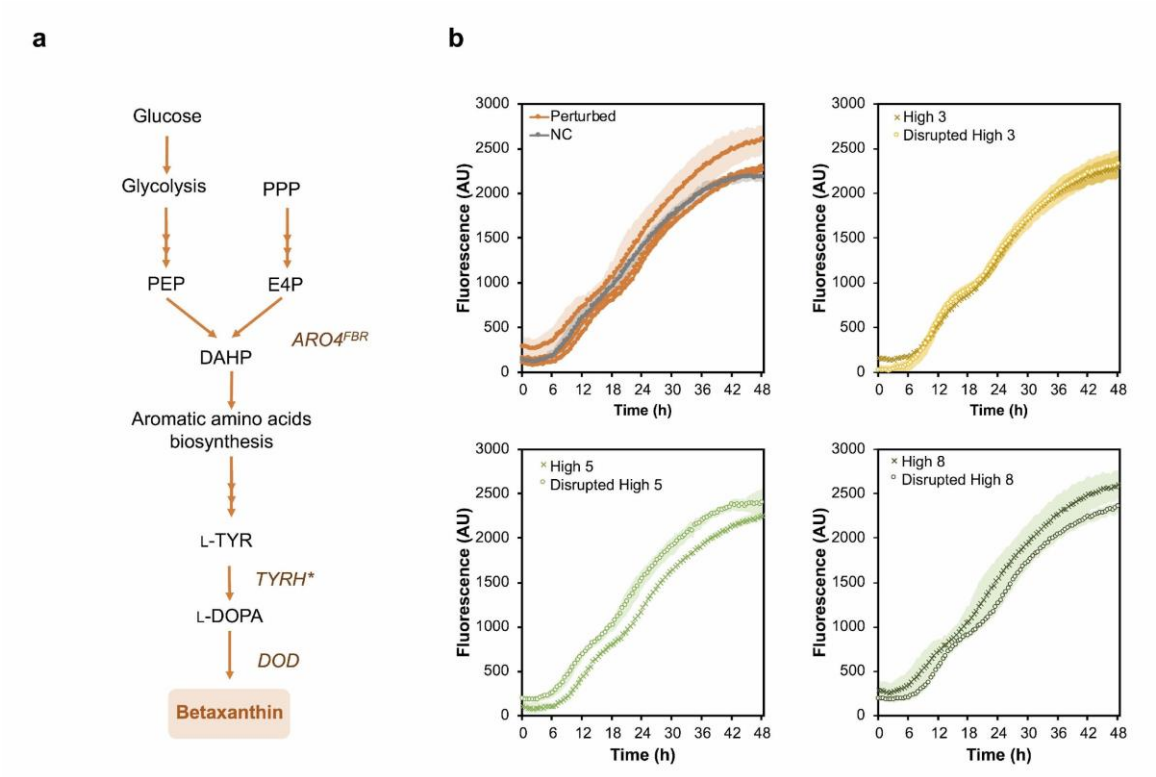

Figure S5. Assessment of perturbation orthogonality by a secondary exemplary pathway. a) The betaxanthin biosynthetic pathway includes a mutant 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase (*ARO4)*, a mutant tyrosine hydroxylase (*TYR**) and a L-3,4-dihydroxyphenylalanine (DOPA) dioxygenase. Metabolites: PEP, phosphoenolpyruvate; E4P, erythrose-4-phosphate; DAHP, 3-deoxy-D-arabinoheptulosonate; L-TYR, tyrosine; L-DOPA, L-3,4-dihydroxyphenylalanine. PPP stands for pentose phosphate pathway. Triple arrows indicate multistep reactions. b) Fluorescence profile for variants carrying the betaxanthin pathway perturbed by sequences P3, P5 and P8 yielded from cellobiose characterization. Orange curves represent strains with intact perturbations. They are compared against the negative control bearing the betaxanthin biosynthetic pathway without perturbation; no significant differences were found. Remaining plots show comparison of strains bearing either intact (cross) or its corresponding disrupted pair (open circle). Only P8 showed significantly higher total fluorescence than disrupted P8 (two sample t-test; $p<0.05$). The shaded area on each curve represents the standard deviation of three biological replicates.

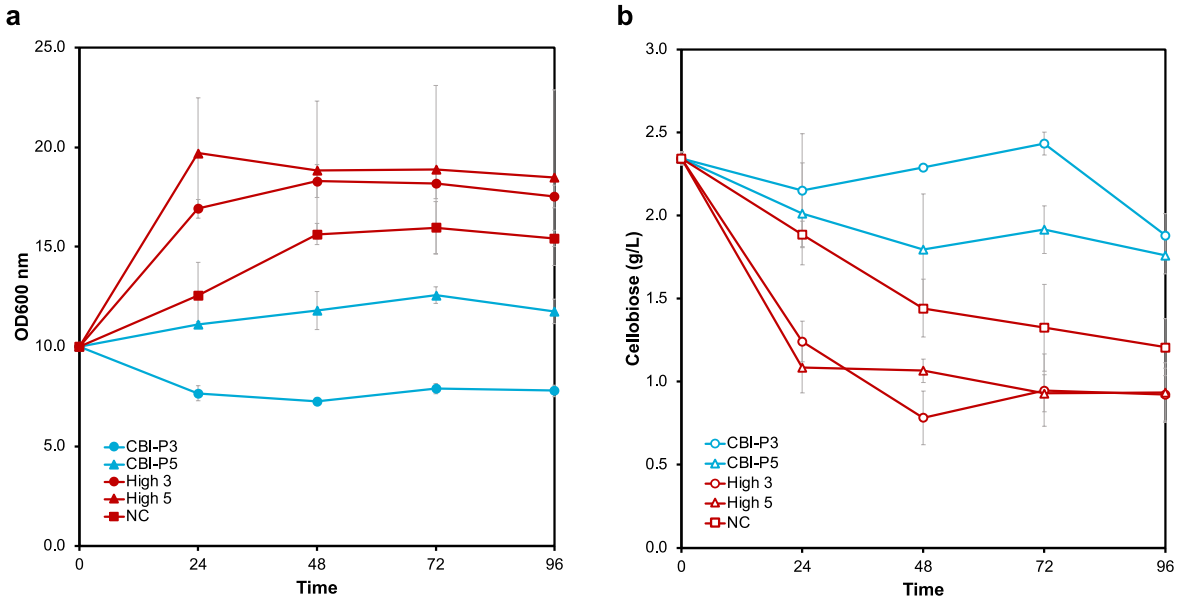

Figure S6. Assessment of the perturbed cellobiose utilization pathway in a genomic context. a) Growth and b) cellobiose-utilizing profiles for the variants during a 96-h fermentation. Red curves represent plasmid-contained pathways and blue curves indicate the pathway integrated to locus 8 reported previously[3]. Growth and consumed cellobiose were significantly higher in plasmid-based systems (two-sample t-test; $p<0.05$). NC represents the cellobiose utilization pathway without perturbation. Error bars represent the standard deviation of three biological replicates.

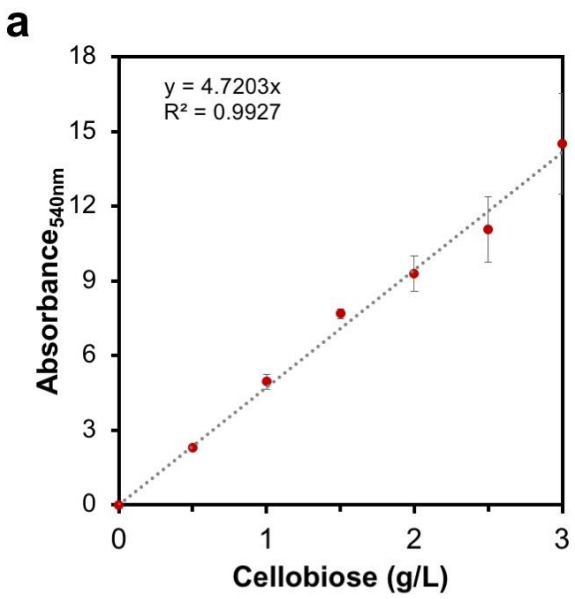

Figure S7. Standard curve for cellobiose estimation. a) Correlation between different concentrations of cellobiose (0 to 3 g/L) and absorbance at 540 nm.

Table S1. Perturbation sequences retrieved in this work. The bold and underlined codes highlight the conserved residues. Italicized letters represent mutated and omitted (*) residues, both of which intended to disrupt the exemplary perturbations.

Gene of interest	Variant	Upstream	Downstream
gfp	P1	AGGTAGATGCTCAGAG AACGTGACTCAGGT	AAAAAGGACGGTGGCT TGCACACTGAAGAT
	P5	GGTTTGATTATGTTTGG CGTCTTGGGGGGG	CAACATGCTGTGCGCCA CACTAAACTCTAA
	P9	TCAAGCATGTTGATAGT GACGGGGATCCGC	AACAATGAGAAGGACC GTCTATATAGCAAG
ghl-1	P1	TAGTGACGTAATTGGTA ATCATTGTAACAC	GAACAATTTGATACCGG TTATTTCCACTCC
	P2	GGAACA**G**TTG**G**GGGT**G** GCATTGGTGAGGTG	TCCTTGTTCCGGGTGTA GGTCTTTGTGCCT
	P3	ATCCCA**G**GAT**G**TCTG**G** GTCCGTCCTATTAC	AGTGCCGGGTTGGTGCT GTGGTTAGTGGGG
	P4	ACGGGGCGCGTGCGGA GACAAGGCGTGGAG	ATGGGCTAACGTGAGCT CCTTGCAATTTTG
	P5	CGGGGT**G**TTG**G**GCTG**G** TGCGTGTCCGGGCC	TGCTCTTGTTTCTACAA GGCAGTATGTTCC
	P6	ACGGGGCGCGTGCGGA GACAAGGCGTGGAG	ATGGGCTAACGTGAGCT CCTTGCAATTTTG
	P7	CGGGGT**G**TTG**G**GCTG**G** TGCGTGTCCGGGCC	TGCTCTTGTTTCTACAA GGCAGTATGTTCC
	P8	CATACG**G**CGC**G**GGGT**G** ATCGTGCAAGCGGT	TATTTTCAGAAATTCGT TCACCAAGTGGAG
	dP3	ATCCCA***T***GAT***A***TCTG***T***G TCCGTCCTATTAC	AGTGCCGGGTTGGTGCT GTGGTTAGTGGGG
	dP5	CGGGGT***C***TTG***C***GCTG***C***T GCGTGTCCGGGCC	TGCTCTTGTTTCTACAA GGCAGTATGTTCC
	dP8	CATACG***A***CG**GGGT**A TCGTGCAAGCGGT	TATTTTCAGAAATTCGT TCACCAAGTGGAG
	P1	GAACAATTTGATACCGG TTATTTCCACTCC	GCCAGGCCCCGACTGCT ATTGCACTCAACG
	P2	TCCTTGTTCCGGGTGTA GGTCTTTGTGCCT	GATCTTGGGCTGAATGT GATGACGTATCAA

cdt-1	P3	AGTGCCGGGTTGGTGCT GTGGTTAGTGGGG	ATCTCCGGTAGACCTAT CTGGACTGGTGCG
	P4	ATGGGCTAACGTGAGCT CCTTGCAATTTTG	TCCGGATGCCGGAGAC AAGCTGGATTAGTG
	P5	TGCTCTTGTTTCTACAA GGCAGTATGTTCC	CTTCACCAAAGCCTGTC GGCGTCCCCCCAC
	P6	ATGGGCTAACGTGAGCT CCTTGCAATTTTG	ATCACGTCTATTAAGCC AAAGCGACTCAAG
	P7	TGCTCTTGTTTCTACAA GGCAGTATGTTCC	CTTCACCAAAGCCTGTC GGCGTCCCCCCAC
	P8	TATTTTCAGAAATTCGT TCACCAAGTGGAG	GCGTATACAAAGGGTTC CAGTTTCGGGACA

Table S2. **Primers used in this work.** Bold letters denote restriction sites. To construct a 90-bp fragment carrying a 30-bp perturbation, a typical PCR was performed using one primer from the first section of the table and its corresponding primer from the second section of the table, e.g., to introduce a perturbation upstream to the TEF2 promoter, the 90-bp primer named *perturbed TEF2p-F* was used with *ext-TEF2p-R*.

Description	Name	Sequence (5' to 3')
Incorporation of perturbations	perturbed TEF2p-F	cctcactaaagggaacaaaagctg**gagctc**nnnnnnnnnnnnnn nnnnnnnnnnnnnnnnggggccgtatacttacatatagtagatgtc
	perturbed TEF2t-R	ggcgaattgggtaccgggcccccc**ctcgag**nnnnnnnnnnnnnn nnnnnnnnnnnnnnnnggggtagcgacggattaatggcatacttgt
	perturbed PYK1p-F	aagctggagctccaccgcggtggcggccgc**tctaga**nnnnnnn nnnnnnnnnnnnnnnnnnnnnnnaacggcgggattcctctatg
	perturbed TEF1p-F	gctcttattgaccacacctctaccggcatg**tccgga**nnnnnnnnn nnnnnnnnnnnnnnnnnnnnnatagcttcaaaatgtttctactcc
	perturbed PGK1t-R	gacggtatcgataagcttgatatcgaattc**ctgcag**nnnnnnnnn nnnnnnnnnnnnnnnnnnnnncaggaagaatacactatactgg
	perturbed PGK1p-F	aacaaaagctggagctccaccgcggtggcggccgc**actag**tnnnnn nnnnnnnnnnnnnnnnnnnnnnnngtgagtaaggaaagagtg
	perturbed TDH3p-F	attaaaaaaaaattcgctcctcttttaatgcct**catatg**nnnnnnnnnnn nnnnnnnnnnnnnnnnnnncagttcgagtttatcattatc
	perturbed CCW12p-F	gagaacgtcggcggttaaaatatattaccctgaacg**cttaag**nnnnnn nnnnnnnnnnnnnnnnnnnnnnnncacccatgaaccacacgg
	disrupted P3-PYK1p-F	aagctggagctccaccgcggtggcggccgc**tctaga**atccc atgatatctg**t**gtccgtcctattacaacggcgggattcctctatg
	disrupted P5-PYK1p-F	aagctggagctccaccgcggtggcggccgc**tctaga**cgggg tcttgcgctgctgcgtgtccgggccaacggcgggattcctctatg
	disrupted P8-PYK1p-F	aagctggagctccaccgcggtggcggccgc**tctaga**catacg acgcagggtaatcgtgcaagcggtaacggcgggattcctctatg
Synthesis of dsDNA fragments	ext-TEF2p-R	atgtaagtatacggcccc
	ext-TEF2t-F	ttaatccgtcgctacccc
	ext-PYK1p-R	catagaggaatcccgccgtt
	ext-TEF1p-R	ggagtagaaacattttgaagc
	ext-PGK1t-F	ccagtatagtgtattcttcc
	ext-PGK1p-R	gtgagtaaggaaagagtg
	ext-TDH3p-R	gttcgagtttatcattatc
	ext-CCW12p-R	cacccatgaaccacacgg
	gfp-F	ggggccgtatacttacat
	gfp-R	ggggtagcgacggattaa
	cdt1-F	gcttcaaaatgtttctactcc
	cdt1-R	aggaagaatacactatactgg
	ghl1-F	aacggcgggattcctctatg
	ghl1-R	catgccggtagaggtgtg

Cloning	ARO4-F	gtgagtaaggaaagagtg
	ARO4-R	acgtcatatgaggcattaaaag
	TYRH-F	cagttcgagtttatcattatc
	TYRH-R	actgcttaagcgttcagggtaatatattttaac
	DOD-F	cacccatgaaccacacgg
	DOD-R	gtaccgggcccccccctcgaggtcgacggt atcgataagcttgaaatggggagcgatttgc
qPCR	LEU-F	cagacaagaacaccgcatttg
	LEU-R	cagacaagatagtggcgatagg
	ALG9-F	gatggtcttcttccaggtgatt
	ALG9-R	ctagtgatcggccactctttac
	q-ghl1-F	gcgtcaagtacttcaacgacta
	q-ghl1-R	catcaagctccaggcaagata
	q-cdt1-F	gtcatcttctccttcacctacac
	q-cdt1-R	gttgacaatgaagccggaaag
	ghl1-P1-F	cgcctttgagtgagctgata
	ghl1-P1-R	agaggcggtttgcgtatt
	ghl1-P2-F	agtgagcgcaacgcaatta
	ghl1-P2-R	cataggagccggaagcataaag
	ghl1-P3-F	ctatggcgtgtgatgtctgtat
	ghl1-P3-R	aatctcccaatctgtcgatgg
	cdt1-P1-F	ttgtcggcttgtctaccttg
	cdt1-P1-R	tcgcttatttagaagtgtcaacaac
	cdt1-P2-F	tgttcttgagtaactctttcctgtag
	cdt1-P2-R	acatgccggtagaggtgt
	cdt1-P3-F	catcgccgtaccacttcaa
	cdt1-P3-R	ttccaaacctttagtacgggtaa
	cdt1-P4-F	aaagagaccgcctcgtttc
	cdt1-P4-R	tcaatgggaggtcatcgaaag
	REC102-F	ccgattgaagtggcaacaaag
	REC102-R	gtcttgggcaatcatgtta

Table S3. Specific growth rates for the strains bearing the cellobiose pathway with the original perturbations and the disrupted perturbations.

Strain	μ (h^{-1})
High 3	0.028 ± 0.002
High 5	0.026 ± 0.002
High 8	0.031 ± 0.003
Disrupted High 3	0.013 ± 0.003
Disrupted High 5	0.020 ± 0.004
Disrupted High 8	0.019 ± 0.003

Table S4. Strains used in this work.

Strain	Genotype	Plasmid
Escherichia coli BW25141	(*lacI*^{q} *rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ Δ*phoBR580 hsdR514* Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$ *galU95 endA*$_{BT333}$ *uidA*(Δ*MluI*)::pir^{+} *recA1*	None
S. cerevisiae YSG50	*MATα (ade2-1, ade3Δ22, ura3-1, his3-11,15, trp1-1, leu2-3,112, can1-100)*	None
YSG50 GFP^{+} High	*MATα (ade2-1, ade3Δ22, ura3-1, his3-11,15, trp1-1, leu2-3,112, can1-100)*	pRS416-PX-TEF2p-GFP-TEF2t
YSG50 GFP^{+} Mut	*MATα (ade2-1, ade3Δ22, ura3-1, his3-11,15, trp1-1, leu2-3,112, can1-100)*	pRS416-MX-TEF2p-GFP-TEF2t
YSG50 CDT^{+}BGL^{+} High	*MATα (ade2-1, ade3Δ22, ura3-1, his3-11,15, trp1-1, leu2-3,112, can1-100)*	pRS415-PX-CB
YSG50 ARO^{+}TYRH^{+}DOD^{+}	*MATα (ade2-1, ade3Δ22, ura3-1, his3-11,15, trp1-1, leu2-3,112, can1-100)*	pRS415-PX-BX

Table S5. Plasmids constructed in this work. The bold code represents a plasmid variant carrying a specific perturbation sequence (See Table S1). Plasmids were assembled by DNA assembler[4].

Plasmid	Characteristics
pRS416-P**X**-TEF2p-GFP-TEF2t	TEF2p-*gfp*-TEF2t cassette in pRS416 backbone; 30-bp perturbations were added to flank the GFP transcriptional unit.
pRS416-M**X**-TEF2p-GFP-TEF2t	TEF2p promoter in the TEF2p-GFP-TEF2t cassette was mutagenized.
pRS415-P**X**-CB	PYK1p-*ghl-1*-ADH1t and TEF1p-*cdt-1*-PGK1t cassettes in pRS415 backbone; 30-bp perturbations flanking each transcriptional unit.
pRS415-dP**X**-CB	PYK1p-*ghl-1*-ADH1t and TEF1p-*cdt-1*-PGK1t cassettes in pRS415 backbone; the conserved nucleotides in the perturbation sequences were mutated.
pRS415-P**X**-BX	PGK1p-*ARO4-PGK1*t, TDH3p-*TYRH*-TDH1t, CCW12p-*DOD*-ADH1t cassettes in pRS415 backbone; 30-bp perturbations flanking each transcriptional unit.

REFERENCES

(1) Xi, L., Fondufe-Mittendorf, Y., Xia, L., Flatow, J., Widom, J., and Wang, J.-P. (2010) Predicting nucleosome positioning using a duration Hidden Markov Model. *BMC Bioinformatics 11*, 346.
(2) Wang, J.-P., Fondufe-Mittendorf, Y., Xi, L., Tsai, G.-F., Segal, E., and Widom, J. (2008) Preferentially Quantized Linker DNA Lengths in Saccharomyces cerevisiae. *PLoS Comput. Biol.* (Stormo, G., Ed.) *4*, e1000175.
(3) Flagfeldt, D. B., Siewers, V., Huang, L., and Nielsen, J. (2009) Characterization of chromosomal integration sites for heterologous gene expression in Saccharomyces cerevisiae. *Yeast 26*, 545–551.
(4) Shao, Z., Zhao, H., and Zhao, H. (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res. 37*, 1–10.