Supplementary information - Manipulating Nucleosome Disfavoring Sequences Allows Fine-Tune Regulation of Gene Expression in Yeast

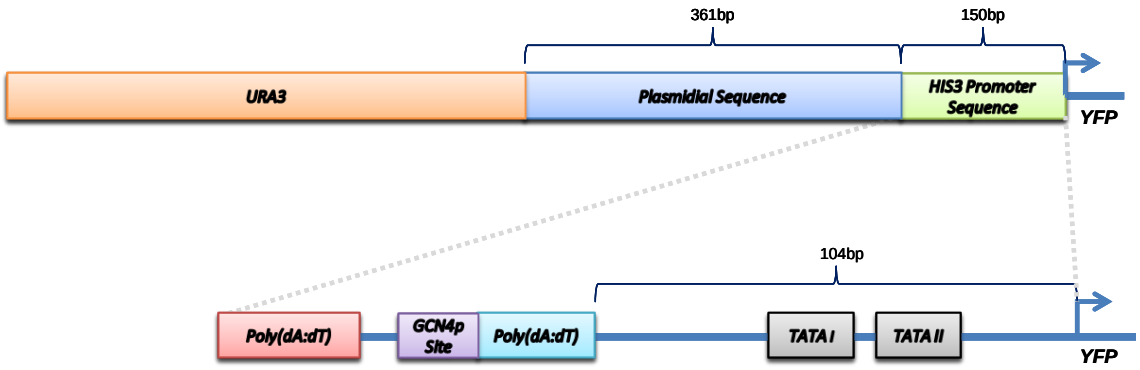
Tali Raveh-Sadka*[1,2], Michal Levo*[1,2], Uri Shabi[3], Boaz Shany[1,2], Leeat Keren[1,2], Maya Lotan-Pompan[1,2], Danny Zeevi[1,2], Eilon Sharon[1,2], Adina Weinberger[1,2†], Eran Segal[1,2†]

Nature Genetics: doi:10.1038/ng.2305

Contents

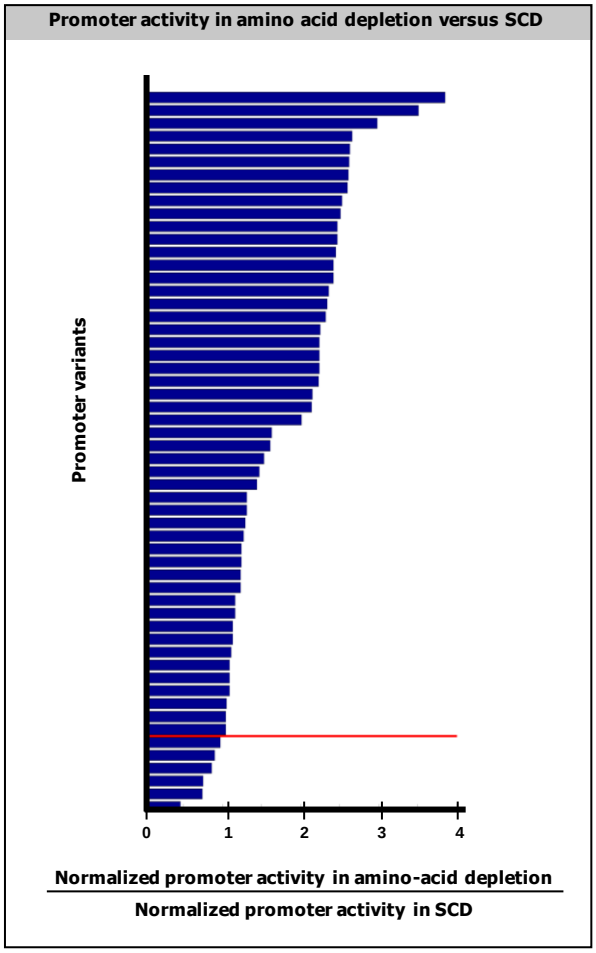
Nature Genetics: doi:10.1038/ng.2305

Supplementary Figures



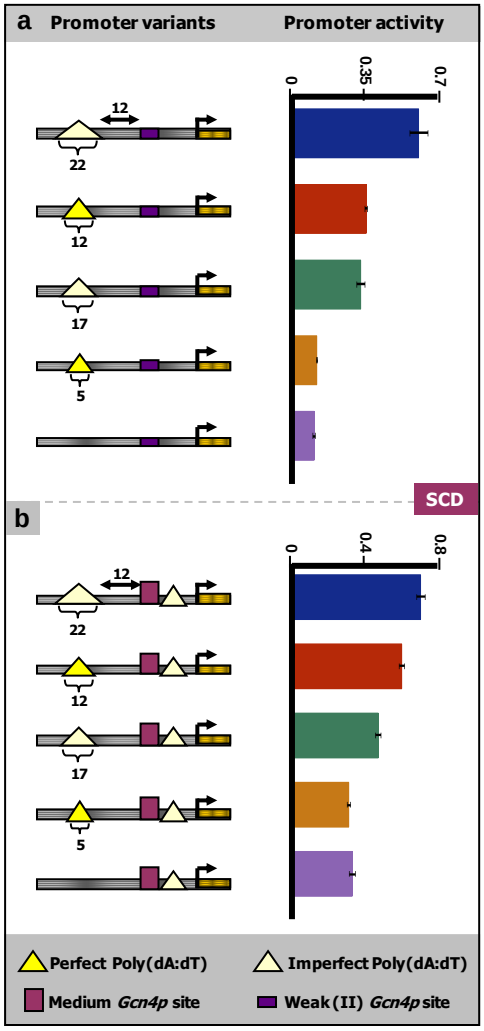
Supplementary Figure 1. Schematic view of the basic design for our promoter constructs

Nature Genetics: doi:10.1038/ng.2305



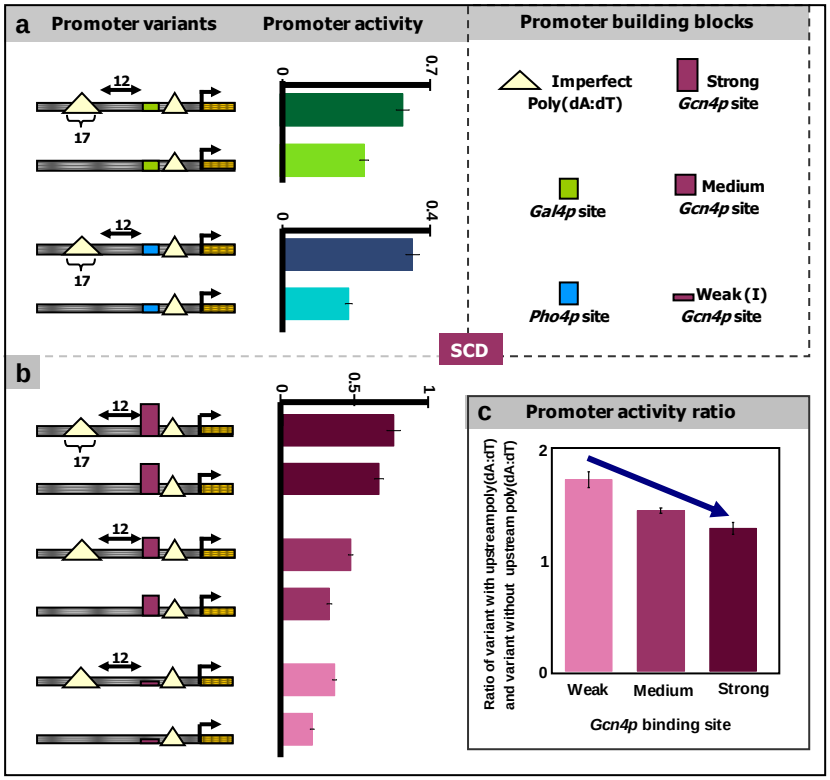
Supplementary Figure 2. Promoter activity in amino acid depletion and in synthetic complete medium. For promoter variants containing a Gcn4p site (variants V1-V8,V13-V60, see **Supp. Table 1**) we compared promoter activity in synthetic complete medium and in amino acid depletion, where the concentration of active Gcn4p is assumed to be higher. To account for general differences between the two growth conditions the measured promoter activity values were first normalized by the change in median mCherry promoter activity between conditions. For each variant, shown is the ratio between the normalized promoter activity values under amino acid depletion and in synthetic complete medium. Ratios larger than one (above red line) signify higher promoter activity under amino acid depletion compared to synthetic complete medium. Promoter variants with ratios smaller than one (below red line) include variants with extremely low promoter activity in both conditions, in which the upstream poly(dA:dT) is relatively far from the Gcn4p site, or is missing completely (V4-5,V30,V34-36). See additional details on this comparison in **Supp. note**.

Nature Genetics: doi:10.1038/ng.2305



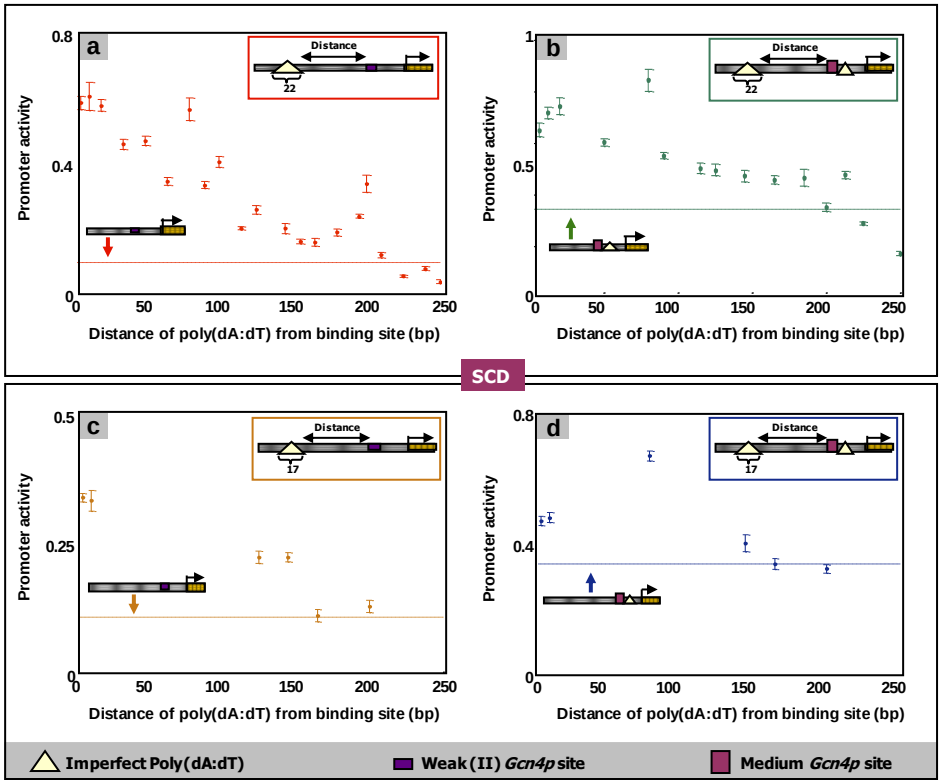
Supplementary Figure 3. Longer and more perfect poly(dA:dT) elements have a stronger transcriptional effect **(a)** Schematic illustration and activities of promoter variants that differ in the length and composition of the poly(dA:dT) element upstream of the Gcn4p site (see sequences used in **Supp. note**). Shown are mean promoter activity values ± two standard errors obtained in 7-14 independent experiments. Cells were grown and measured in synthetic complete medium. **(b)** Same as (a), but for a different sequence context, with a different Gcn4p site and with a poly(dA:dT) element downstream of the site, as in the native *HIS3* promoter.

Nature Genetics: doi:10.1038/ng.2305



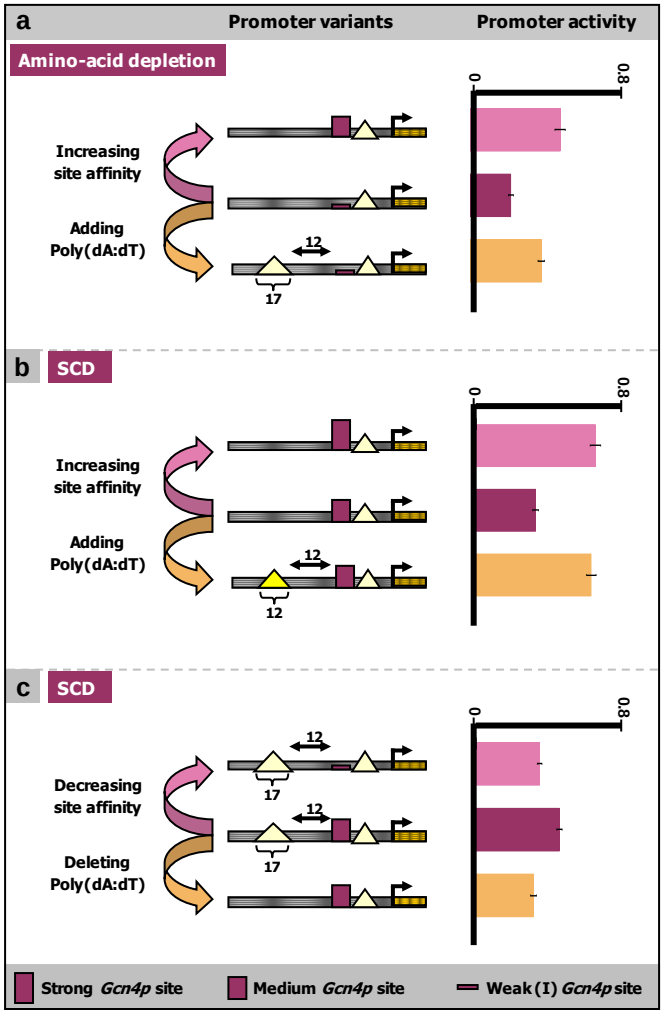
Supplementary Figure 4. The stimulatory nature of the effect of poly(dA:dT) tracts is independent of the regulating factor identity but its magnitude is inversely proportional to the affinity of the factor site (a) Schematic illustrations and activity measurements for promoter variants with or without a poly(dA:dT) tract upstream of either a Gal4p (top) or a Pho4p (bottom) binding site. Measurements are shown as the mean promoter activity ± two standard errors obtained in 7 independent experiments when cells were grown and measured in synthetic complete medium. **(b)** Schematic illustrations and activity measurements for promoter variants with or without a poly(dA:dT) tract upstream of a strong (top), medium (middle) or weak (bottom) binding site for Gcn4p. Measurements are shown as the mean promoter activity ± two standard errors obtained in 7-14 independent experiments when cells were grown and measured in synthetic complete medium. **(c)** For each pair of promoter variants with the same Gcn4p binding site from (a), shown is the ratio of promoter activity values between the variant containing the upstream poly(dA:dT) element and the variant lacking this element. The blue arrow indicates that the magnitude of the transcriptional effect of the poly(dA:dT) tract is inversely proportional to the affinity of the Gcn4p site.

Nature Genetics: doi:10.1038/ng.2305



Supplementary Figure 5. The transcriptional effect of poly(dA:dT) tracts depends on their distance from transcription factor binding sites. **(a)** Activity measurements for promoter variants that differ in the location of an upstream 22bp poly(dA:dT) tract (see illustration in the inset). For each variant, shown is the mean promoter activity ± two standard errors obtained from 7-14 independent experiments when cells were grown in synthetic complete medium. Horizontal line represent the promoter activity of a variant that lacks the upstream poly(dA:dT) tract. **(b)** As in (a), but for promoter variants constructed in another promoter context, with a different Gcn4p site and with a poly(dA:dT) tract downstream of the site (see illustration in the inset). **(c)** As in (a), but for promoter variants constructed in another promoter context, with an upstream poly(dA:dT) tract of length 17bp, identical to the upstream tract in the native *HIS3* promoter (see illustration in the inset). **(d)** As in (a), but for promoter variants constructed in another promoter context, with an upstream poly(dA:dT) tract of length 17bp , identical to the upstream tract in the native *HIS3* promoter (see illustration in the inset).

Nature Genetics: doi:10.1038/ng.2305

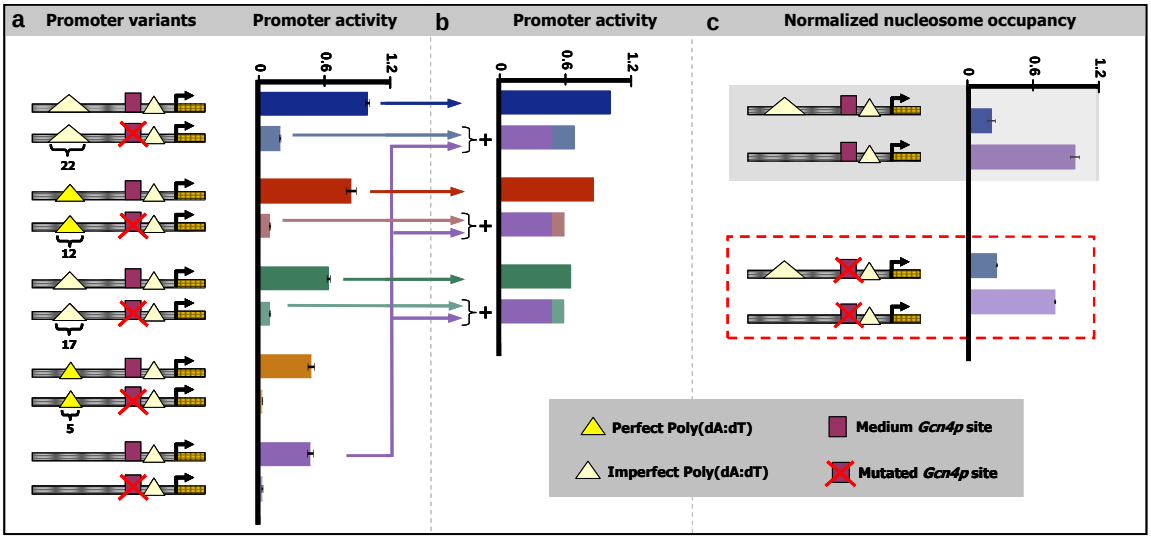


Supplementary Figure 6. The transcriptional effect of manipulations to poly(dA:dT) tracts can be comparable in magnitude to the transcriptional effect of alterations in the affinity of transcription factor binding sites. **(a)** Schematic illustrations and activity measurements for promoter variants with (bottom) or without (top, middle) a poly(dA:dT) tract upstream of a medium (top) or weak (middle, bottom) binding site for Gcn4p. Measurements are shown as mean promoter activity ± two standard errors obtained in 6-12 independent experiments when cells were grown and measured under conditions of amino-acid depletion. **(b)** Schematic illustrations and activity measurements for promoter variants with (bottom) or without (top, middle) a poly(dA:dT) tract upstream of a strong (top) or medium (middle, bottom) binding site for Gcn4p. Measurements are shown as the mean promoter activity ± two standard errors obtained in 7-14 independent experiments when cells were grown and measured in synthetic complete medium. **(c)** Schematic illustrations and activity measurements for

Nature Genetics: doi:10.1038/ng.2305

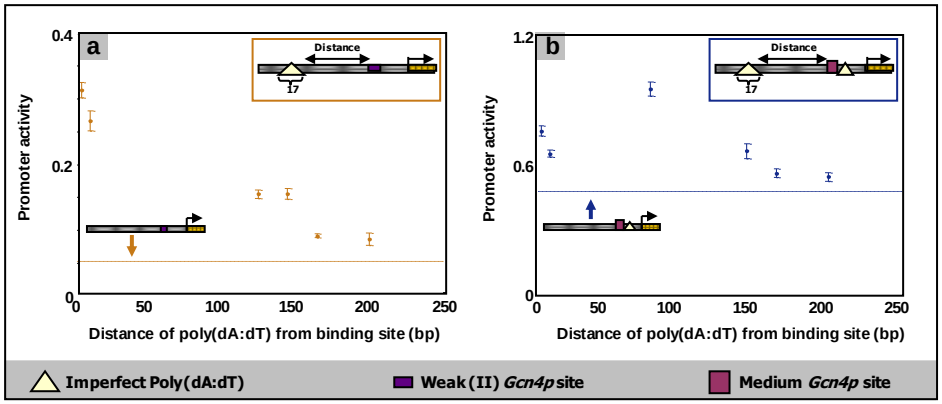
promoter variants with (top, middle) or without (bottom) a poly(dA:dT) tract upstream of a weak (top) or medium (middle, bottom) binding site for Gcn4p. Measurements are shown as the mean promoter activity ± two standard errors obtained in 7-14 independent experiments when cells were grown and measured in synthetic complete medium.

Nature Genetics: doi:10.1038/ng.2305



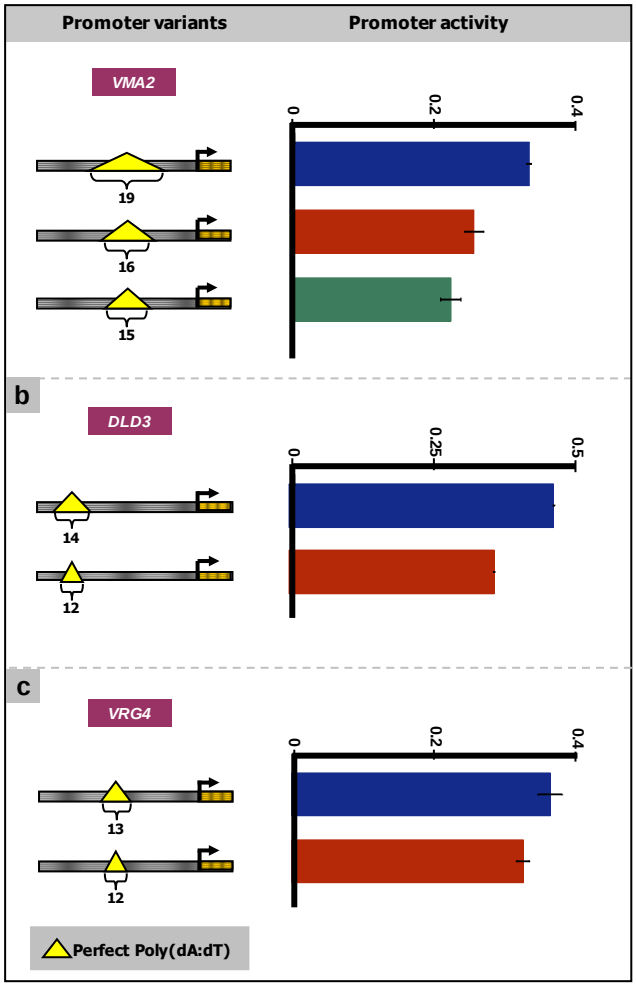
Supplementary Figure 7. An intact Gcn4p site results in amplification of the transcriptional effect of the poly(dA:dT) tract, but is not required for the observed change in nucleosome occupancy in the vicinity of the tract. See detailed discussion on these measurements in supplementary note. **(a)** Schematic illustration and promoter activity values for the promoter variants in **Fig. 2b** and for similar variants in which the Gcn4p site was mutated (see **Supp. note**). Shown are the mean promoter activity ± two standard errors obtained in 6-12 independent experiments. Cells were grown and measured under conditions of amino-acid depletion. For similar results in the sequence context in which the downstream poly(dA:dT) element is deleted, see **Supp. Table 1 (b)** Shown is a comparison of the promoter activity for variants containing both an upstream poly(dA:dT) tract and the non-mutated Gcn4p site (presented also in (a)) versus the sum of promoter activities for variants containing only one of these elements (the sum of two variants presented in (a)) **(c)** Same as **Fig. 2d**, for the variants containing a long (22bp) upstream poly(dA:dT) or lacking this tract, in the presence of the non-mutated or the mutated Gcn4p site.

Nature Genetics: doi:10.1038/ng.2305



Supplementary Figure 8. The transcriptional effect of poly(dA:dT) tracts depends on their distance from other promoter elements. **(a)** Similar to **Fig. 4**, shown are the activity measurements for promoter variants that differ in the location of an upstream 17bp poly(dA:dT) tract, identical to the upstream tract in the native *HIS3* promoter (see illustration in the inset). For each variant, shown is the mean promoter activity ± two standard errors obtained from 6-12 independent experiments when cells were grown under conditions of amino-acid depletion. A horizontal line represents the promoter activity of a variant that lacks the upstream poly(dA:dT) tract. **(b)** As in (a), but for promoter variants constructed in another promoter context, with a different Gcn4p site and with a poly(dA:dT) tract downstream of the site (see illustration in the inset).

Nature Genetics: doi:10.1038/ng.2305

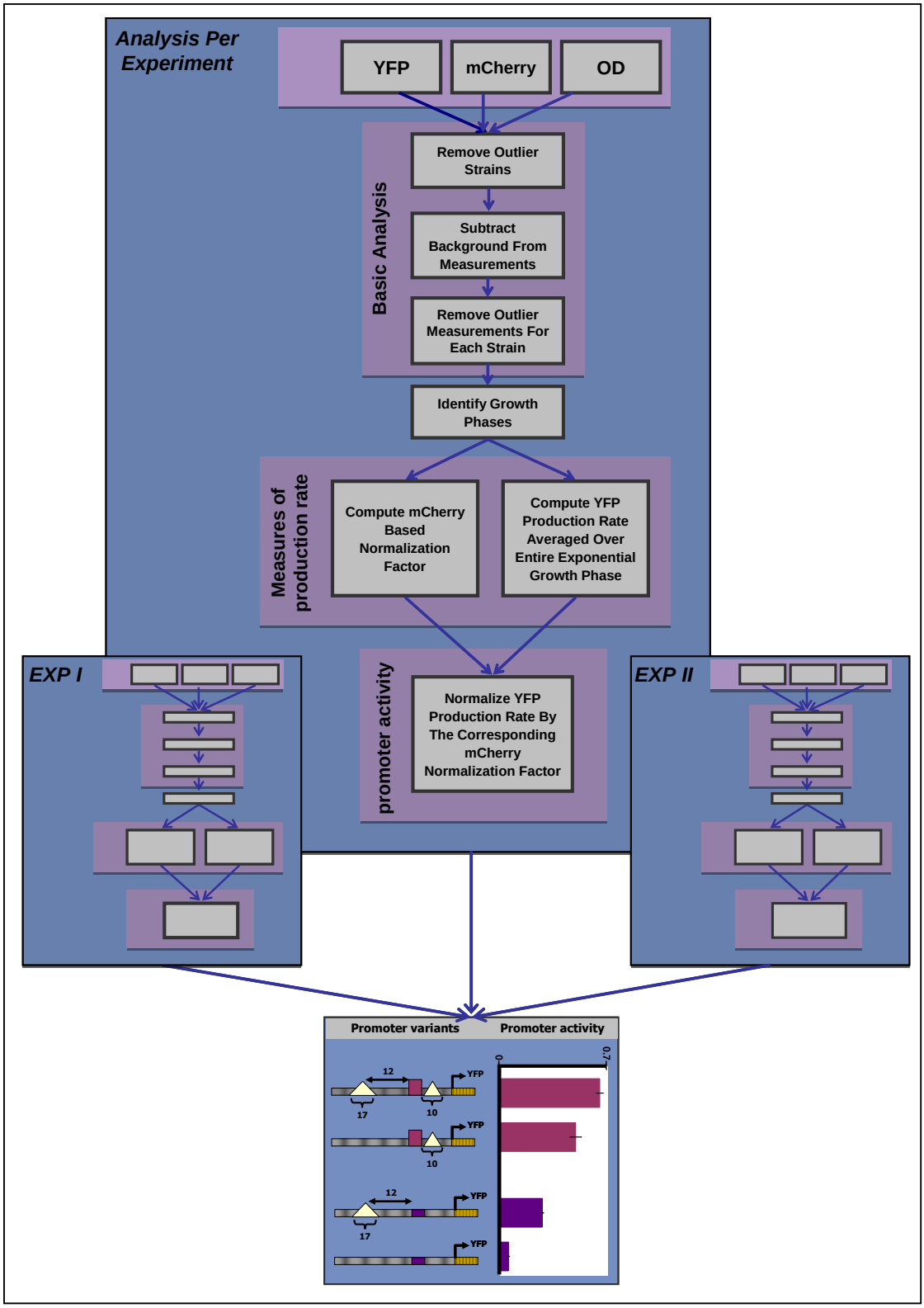


Supplementary Figure 9. Small changes in the length of poly(dA:dT) elements can significantly affect promoter activity **(a)** Shown are promoter activity measurements of the native VMA2 promoter, containing a perfect 19bp poly(dA:dT) tract (blue bar), and two additional promoter variants in which this tract is slightly shorter (16bp, red bar, and 15bp, green bar). Measurements are shown as the mean promoter activity ± two standard errors obtained in 3 independent experiments when cells were grown and measured in synthetic complete medium **(b)** Shown are promoter activity measurements of two variants of the native DLD3 promoter, containing a perfect 13bp poly(dA:dT) tract, one in which the tract is slightly shorter than the native tract (12bp, red bar), and one in which the tract is slightly longer (14bp, blue bar). Measurements are shown as the mean promoter activity ± two standard errors obtained in 3 independent experiments when cells were grown and measured in synthetic complete medium **(c)** Shown are promoter activity measurements of the native VRG4 promoter, containing a perfect 13bp

Nature Genetics: doi:10.1038/ng.2305

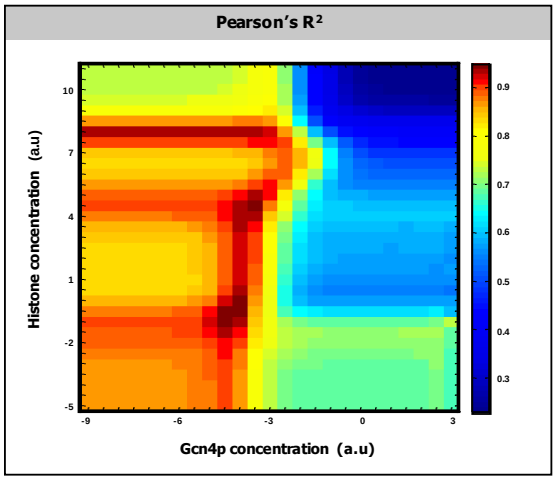
poly(dA:dT) tract (blue bar), and an additional promoter variant in which this tract is slightly shorter (12bp tract, red bar). Measurements are shown as the mean promoter activity ± two standard errors obtained in 3 independent experiments when cells were grown and measured in synthetic complete medium

Nature Genetics: doi:10.1038/ng.2305



Supplementary Figure 10. General scheme of the plate reader measurements analysis pipeline. See **Supp. note** for detailed description of each step.

Nature Genetics: doi:10.1038/ng.2305

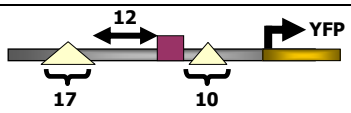
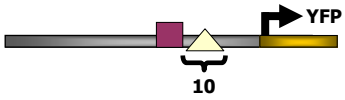
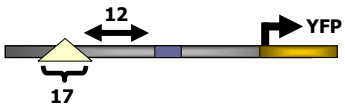


Supplementary Figure 11. Pearson's R^2 values for scanned parameter settings. Shown is a heatmap of Pearson's R^2 values that were obtained when the model was applied to variants V1-V8, V13-V16, V18, V38 using each set of scanned parameters (for histone and Gcn4p concentrations in $\log_{10}$ scale). See **Supp. note** for additional details.

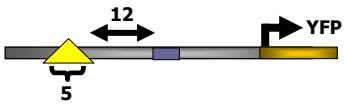
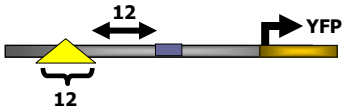
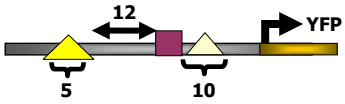
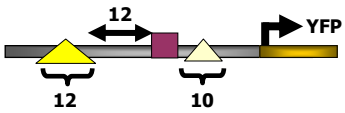
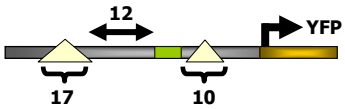
Nature Genetics: doi:10.1038/ng.2305

Supplementary Tables

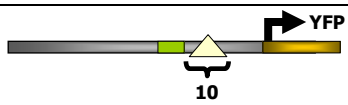
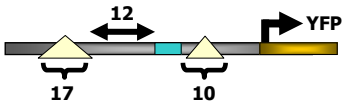
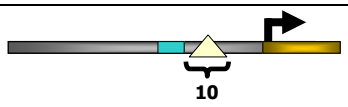
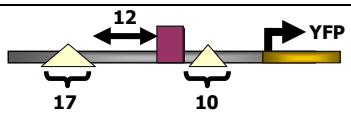
Supplementary Table 1. Promoter library.

Promoter Variant Number	Strain Illustration	Promoter activity in activating condition	Standard error of promoter activity in activating condition
V1		**0.652**	**0.008**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTCA TTTTTTTTTTTCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATA AAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V2		**0.482**	**0.014**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGCCACC TAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTGATTTC TTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V3		**0.266**	**0.008**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTCA TTTTTTTTTTTCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTG ATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V4		**0.005**	**0.004**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGCCACC		

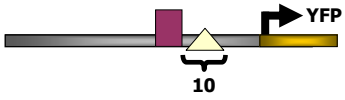
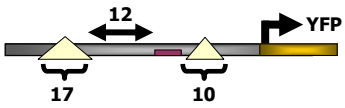
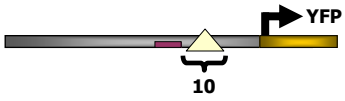
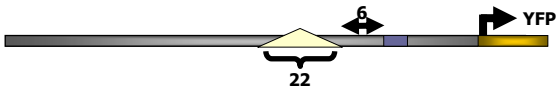
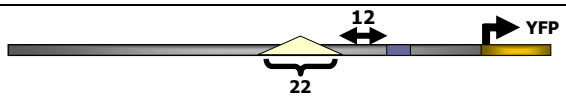
Nature Genetics: doi:10.1038/ng.2305

	TAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTGATTTCTTCGAAGAAT ATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V5		**0.070**	**0.003**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTTC CACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTGATTTCTTCGAA GAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V6		**0.295**	**0.010**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTTT TTTTTTCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTGATTTC TTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V7		**0.491**	**0.015**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTTC CACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTGA TTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V8		**0.861**	**0.024**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTTT TTTTTTCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGT AATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V9		**1.902**	**0.030**

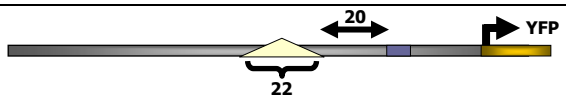
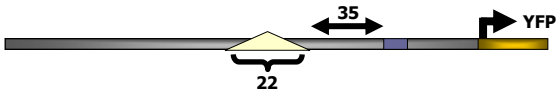
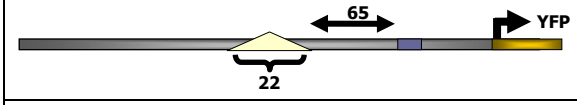
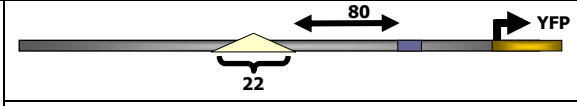
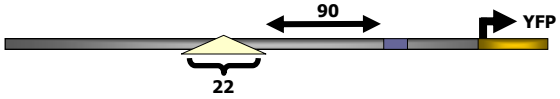
Nature Genetics: doi:10.1038/ng.2305

	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTCA TTTTTTTTTTTCCACCTAGCGGACGGAAGACTCTCCTCCGTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTAT ACATTATATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V10		**1.604**	**0.063**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGCCACC TAGCGGACGGAAGACTCTCCTCCGTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTA ATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V11		**0.648**	**0.011**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTCA TTTTTTTTTTTCCACCTAGCGGACACGTGCTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATA AAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V12		**0.344**	**0.003**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGCCACC TAGCGGACACGTGCTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTGATTTC TTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V13		**1.703**	**0.037**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTCA TTTTTTTTTTTCCACCTAGCGGATGACTCATTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATA AAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		

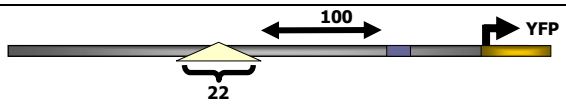
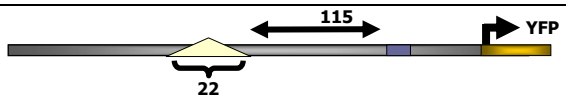
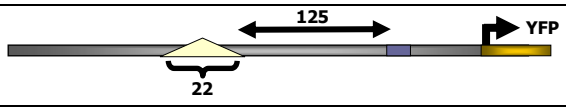
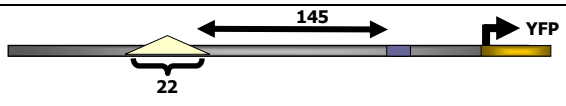
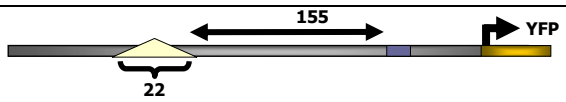
Nature Genetics: doi:10.1038/ng.2305

V14		**1.632**	**0.056**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGCCACC TAGCGGATGACTCATTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTGATTTC TTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V15		**0.382**	**0.009**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTCA TTTTTTTTTTTCCACCTAGCGGATGACTCGTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATA AAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V16		**0.219**	**0.007**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGCCACC TAGCGGATGACTCGTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTGATTTC TTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V17		**0.602**	**0.010**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGCCACC TTTTTTTTTTCATTTTTTTTTTTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTA ATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V18		**0.492**	**0.011**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTTT		

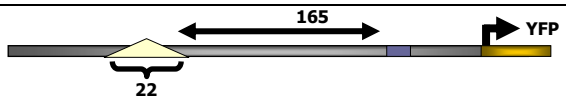
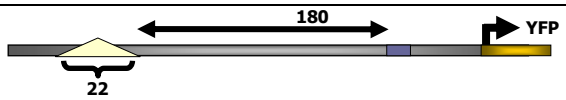
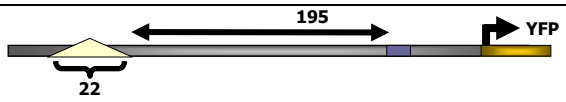
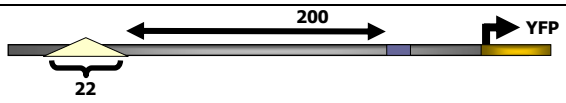
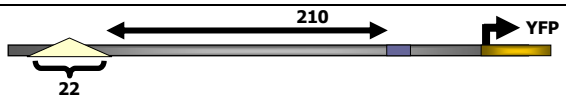
Nature Genetics: doi:10.1038/ng.2305

	TTTCATTTTTTTTTTTTCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTA ATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V19		**0.454**	**0.007**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGATTTTTTTTTCATTTT TTTTTTTAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTA ATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V20		**0.358**	**0.003**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGTTTTTTTTTCATTTTTTTTTTTCGAGAAAGG AAGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTA ATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V21		**0.272**	**0.003**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTATTTTTTTTTCATTTTTTTTTTTTGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGG AAGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTA ATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V22		**0.392**	**0.009**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGTTTTTTTTTCATTTTTTTTTTTTGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGG AAGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTA ATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V23		**0.243**	**0.004**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGTTTTTTTT		

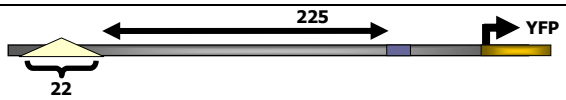
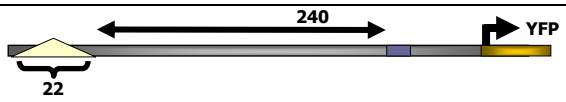
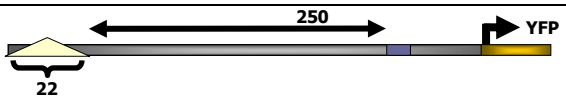
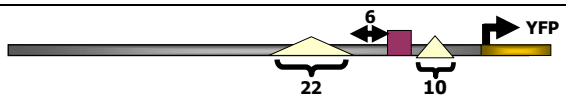
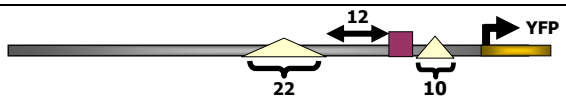
Nature Genetics: doi:10.1038/ng.2305

	TCATTTTTTTTTTTAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGG AAGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTA ATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V24		**0.286**	**0.007**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCTTTTTTTTTCATTTTTTTT TTTACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGG AAGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTA ATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V25		**0.175**	**0.003**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTTGGGGTCTTTTTTTTTTCATTTTTTTTTTTGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAA TGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V26		**0.196**	**0.005**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTTTTTTCATTTTTTTTTTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAA TGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V27		**0.165**	**0.006**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGATTTTTTTTTCATTTTTTTTTTTTACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAA TGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V28		**0.126**	**0.003**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG		

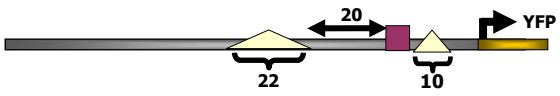
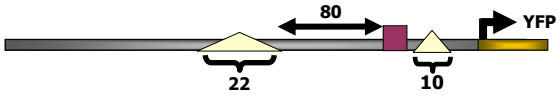
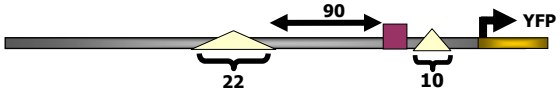
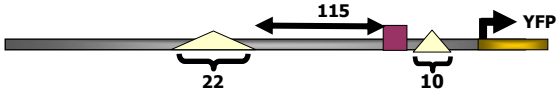
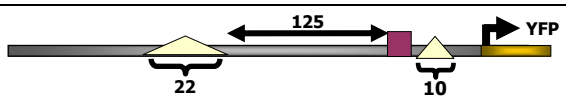
Nature Genetics: doi:10.1038/ng.2305

	ATGGCCTTTTTTTTTTCATTTTTTTTTTTTCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAA TGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V29		**0.121**	**0.005**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGTTTTT TTTTCATTTTTTTTTTTTGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAA TGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V30		**0.122**	**0.003**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGATTTTTTTTTCATTTTTTTTTTT AAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAA TGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V31		**0.225**	**0.006**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCTTTTTTTTTCATTTTTTTTTTTTAACGTCAAAGGGCGA AAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAA TGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V32		**0.228**	**0.008**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGTTTTTTTTTTCATTTTTTTTTTTTACTCCAACGTCAAAGGGCGA AAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAA TGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V33		**0.086**	**0.003**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT		

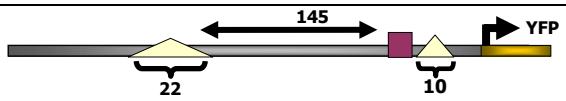
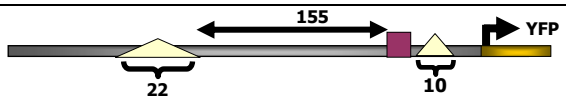
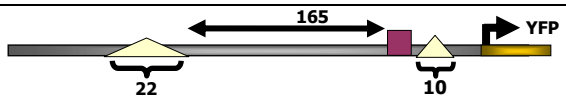
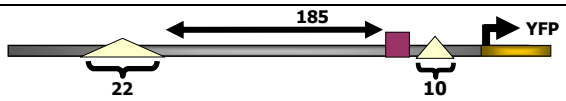
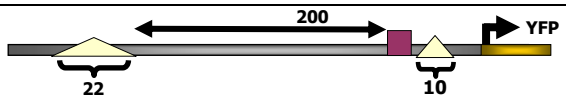
Nature Genetics: doi:10.1038/ng.2305

	TCCAGTTTGGAACAAGAGTCCACTATTATTTTTTTTTCATTTTTTTTTTTTAAGAACGTGGACTCCAACGTCAAAGGGCGA AAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAA TGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V34		**0.037**	**0.002**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACTTTTTTTTTCATTTTTTTTTTTTAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGA AAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAA TGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V35		**0.043**	**0.004**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTTT TTTTTTTCATTTTTTTTTTTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGA AAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAA TGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V36		**0.015**	**0.004**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGTTTTTTTTTCATT TTTTTTTTTGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGA AAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAA TGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V37		**1.220**	**0.029**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGCCACC TTTTTTTTTTCATTTTTTTTTTTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATT ATATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V38		**1.013**	**0.010**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC		

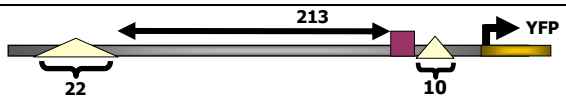
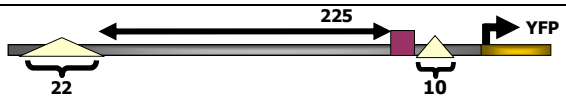
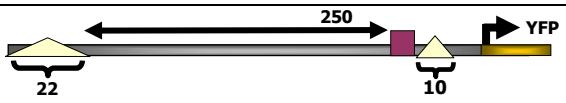
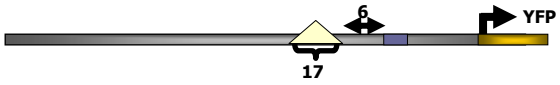
Nature Genetics: doi:10.1038/ng.2305

	ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTTT TTTCATTTTTTTTTTTTCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATT ATATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V39		**1.047**	**0.022**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGATTTTTTTTTCATTTT TTTTTTTAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATT ATATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V40		**1.066**	**0.020**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGTTTTTTTTTCATTTTTTTTTTTGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGG AAGGGAAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATT ATATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V41		**0.840**	**0.010**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGTTTTTTTT TCATTTTTTTTTTTTAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGG AAGGGAAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATT ATATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V42		**0.819**	**0.005**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCTTTTTTTTTTCATTTTTTTTTTTGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V43		**0.769**	**0.011**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA		

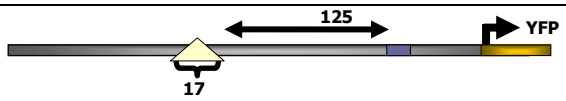
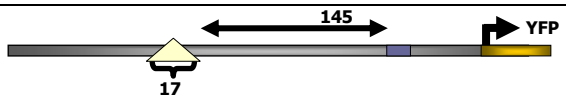
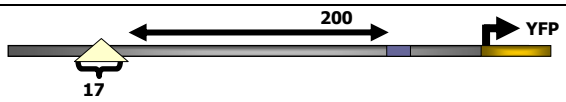
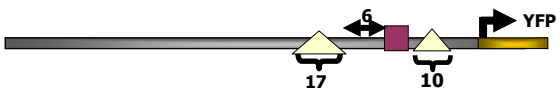
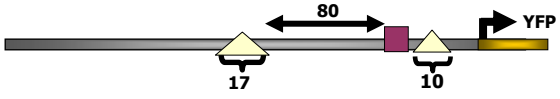
Nature Genetics: doi:10.1038/ng.2305

	AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTTTTTTCATTTTTTTTTTTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V44		**0.726**	**0.012**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGATTTTTTTTTCATTTTTTTTTTTTACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V45		**0.780**	**0.015**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCTTTTTTTTTCATTTTTTTTTTTCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V46		**0.666**	**0.018**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGTTTTT TTTTCATTTTTTTTTTTTGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V47		**0.690**	**0.013**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGTTTTTTTTTCATTTTTTTTTTTTGGCGA AAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V48		**0.544**	**0.020**

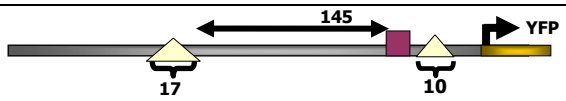
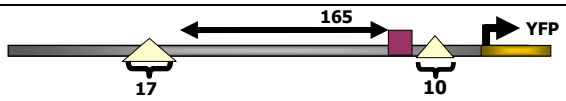
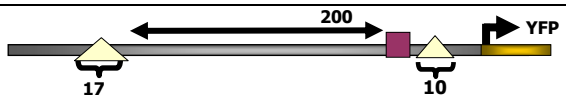
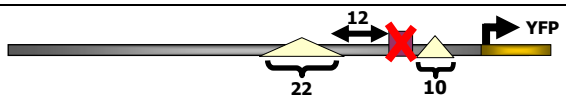
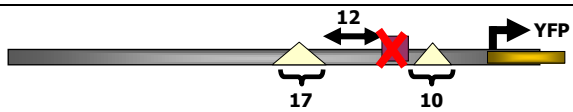
Nature Genetics: doi:10.1038/ng.2305

	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGTTTTTTTTTCATTTTTTTTTTTTACTCCAACGTCAAAGGGCGA AAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V49		**0.701**	**0.027**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTTTTTTTTTCATTTTTTTTTTTTTTAAAGAACGTGGACTCCAACGTCAAAGGGCGA AAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V50		**0.476**	**0.008**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACTTTTTTTTTCATTTTTTTTTTTTAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGA AAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V51		**0.228**	**0.006**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGTTTTTTTTTCATT TTTTTTTTTGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGA AAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAA GCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA AGGGAAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V52		**0.313**	**0.006**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGCCACC TTTTTCATTTTTTTTTTTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGT GATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		

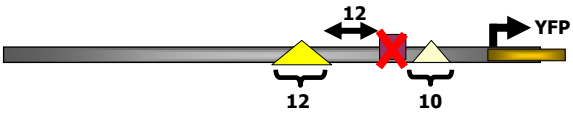
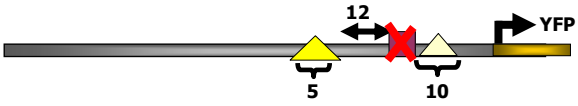
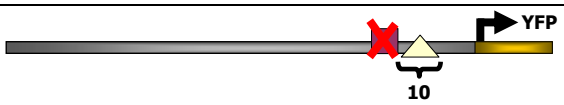
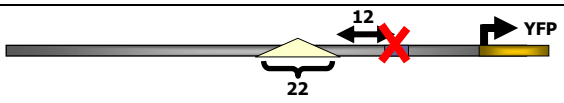
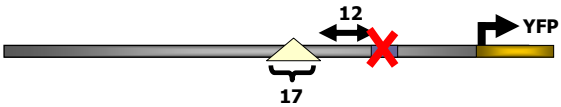
Nature Genetics: doi:10.1038/ng.2305

V53		**0.155**	**0.003**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTCATTTTTTTTTTTTTTTGGGGTCGAGGTGCCGTAAAGCAC TAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG AAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTG ATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V54		**0.155**	**0.004**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGATTTTCATTTTTTTTTTTACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCAC TAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG AAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTG ATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V55		**0.086**	**0.004**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGTTTTCATTTTTTTTTTTTACTCCAACGTCAAAGGGCGAAAAA CCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACT AAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG AAGAAAGCGCCACCTAGCGGATGACTCTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTG ATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V56		**0.753**	**0.011**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGCCACC TTTTTCATTTTTTTTTTTTAGCGGATGACTCTTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATAT AAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V57		**0.943**	**0.016**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGTTTTCATTTTTTTTTTTTGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGG GAAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATAT		

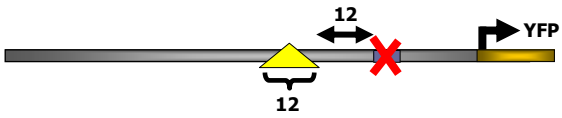
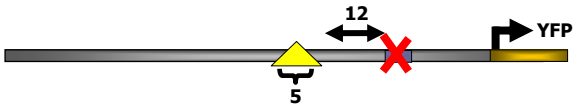
Nature Genetics: doi:10.1038/ng.2305

	AAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V58		**0.665**	**0.017**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGATTTTCATTTTTTTTTTTACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCAC TAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG AAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATA AAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V59		**0.565**	**0.009**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGTTTTC ATTTTTTTTTTTTGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACT AAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG AAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATA AAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V60		**0.548**	**0.009**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGTTTTCATTTTTTTTTTTACTCCAACGTCAAAGGGCGAAAAA CCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACT AAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG AAGAAAGCGCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATA AAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V61		**0.202**	**0.003**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTTT TTTCATTTTTTTTTTTTCCACCTAGCGGATAACACTATTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATT ATATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V62		**0.107**	**0.003**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTCA		

Nature Genetics: doi:10.1038/ng.2305

	TTTTTTTTTTTCCACCTAGCGGATAACACTATTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATA AAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V63		**0.107**	**0.004**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTTT TTTTTTCCACCTAGCGGATAACACTATTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGT AATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V64		**0.038**	**0.002**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTTC CACCTAGCGGATAACACTATTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTG ATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V65		**0.037**	**0.003**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGCCACC TAGCGGATAACACTATTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTGATTTC TTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V66		**0.247**	**0.004**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTTT TTTCATTTTTTTTTTTCCACCTAGCGGATAACACTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTA ATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V67		**0.131**	**0.002**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC		

Nature Genetics: doi:10.1038/ng.2305

	ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTCA TTTTTTTTTTTCCACCTAGCGGATAACACTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTG ATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V68		**0.128**	**0.005**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTTT TTTTTTCCACCTAGCGGATAACACTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTGATTT CTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V69		**0.044**	**0.003**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGTTTTTC CACCTAGCGGATAACACTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTGATTTCTTCGAA GAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		
V70		**0.026**	**0.004**
	CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCTATGCGGTGTGAAATACCGCACAGATGCGTA AGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGT TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTA AAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGCCACC TAGCGGATAACACTAGCGATTGGCATTATCACATAATGAATTATACATTATATAAAGTAATGTGATTTCTTCGAAGAAT ATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG		

	***Gcn4p* binding site (strong)**		**Imperfect poly(dA:dT) tract**
	***Gcn4p* binding site (Medium)**		**perfect poly(dA:dT) tract**
	***Gcn4p* binding site (Weak I)**		***Gal4p* binding site**
	***Gcn4p* binding site (Weak II)**		***Pho4*p binding site**
	Mutated *Gcn4p* binding site		

- Double-headed arrows show the edge to edge distance (in basepairs) between the binding site and the poly(dA:dT) tract.

Nature Genetics: doi:10.1038/ng.2305

Supplementary Table 2. Primers list for construction of promoter variants

Primer Name	Sequence
Promoter fw	CTAAACTCACAAATTAGAGCTTC
Promoter rv	TCGTTTATCTTGCCTGCT
URA3 fw (from pRS426)	CGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGCATCAGAGCAGATTG
URA3 rv (from pRS426)	TCTGTGCGGTATTTCACACC

Supplementary Table 3. Primers list for nucleosome occupancy assay.

Primer Name	Sequence
1fw.1	TAAAGGGAGCCCCCGATTTAG
1fw.2	AACCCTAAAGGGAGCCCCC
1fw.3	AGCCCCCGATTTAGAGCTT
1rv.1	AGAGTCATCCGCTAGGTGG
2fw.1	AACGTGGCGAGAAAGGAAG
2fw.2	GGGAAAGCCGGCGAAC
2fw.3	TTTAGAGCTTGACGGGGAAA
2rv.1	ATTATGTGATAATGCCAATC
3fw.1	CCACCTAGCGGATGACTCT
3rv.1	GTATATTCTTCGAAGAAATCAC
3rv.2	GCCTGCTCATTTTTTAGTATATTC
4fw.1	AGCGATTGGCATTATCA
4rv.1	GCCTTCGTTTATCTTGCCTGCTC
5fw.1	CATTATATAAAGTAATGTGATTTCTTCG
5rv.1	CAACACCAGTGAATAATTCTTCACC
6fw.1	GAATATACTAAAAAATGAGCAGGC
6rv.1	CATCACCATCTAATTCAACC
PHO5_fw	TTTGAATTGTCGAAATGAAACG
PHO5_rv	CTTGCTCTATTTGTTGTTGTTCTT

Nature Genetics: doi:10.1038/ng.2305

Supplementary Table 4. Primers pairs used for each strain in nucleosome occupancy assays. For each strain measured (**See Supp. Table 1**. for visual description and sequence) listed are the primer pairs used (see primer sequences in **Supp. Table 3**).

Strain/Primer Pair	**1**	**2**	**3**	**4**	**5**	**6**
V1	1fw.3-1rv.1	2fw.1- 2rv.1	3fw.1- 3rv.1	4fw.1- 4rv.1	5fw.1- 5rv.1	6fw.1- 6rv.1
V2,V7	1fw.2-1rv.1	2fw.2-2rv.1	3fw.1-3rv.1	4fw.1- 4rv.1	5fw.1- 5rv.1	6fw.1- 6rv.1
V3,V18	1fw.3-1rv.1	2fw.2-2rv.1	3fw.1-3rv.2	4fw.1- 4rv.1	5fw.1- 5rv.1	6fw.1- 6rv.1
V4,V5	1fw.2-1rv.1	2fw.3-2rv.1	3fw.1-3rv.2	4fw.1- 4rv.1	5fw.1- 5rv.1	6fw.1- 6rv.1
V6	1fw.1-1rv.1	2fw.2-2rv.1	3fw.1-3rv.2	4fw.1- 4rv.1	5fw.1- 5rv.1	6fw.1- 6rv.1
V8	1fw.1-1rv.1	2fw.1-2rv.1	3fw.1-3rv.1	4fw.1- 4rv.1	5fw.1- 5rv.1	6fw.1- 6rv.1
V24,V26-27,V29,V33-36	1fw.2-1rv.1		3fw.1-3rv.2		5fw.1- 5rv.1	

Supplementary Table 5. Statistics for the application of the full and simplified models to our promoter variants. Similar improvements in model performance are obtained when modeling nucleosomes with their full known sequence preferences or using only poly(dA:dT)-related sequence preferences (see **Supp. note** for details on the affinity models used). Values represent application of the model to a subset of the variants, with qualitatively predicted effects for manipulations to poly(dA:dT) tracts (V1-V8, V13-V16, V18, V38 – see description and sequences in **Supp. Table 1**). This set was used for the selection of optimal parameters). Values in parentheses denote the statistics computed for the application of the model (with the same parameters) to the entire set of Gcn4p-regulated promoter variants (V1-V8 and V13-V60).

Model \ Statistic	Pearson's R^2	Spearman's rank correlation coefficient	Normalized mutual information
Full model	**0.94 (0.75)**	**0.99 (0.82)**	**0.83 (0.53)**
Simplified based on Kaplan et al. 5-mer statistics	**0.94 (0.81)**	**0.98 (0.87)**	**0.70 (0.66)**
Simplified based only on poly(dA:dT) tracts	**0.93 (0.79)**	**0.96 (0.82)**	**0.65 (0.51)**

Nature Genetics: doi:10.1038/ng.2305

Supplementary Note

Extended information on promoter library design

Detailed Description of Promoter Building Blocks

Gcn4p Sites. The large majority of promoter variants in our library contain a site for the transcriptional activator Gcn4p, a basic leucine zipper protein which regulates amino acid biosynthetic genes and recruits histone acetyltransferases complexes[1,2]. Although the affinity of the transcription factor Gcn4p to different sequences has been studied in multiple papers[3–7], there are still ambiguities regarding the length of the binding motif that it recognizes, the quantitative contribution of each binding site position to the overall affinity, and the exact probability distribution at each position. According to a previous study[6], which performed 57 point mutations on the site and assayed both the *in-vitro* binding of Gcn4p and the *in-vivo* induction level of the downstream gene, the promoters that we designed in this study contain four different Gcn4p sites, as follows:

1. Strong affinity site – ATGACTC**A**T
2. Medium affinity site – ATGACTCTT (the native site of the *HIS3* promoter)
3. Weak affinity site I – ATGACTC**G**T
4. Weak affinity site II – ATGACTCT**A** (created upon deletion of TTTTTTTTCTT, including the downstream poly(dA:dT) tract).

Importantly, our promoter activity measurements are consistent with this ranking of the affinities of the Gcn4p sites.

Pho4p Site. Site used is CACGTGC

Gal4p Site. Site used is CGGAAGACTCTCCTCCG

Poly(dA:dT) Tracts. In this study we use a small number of fixed poly(dA:dT) tracts (see sequences below). Specifically, in our variants:

- The minimal tract is of length 5bp
- *A Perfect* poly(dA:dT) of length ‘k’ is a block of ‘k’ consecutive Thymines

Nature Genetics: doi:10.1038/ng.2305

- *Imperfect* poly(dA:dT) tracts contain only one mismatch block (with 1-2 non Thymines) with at least two Thymines downstream of the mismatch and four Thymines upstream of the mismatch

Some of our promoter variants include one or both of the two poly(dA:dT) tracts that are present in the native *HIS3* promoter, as follows:

- A 17bp imperfect tract TTTTCATTTTTTTTTT upstream of the site
- A 10bp imperfect tract TTTTTTTCTT downstream of the site

In addition, some of our designed promoters contain one of the following tracts:

- A 22bp imperfect tract – TTTTTTTTTCATTTTTTTTTT
- A 12bp perfect tract - TTTTTTTTTTT
- A 5bp perfect tract - TTTTT

See **Supp. Table 1** for further details on the design of each promoter variants.

Sequence of master strain (in genbank format):

```
 LOCUS       MasterStrain              4364 bp    DNA     linear   22-NOV-2009 FEATURES             Location/Qualifiers
primer_bind     1..49
                /note="Chromosome 15 insertion site (forward primer)"
     terminator      complement(63..293)
                /note="ADH1 terminator (reverse orientation)"
     gene            complement(304..1014)
                /note="mCherry gene (reverse orientation)"
     promoter        complement(1015..1587)
                /note="TEF2 promoter (reverse orientation)"
     primer_bind     1604..1653
                /note="Recombination site for URA3-promoter construct (forward)"
     primer_bind     1940..2032
                /note="Recombination site for URA3-promoter construct (reverse)"
     promoter        1940..2040
                /note="HIS3 Proximal promoter"
     gene            2041..2757
                /note="YFP Gene (yEVenus)"
     terminator      2766..2994
                /note="ADH1 terminator"
     promoter        3005..3383
                /note="TEF promoter"
     gene            3384..3956
                /note="Nat1 selection marker"
     terminator      3957..4196
                /note="TEF terminator"
     primer_bind     complement(4325..4364)
```


Nature Genetics: doi:10.1038/ng.2305

/note="Chromosome 15 insertion site (reverse primer)"

BASE COUNT 1116 A 1030 C 1071 G 1147 T

ORIGIN

1 TCTTGGCCTC CTCTAGTACA CTCTATATTT TTTTATGCCT CGGTAATGAT GCCACCTGAC
61 GTCATCTATA TTACCCTGTT ATCCCTAGCG GATCTGCCGG TAGAGGTGTG GTCAATAAGA
121 GCGACCTCAT ACTATACCTG AGAAAGCAAC CTGACCTACA GGAAAGAGTT ACTCAAGAAT
181 AAGAATTTTC GTTTTAAAAC CTAAGAGTCA CTTTAAAATT TGTATACACT TATTTTTTTT
241 ATAACTTATT TAATAATAAA AATCATAAAT CATAAGAAAT TCGCTTATTT AGAAGTCGGT
301 CCGTTACTTG TACAGCTCGT CCATGCCGCC GGTGGAGTGG CGGCCCTCGG CGCGTTCGTA
361 CTGTTCCACG ATGGTGTAGT CCTCGTTGTG GGAGGTGATG TCCAACTTGA TGTTGACGTT
421 GTAGGCGCCG GGCAGCTGCA CGGGCTTCTT GGCCTTGTAG GTGGTCTTGA CCTCAGCGTC
481 GTAGTGGCCG CCGTCCTTCA GCTTCAGCCT CTGCTTGATC TCGCCCTTCA GGGCGCCGTC
541 CTCGGGGTAC ATCCGCTCGG AGGAGGCCTC CCAGCCCATG GTCTTCTTCT GCATTACGGG
601 GCCGTCGGAG GGGAAGTTGG TGCCGCGCAG CTTCACCTTG TAGATGAACT CGCCGTCCTG
661 CAGGGAGGAG TCCTGGGTCA CGGTCACCAC GCCGCCGTCC TCGAAGTTCA TCACGCGCTC
721 CCACTTGAAG CCCTCGGGGA AGGACAGCTT CAAGTAGTCG GGGATGTCGG CGGGGTGCTT
781 CACGTAGGCC TTGGAGCCGT ACATGAACTG AGGGGACAGG ATGTCCCAGG CGAAGGGCAG
841 GGGGCCACCC TTGGTCACCT TCAGCTTGGC GGTCTGGGTG CCCTCGTAGG GGCGGCCCTC
901 GCCCTCGCCC TCGATCTCGA ACTCGTGGCC GTTCACGGAG CCCTCCATGT GCACCTTGAA
961 GCGCATGAAC TCCTTGATGA TGGCCATGTT ATCCTCCTCG CCCTTGCTCA CCATGGTACT
1021 AGTGTTTAGT TAATTATAGT TCGTTGACCG TATATTCTAA AAACAAGTAC TCCTTAAAAA
1081 AAAACCTTGA AGGGAATAAA CAAGTAGAAT AGATAGAGAG AAAAATAGAA AATGCAAGAG
1141 AATTTATATA TTAGAAAGAG AGAAAGAAAA ATGGAAAAAA AAAAATAGGA AAAGCCAGAA
1201 ATAGCACTAG AAGGAGCGAC ACCAGAAAAG AAGGTGATGG AACCAATTTA GCTATATATA
1261 GTTAACTACC GGCTCGATCA TCTCTGCCTC CAGCATAGTC GAAGAAGAAT TTTTTTTTTC
1321 TTGAGGCTTC TGTCAGCAAC TCGTATTTTT TCTTTCTTTT TTGGTGAGCC TAAAAAGTTC
1381 CCACGTTCTC TTGTACGACG CCGTCACAAA CAACCTTATG GGTAATTTGT CGCGGTCTGG
1441 GTGTATAAAT GTGTGGGTGC AACATGAATG TACGGAGGTA GTTTGCTGAT TGGCGGTCTA
1501 TAGATACCTT GGTTATGGCG CCCTCACAGC CGGCAGGGGA AGCGCCTACG CTTGACATCT
1561 ACTATATGTA AGTATACGGC CCCATATATA GGCCCTTTCG TCTCGCGCGT TTCGGTGATG
1621 ACGGTGAAAA CCTCTGACAC ATGCAGCTCC CGGAGACGGT CACAGCTTGT CTGTAAGCGG
1681 ATGCCGGGAG CAGACAAGCC CGTCAGGGCG CGTCAGCGGG TGTTGGCGGG TGTCGGGGCT
1741 GGCTTAACTA TGCGGCATCA GAGCAGATTG TACTGAGAGT GCACCATATG GACATATTGT
1801 CGTTAGAACG CGGCTACAAT TAATACATAA CCTTATGTAT CATACACATA CGATTTAGGT
1861 GACACTATAG AACGCGGCCG CCAGCTGAAG CTTCGTACGC TGCAGGTCGA CGGATCGGTG
1921 ACGGTGCTGG TTTAATTAAG ATTGGCATTA TCACATAATG AATTATACAT TATATAAAGT
1981 AATGTGATTT CTTCGAAGAA TATACTAAAA AATGAGCAGG CAAGATAAAC GAAGGCAAAG
2041 ATGTCTAAAG GTGAAGAATT ATTCACTGGT GTTGTCCCAA TTTTGGTTGA ATTAGATGGT
2101 GATGTTAATG GTCACAAATT TTCTGTCTCC GGTGAAGGTG AAGGTGATGC TACTTACGGT
2161 AAATTGACCT TAAAATTGAT TTGTACTACT GGTAAATTGC CAGTTCCATG GCCAACCTTA
2221 GTCACTACTT TAGGTTATGG TTTGCAATGT TTTGCTAGAT ACCCAGATCA TATGAAACAA
2281 CATGACTTTT TCAAGTCTGC CATGCCAGAA GGTTATGTTC AAGAAAGAAC TATTTTTTTC
2341 AAAGATGACG GTAACTACAA GACCAGAGCT GAAGTCAAGT TTGAAGGTGA TACCTTAGTT
2401 AATAGAATCG AATTAAAAGG TATTGATTTT AAAGAAGATG GTAACATTTT AGGTCACAAA
2461 TTGGAATACA ACTATAACTC TCACAATGTT TACATCACTG CTGACAAACA AAAGAATGGT
2521 ATCAAAGCTA ACTTCAAAAT TAGACACAAC ATTGAAGATG GTGGTGTTCA ATTAGCTGAC
2581 CATTATCAAC AAAATACTCC AATTGGTGAT GGTCCAGTCT TGTTACCAGA CAACCATTAC

Nature Genetics: doi:10.1038/ng.2305

2641 TTATCCTATC AATCTGCCTT ATCCAAAGAT CCAAACGAAA AGAGAGACCA CATGGTCTTG
2701 TTAGAATTTG TTACTGCTGC TGGTATTACC CATGGTATTG ATGAATTGTA CAAATAAGGC
2761 GCGCCACTTC TAAATAAGCG AATTTCTTAT GATTTATGAT TTTTATTATT AAATAAGTTA
2821 TAAAAAAAAT AAGTGTATAC AAATTTTAAA GTGACTCTTA GGTTTTAAAA CGAAAATTCT
2881 TATTCTTGAG TAACTCTTTC CTGTAGGTCA GGTTGCTTTC TCAGGTATAG TATGAGGTCG
2941 CTCTTATTGA CCACACCTCT ACCGGCAGAT CCGCTAGGGA TAACAGGGTA ATATAGATCT
3001 GTTTAGCTTG CCTTGTCCCC GCCGGGTCAC CCGGCCAGCG ACATGGAGGC CCAGAATACC
3061 CTCCTTGACA GTCTTGACGT GCGCAGCTCA GGGGCATGAT GTGACTGTCG CCCGTACATT
3121 TAGCCCATAC ATCCCCATGT ATAATCATTT GCATCCATAC ATTTTGATGG CCGCACGGCG
3181 CGAAGCAAAA ATTACGGCTC CTCGCTGCAG ACCTGCGAGC AGGGAAACGC TCCCCTCACA
3241 GACGCGTTGA ATTGTCCCCA CGCCGCGCCC CTGTAGAGAA ATATAAAAGG TTAGGATTTG
3301 CCACTGAGGT TCTTCTTTCA TATACTTCCT TTTAAAATCT TGCTAGGATA CAGTTCTCAC
3361 ATCACATCCG AACATAAACA ACCATGGGTA CCACTCTTGA CGACACGGCT TACCGGTACC
3421 GCACCAGTGT CCCGGGGGAC GCCGAGGCCA TCGAGGCACT GGATGGGTCC TTCACCACCG
3481 ACACCGTCTT CCGCGTCACC GCCACCGGGG ACGGCTTCAC CCTGCGGGAG GTGCCGGTGG
3541 ACCCGCCCCT GACCAAGGTG TTCCCCGACG ACGAATCGGA CGACGAATCG GACGACGGGG
3601 AGGACGGCGA CCCGGACTCC CGGACGTTCG TCGCGTACGG GGACGACGGC GACCTGGCGG
3661 GCTTCGTGGT CGTCTCGTAC TCCGGCTGGA ACCGCCGGCT GACCGTCGAG GACATCGAGG
3721 TCGCCCCGGA GCACCGGGGG CACGGGGTCG GGCGCGCGTT GATGGGGCTC GCGACGGAGT
3781 TCGCCCGCGA GCGGGGCGCC GGGCACCTCT GGCTGGAGGT CACCAACGTC AACGCACCGG
3841 CGATCCACGC GTACCGGCGG ATGGGGTTCA CCCTCTGCGG CCTGGACACC GCCCTGTACG
3901 ACGGCACCGC CTCGGACGGC GAGCAGGCGC TCTACATGAG CATGCCCTGC CCCTAATCAG
3961 TACTGACAAT AAAAAGATTC TTGTTTTCAA GAACTTGTCA TTTGTATAGT TTTTTTATAT
4021 TGTAGTTGTT CTATTTTAAT CAAATGTTAG CGTGATTTAT ATTTTTTTTC GCCTCGACAT
4081 CATCTGCCCA GATGCGAAGT TAAGTGCGCA GAAAGTAATA TCATGCGTCA ATCGTATGTG
4141 AATGCTGGTC GCTATACTGC TGTCGATTCG ATACTAACGC CGCCATCCAG TGTCGAAAAC
4201 GAGCTCGAAT TCATCGATGA TATCAGATCC ACTAGTGGCC TATGCGGCCG CGGATCTGCC
4261 GGTCTCCCTA TAGTGAGTCG TATTAATTTC GATAAGCCAG GTTAACCTGC ATTAATGAAT
4321 CGGCCCGTAG TGAGAGTGCG TTCAAGGCTC TTGCGGTTGC CATA
//

Nature Genetics: doi:10.1038/ng.2305

Detailed data analysis pipelines.

Plate Reader Measurements

The data analysis is done automatically, with a few steps subjected to manual curation as means of quality control (for a general scheme see **Supp. Fig. 10**).

Raw Data. Experiments are performed in 96-well plates. The data for each experiment consists of OD (optical density, indicative of cell population size), YFP, and mCherry reads for each of the strains, measured every 20 minutes.

Basic Analysis. The analysis pipeline consists of several steps, as follows:

1. Removing outlier strains. Since all of the constructed strains are identical, except for the promoter upstream of the YFP reporter, they are expected to grow similarly. For this reason, strains displaying significantly different growth curves are suspected to have undergone unrelated mutations and were thus excluded from the analysis.
2. Subtracting background levels from measurements. To account for auto-fluorescence of yeast cells, we use a control strain lacking the mCherry gene and a promoter upstream of the YFP gene, and subtract from all YFP and mCherry measurements the corresponding measurements of this strain. Similarly, we subtract the OD reads of a control well containing only medium from the OD measurements of all strains, in order to account for the OD baseline which corresponds to the absorbance properties of the medium and is not indicative of cell population size.
3. Removing outlier measurements for each strain. Due to technical reasons, individual measurements within the set of measured time points for each strain may deviate significantly from the trend portrayed by all other measurements for that strain. Such outliers are automatically removed by comparing the measured value to the two previous and two subsequent measurements, and discarding measurements whose value is much lower or much higher than that measured in the neighboring time points (more than 2 standard deviations from the mean). Note, however, that cases in which such outliers were detected are rare.

Identifying Growth Phases. In order to identify the exponential growth phase, we search for a region in the OD curve that can be well fitted to an exponential function, and that is preceded by a region in the curve that is relatively flat (lag phase).

Nature Genetics: doi:10.1038/ng.2305

Measures of Promoter Activity. The degradation rate of YFP is known to be relatively small (consistent with a monotonically increasing pattern that we observed in all YFP measurements), allowing us to make the simplifying assumption that the difference in YFP level for any two measurements is attributed to the population-mean production of YFP in the corresponding time interval. In addition, since the 5' UTR and 3' UTR for both YFP and mCherry in all of our variants is identical, and thus any existing post-translational regulation on the YFP and mCherry proteins is the same for all strains, we assume that the rates of YFP and mCherry production are proportional to their rates of transcription. This assumption is supported by q-PCR experiments on 8 yeast strains with similar characteristics to the strains in our promoter library[8], that show high correspondence between YFP mRNA levels and YFP protein fluorescence reads (R^2=0.92). Under the above assumptions, given some time interval, we compute the average production rate of either YFP or mCherry per cell per second, by dividing the difference in YFP or mCherry levels in the given time interval by an integral over OD levels in the same interval. Due to the overall similarity in the shape of all of the observed transcriptional responses (**Fig. 1**), we facilitate the comparison between promoter variants and between different experimental conditions, by extracting a single robust measure for each transcriptional response. This measure is calculated as the YFP production rate per cell per second averaged over the entire exponential phase. To account for technical variation between experiments, we use additional measurements of an mCherry reporter driven by a constant promoter. The median mCherry production rate over the exponential phase for a subset of strains that appear in all measurement plates (termed control subset) is not expected to change between different experiments done under the same experimental conditions. Thus, differences in this median value are indicative of technical rather than biological variation. To remove this source of variation from our YFP production measurements, we define a normalization factor per experiment which is the median mCherry production rate of the control subset in that experiment divided by the same median in the reference experiment. We then divide the YFP production rate for each strain in a certain experiment by the corresponding normalization factor. The resulting measure is referred to throughout the manuscript as "promoter activity".

Notably, we define our measure of "promoter activity" by averaging over a relatively large window, the entire exponential phase, enabling us to obtaining a highly robust

Nature Genetics: doi:10.1038/ng.2305

measure. Interestingly, though, when averaging on much smaller time intervals, with a one hour window, we found that the promoter activity values calculated based on the entire exponential phase are actually obtained for all strains at the same one hour window in late exponential phase. This suggests that the values computed over the entire exponential phase can be viewed as representing the YFP production rate for all strains at a specific one hour window.

Flow Cytometry Measurements

The data analysis is done automatically, with a few steps subjected to manual quality control. Some of the processing steps were adapted from a pipeline developed by Gil Hornung of the Barkai Lab.

Raw Data. FCS files are imported to Matlab using an available script written by Laszlo Balkay. The data is then filtered as follows:

1. Removal of outlier wells. Wells in which cells did not grow for some reason (less than 1000 cells counted) or wells with abnormal forward scatter, side scatter or mCherry distributions were discarded.
2. Removal of outlier cells. To avoid various artifacts, cells collected in the first or last 0.5 seconds in each well were discarded from further analysis, as well as cells for which negative or saturated values were measured in one of the parameters. Additionally, to avoid artifacts stemming from flow instability (bubbles, etc.) cells collected in time intervals in which the mean fluorescence levels differ significantly from the mean fluorescence over the entire collection time were also removed.
3. Gating cells based on FSC and SSC values. To reduce the extent to which variability in cell size and shape influence the measured YFP values per cell (and thereby the calculated values of noise and noise strength across the population), we increase the homogeneity of the cell population by gating cells based on their FSC-A and SSC-A values. On the SSC-A vs. FSC-A plane, two populations can be observed. For our analysis we focus on the lower population which is enriched for G1 cells.

For the derivation of statistical measures (mean YFP intensity - μ, variance - σ^2, noise - σ^2/μ^2 and noise strength - σ^2/μ) YFP values are corrected to further account for changes in cell size within the gated population using robust multiple linear regression of YFP vs FSC-A and SSC-A.

Nature Genetics: doi:10.1038/ng.2305

Merging of experiments. To account for technical variation between experiments, we use the corrected mean YFP values for a subset of the strains that were present in all plates (termed control subset). The median over these values is not expected to change between different experiments done under the same experimental conditions, and thus, differences in this median value are indicative of technical rather than biological variation. To remove this source of variation, we normalize this median to be the same in all experiments and equal to the median in some randomly selected reference plate. This is done be normalizing the corrected YFP mean for each strain in each plate by a normalization factor which is defined to be the median of mean corrected YFP values over the control subset in that experiment, divided by the median of mean corrected YFP values over the control subset in some reference plate. The noise (σ^2/μ^2) from different experiments is similarly normalized by the same normalization factor.

Nature Genetics: doi:10.1038/ng.2305

Detailed Model Description

Model parameters

Gcn4p affinity model. As discussed above, our designed promoters contain four Gcn4p sites. For modeling purposes, we use a published position weight matrix[4]. Although this matrix is consistent with the ranking of sites 1-3 above (see **Detailed Description of Promoter Building Blocks)**, it cannot distinguish between Weak affinity site I and Weak affinity site II that differ in the 9th bp, since the matrix is defined as a shorter, 8bp, motif. However, this basepair has been reported as important for binding by several other papers[3,6,7], and since this is also supported by our measurements, we decided to increase the length of the Gcn4p weight matrix to 9bp. The probability distribution for this position was selected naively to give a high (0.5) probability to a Thymine (similar to the reported values[3,7]), and a lower and uniform (0.16666) probability for all other A, C, and G bases. When applying the model to our promoter variants, Gcn4p binding was restricted to its known site and the binding affinity was computed using this position weight matrix.

Histone affinity model. For modeling purposes we use a slightly updated version of the binding model presented in Kaplan et al.[9] (available on demand).

Concentration parameters. To estimate the values representing histone and Gcn4p concentrations, we scanned settings to these parameters between 10^{-5} to 10^{11}, and 10^{-9} to 10^{3}, respectively, with multiplicative jumps of 5.

Estimating the quality of the model predictions

For each given set of input concentrations, our model outputs the predicted binding probability of Gcn4p to its site. We employ the commonly used Pearson R^2 value to estimate the values for our parameters (selecting parameters that maximize this statistic). Notably, Pearson's R^2 allows for linear transformations between the predicted and observed values, which is suitable for our purposes, since the binding probability of Gcn4p to its site is only a proxy for the measured promoter activity values. When applying this framework to the promoter variants in which manipulations to poly(dA:dT) tracts resulted in qualitatively predicted effects on promoter activity, we found that the best R^2 obtained is 0.94 (**Fig. 6b, right panel**), with parameter settings of $10^{-0.5}$ for the histone concentration, and $10^{-4.5}$ for the Gcn4p concentration (see **Supp. Fig. 3** for the heatmap of R^2 values for the scanned parameter settings). This parameter setting was

Nature Genetics: doi:10.1038/ng.2305

also applied to the entire set of promoter variants (**Fig. 6c**), thus avoiding selection of parameters biased by promoter activity values of variants that are not qualitatively understood, and gaining a lower limit assessment of our ability to explain the variability in our measurements.

Despite its significant utility, the Pearson R^2 value has the undesired property of yielding high values for relatively binned datasets (see Anscombe's quartet). Therefore, to better quantify how the incorporation of nucleosomes allows us to "break" the binned structure obtained when predictions are based only on binding site affinity (**Fig. 6b** left graph), we employed two additional statistics: Spearman's rank correlation coefficient, and a normalized mutual information score (defined as the mutual information between the measured and predicted values divided by the entropy of the data). Both the Pearson R^2 value and these two additional measures are specified for each of the graphs in Figure 6. Combined with the visual examination of the graphs presented, these statistics give the reader a better understanding of the extent to which different models explain the data, and demonstrate the improvement that can be gained by incorporating nucleosomes into the model.

Distilling the contribution of poly(dA:dT) tracts to the performance of the model

The model described above employs published sequence preferences of histones[9] that are, of course, reflective of diverse sequence features that influence nucleosome occupancy. In our setting, it is of course interesting to distill the contribution of poly(dA:dT) tracts to the performance of the model. However, this task is not trivial, as it is difficult to decouple the poly(dA:dT) part of the preferences from those of the other sequence features. We have thus employed two different approaches towards this task.

The first approach relies on the fact that the model of Kaplan et al.[9] consists of two components, one capturing the contribution of di-nucleotide periodicity and the other the contribution of different 5-mers to nucleosome occupancy, with the latter component largely representing the disfavoring nature of poly(dA:dT) tracts. We therefore constructed an alternative model that only used this latter component as the histone binding affinity model. When applying this model to our promoter variants (and scanning the parameter space in the same way that we did in the case of the full model, we

Nature Genetics: doi:10.1038/ng.2305

obtained nearly identical results both in terms of the parameters selected (10^{-1} for the histone concentration and $10^{-4.5}$ for the Gcn4p concentration), and the quality of the statistical measures (see **Supp. Table 5**). These results imply that poly(dA:dT) tracts are indeed the main component that increases our ability to explain the data.

Although the above model is dominated by the effect of poly(dA:dT) tracts, it may still capture additional influences that may be represented in the 5-mer statistics. For this reason, we employed another approach based solely on poly(dA:dT) tracts, in which we model the affinity of histones to every 147bp sequence as the sum of Thymines (T) or Adenines (A) contained in poly(dA:dT) tracts identified within that window (length > 5bp, and at most two mismatches). We then multiply this sum by a scaling factor (-0.1) to match the scale of the affinity scores to those obtained with the full affinity model from Kaplan et al.[9], as this allows a comparison of the values predicted by both models for similar Gcn4p and histone concentration parameters. This affinity model also yielded similar results to those of the full model, with identical values selected for the parameters ($10^{-0.5}$ for the histone concentration and $10^{-4.5}$ for the Gcn4p concentration), and with similar statistical measures (see **Supp. Table 5**).

Together, these results demonstrate that poly(dA:dT) tracts are indeed the main determinant of our improved ability to explain the data upon the incorporation of nucleosomes. Nevertheless, we note that since our library of synthetic promoters was specifically designed to elucidate the role of poly(dA:dT) tracts in transcription, the possible transcriptional effects of changes to other nucleosome favoring or disfavoring sequence signals were not tested in our set of variants, and thus, our results do not imply in any way that such signals have negligible transcriptional effects.

Nature Genetics: doi:10.1038/ng.2305

Additional promoter activity measurements

Measurements in synthetic complete medium

Promoter variants were measured also in synthetic complete medium (SCD) instead of under amino acid depletion (see **Supp Fig. 3-6** corresponding to **Fig. 2-4 and Supp. Fig. 8**). Importantly, the qualitative trends reported in the main text, hold in this condition as well.

We note that these conditions likely differ in the amount of active of the Gcn4p, yet it seems that this is not the sole difference between them, and thus the comparison between the promoter activity values measured in the two conditions do not simply reflect changes in the concentration of the regulator. Comparison of mCherry measurements in these two conditions suggests a possible change in the general transcriptional activity in the cell. Though in each condition, the mCherry promoter activity is very similar for all strains (as expected by the fact that in all strains mCherry is driven by the same constant promoter), there is ~1.6 fold increase in mCherry levels measured in SCD media relative to mCherry levels measured under amino acid depletion. While this increase may be specific to the activation of the TEF2 promoter driving the expression of the mCherry, it may also be indicative of some general change in the efficiency of processes like transcription and translation (for instance, due to differences in the amount of cellular resources between these two conditions), leading to a global increase in promoter activity levels. Measurements of promoter activity of ribosomal protein promoters in SCD and amino acid depletion conducted in our lab (unpublished data) are compatible with this possibility.

When attempting to remove such global effects by normalizing the promoter activity values by the change in median mCherry levels between conditions, the expected effect of an increased Gcn4p concentration in amino acid depletion compared to SCD becomes apparent for almost all variants (**Supp. Fig. 2**). The normalized promoter activity of the vast majority of our variants is higher under amino acid depletion, with the only exceptions being either variants that do not contain a Gcn4p site or variants with an extremely low promoter activity in both conditions (in these variants it may be the case that the contribution of Gcn4p to the measured promoter activity values is relatively low, compared to that of the basal transcription machinery). Still, we suspect that even when applying such normalization we might not be accounting properly for the various

Nature Genetics: doi:10.1038/ng.2305

differences between these conditions, and we thus refrain from considering them as differing only in the Gcn4p concentration.

Measurements of promoter variants with a mutated Gcn4p binding site

To verify the role of Gcn4p as the main regulator in most of our promoter variants (except for those containing sites for either Gal4p or Pho4p), we constructed a set of variants corresponding to those of **Fig. 2**, in which we replaced the Gcn4p site with a mutated site whose affinity is predicted to be comparable to that of background sequences[4] (see **Detailed Description of Promoter Building Blocks** above). As expected from the known role of Gcn4p as a regulator of the *HIS3* promoter[10], we found a sharp reduction in the promoter activity measurements for all variants with a mutated Gcn4p site (80% to 93% reduction for the variants in **Supp. Fig. 7a**). Although the promoter activity of these variants is relatively low, a mild increase can be found with longer and more perfect poly(dA:dT) tracts. This might be indicative of the effect that these tracts have not only on nearby binding sites, but also on the accessibility of other promoter elements such as the TATA box. Notably, a comparison of these promoter variants to the variants in which the site was not mutated shows that the effect of the poly(dA:dT) tracts on promoter activity is amplified in the presence of the Gcn4p site, in what can be referred to as a cooperative relation[11,12], whereby the promoter activity of variants with both a strong poly(dA:dT) tract and a non-mutated binding site is significantly higher than the sum of promoter activities of the variants containing only one of these elements (**Supp. Fig. 7b**).

For the two 'extreme' variants in this set, the one lacking the upstream poly(dA:dT) tract and the one with the longest poly(dA:dT) tract, we also measured nucleosome occupancy, focusing on the region surrounding the mutated site (a few additional measurements of more downstream regions showed that this region had the most pronounced change in nucleosome occupancy also among these variants). Similar to our measurements in **Fig. 2d.,** we found that even in the presence of a mutated Gcn4p site, there is a significant reduction in nucleosome occupancy over the Gcn4p site upon the introduction of the poly(dA:dT) tract **(Supp. Fig. 7c**). This suggests that binding of Gcn4p is not required for the reduction of nucleosome occupancy observed upon the

Nature Genetics: doi:10.1038/ng.2305

introduction of poly(dA:dT) tracts, consistent with conclusions derived from genome-wide analyses[13].

Thus, these measurements support the role of Gcn4p as the main regulator in our promoters. In addition, it demonstrates that although an intact biding site for the factor is not required to induce the nucleosome occupancy change conferred by the poly(dA:dT) tract, the presence of such a site allows for an amplification in the transcriptional effect of this tract.

Nature Genetics: doi:10.1038/ng.2305

References

1. Knutson, B.A. & Hahn, S. Domains of Tra1 important for activator recruitment and transcription coactivator functions of SAGA and NuA4 complexes. *Molecular and cellular biology* **31**, 818-31 (2011).

2. Kuo, M.H., vom Baur, E., Struhl, K. & Allis, C.D. Gcn4 activator targets Gcn5 histone acetyltransferase to specific promoters independently of transcription. *Molecular cell* **6**, 1309-20 (2000).

3. Zhu, C. *et al.* High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome research* **19**, 556-66 (2009).

4. MacIsaac, K.D. *et al.* An improved map of conserved regulatory sites for Saccharomyces cerevisiae. *BMC bioinformatics* **7**, 113 (2006).

5. Harbison, C.T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99-104 (2004).

6. Hill, D.E., Hope, I.A., Macke, J.P. & Struhl, K. Saturation mutagenesis of the yeast his3 regulatory site: requirements for transcriptional induction and for binding by GCN4 activator protein. *Science (New York, N.Y.)* **234**, 451-7 (1986).

7. Fordyce, P.M. *et al.* De novo identification and biophysical characterization of transcription-factor binding sites with microfluidic affinity analysis. *Nature biotechnology* **28**, 970-5 (2010).

8. Zeevi, D. *et al.* Compensation for differences in gene copy number among yeast ribosomal proteins is encoded within their promoters. *Genome research* **21**, 2114-28 (2011).

9. Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362-6 (2009).

10. Hope, I.A. & Struhl, K. GCN4 protein, synthesized in vitro, binds HIS3 regulatory sequences: implications for general control of amino acid biosynthetic genes in yeast. *Cell* **43**, 177-88 (1985).

11. Raveh-Sadka, T., Levo, M. & Segal, E. Incorporating nucleosomes into thermodynamic models of transcription regulation. *Genome research* **19**, 1480-96 (2009).

12. Miller, J.A. & Widom, J. Collaborative competition mechanism for gene activation in vivo. *Molecular and cellular biology* **23**, 1623-32 (2003).

13. Field, Y. *et al.* Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS computational biology* **4**, e1000216 (2008).

Nature Genetics: doi:10.1038/ng.2305