

Biophysical Journal, Volume 112

Supplemental Information

**Specificity-Determining DNA Triplet Code for Positioning of Human
Preinitiation Complex**

Matan Goldshtein and David B. Lukatsky

Supporting Material

Supporting Figures

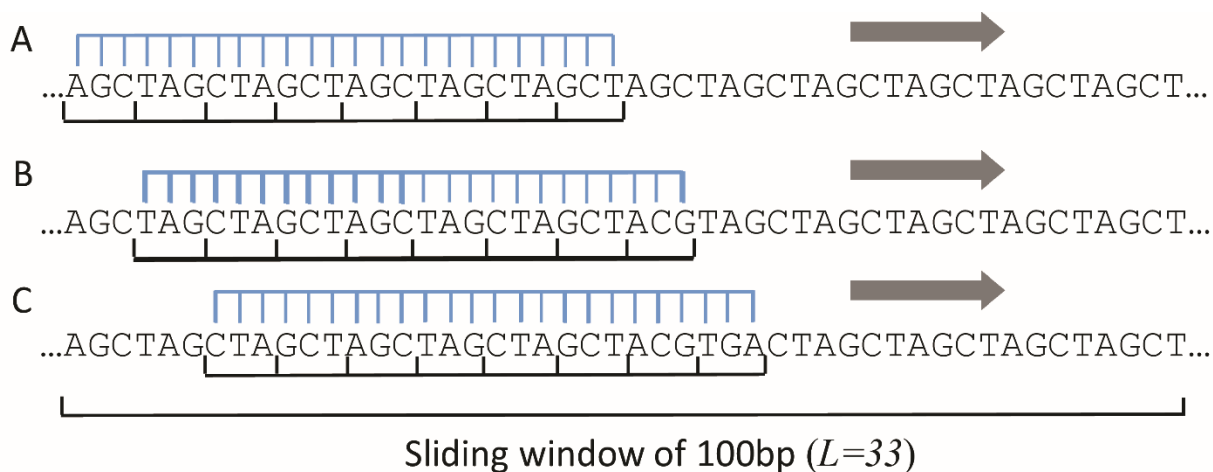


FIGURE S1 Cartoon illustrating the calculation of the nonconsensus protein-DNA binding energies, U , as a model random binder slides along the sliding window. The interaction contacts of a model protein TF with all DNA nucleotide bases are depicted in blue. The corresponding nucleotide triplets are depicted in black below the DNA strand. In our model we used TF that forms 24 contacts with nucleotide bases (blue), which corresponds to $M=8$ contacts with nucleotide triplets (black). Each model TF slides (gray arrow) along the DNA sequence by 3 bp steps. We used the sliding window with the width 100 bp, which corresponds to $L=33$ nucleotide triplets. The following three examples illustrate the energy calculation as TF slides three consecutive steps along the sliding window: (A) $U(1)=2K_{AGC}+2K_{TAG}+2K_{CTA}+2K_{GCT}$; (B) $U(2)=2K_{TAG}+2K_{CTA}+2K_{GCT}+K_{AGC}+K_{ACG}$; (C) $U(3)=2K_{CTA}+2K_{GCT}+K_{AGC}+K_{ACG}+K_{TAG}+K_{TGA}$. The 64 random energy parameters K_{α} are drawn from the Gaussian distribution with the zero mean and the standard deviation $\sigma=2k_{\text{B}}T$. These parameters uniquely define a given random binder. In all our calculations we used the free energy averaged over 250 random binders. Therefore, for each DNA sliding window, the procedure described above was repeated for all 250 random binders, each characterized by a different set of K_{α} .

A

First reading frame

...AGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT...

Second reading frame

...AGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT...

Third reading frame

...AGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT...

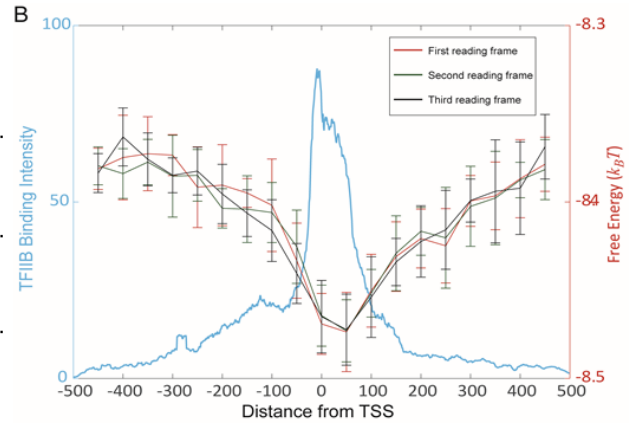


FIGURE S2 Robustness of the nonconsensus protein-DNA binding free energy landscape computed for different DNA reading frames. This figure is complementary to Fig. 1 of the main text, and all the definitions and the axes labels are identical to those defined in Fig. 1. (A) Three possible DNA reading frames for a sliding random binder are illustrated. (B) The average free energy of nonconsensus TFIIB-DNA binding for all three possible DNA reading frames, and the measured profile of average TFIIB occupancy around the TSSs of 6097 genes. For each reading frame, the average free energy was calculated every 50 bp, within the interval (-450 bp; 450 bp). The rest of the parameters are identical to those defined in Fig. 1 of the main text.

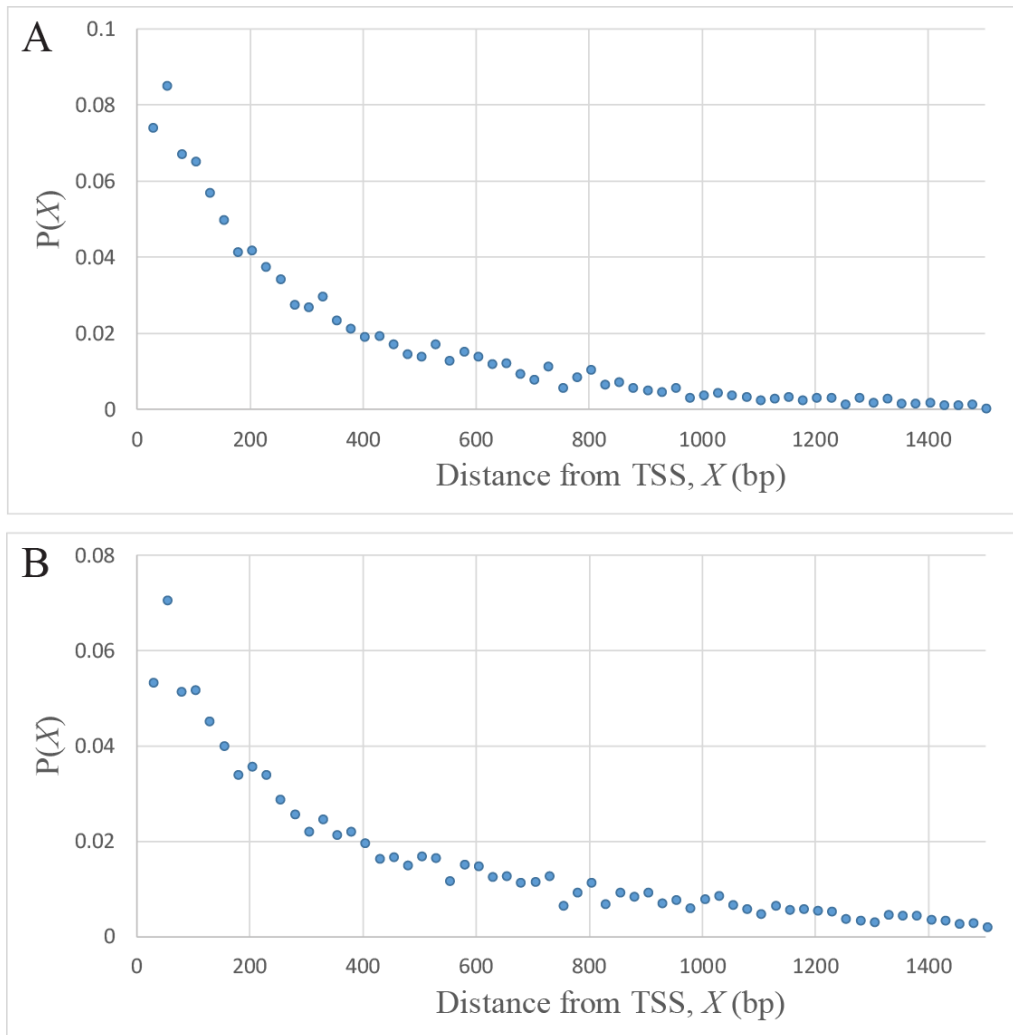


FIGURE S3 Probability distribution of the Kozak translation start site consensus sequence as a function of the distance X after TSS for 6097 genes used in our analysis. The Kozak sequence is the consensus sequence recognized by ribosomes to begin translation, gccRccATGG, where R represents a purine (i.e. A or G), and lower-case letters represent a lower significance parts of the consensus motif (M. Kozak, Nucl. Acid Res. 15(20), 8185 (1987)). (A) Probability distribution for the consensus sequence ATGG: 29% of the data falls within the first 100bp after the transcription start site (TSS); (B) Probability distribution for the consensus sequence RnnATGG: 22% of the data falls within the first 100bp after TSS.

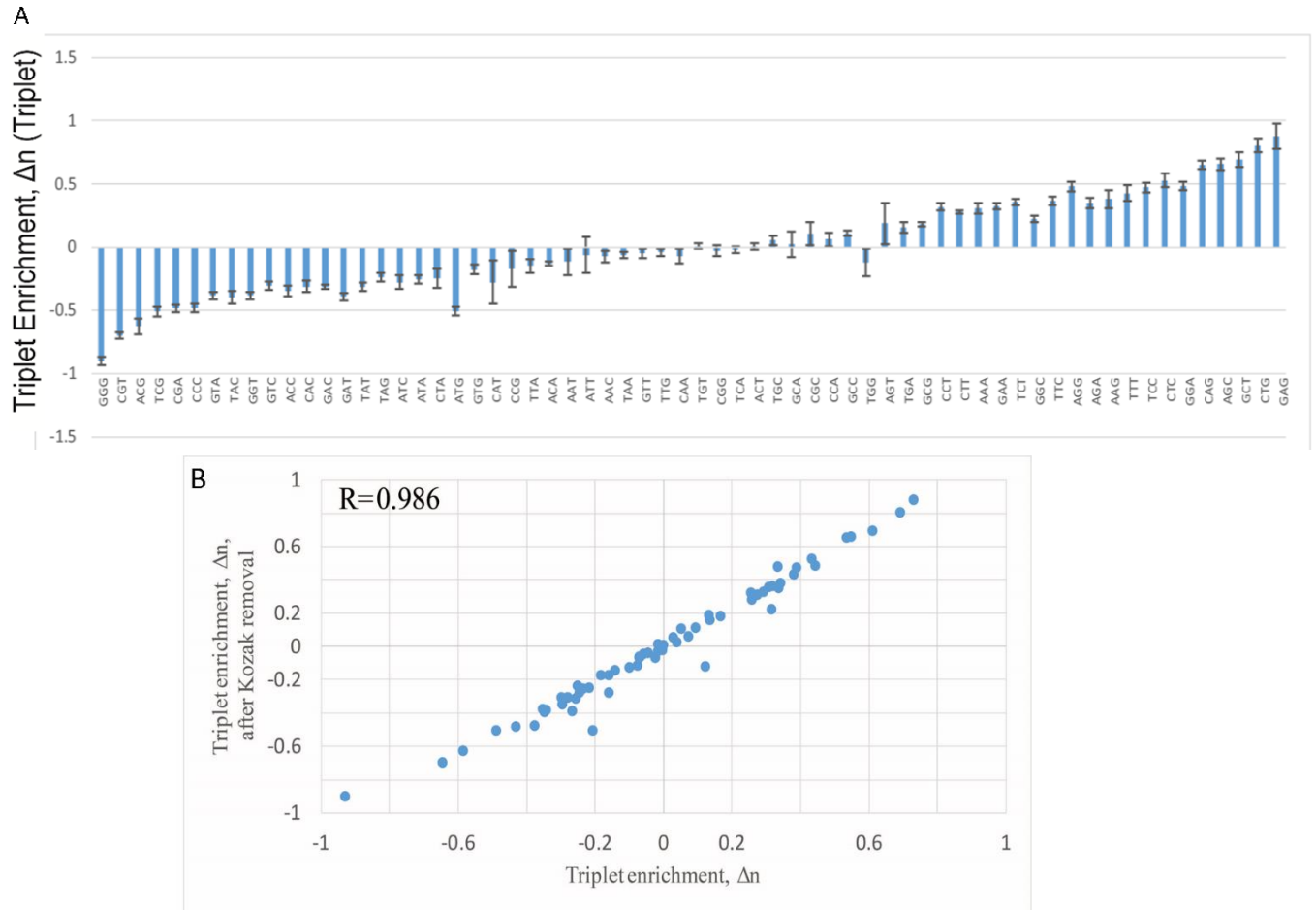


FIGURE S4 Enrichment levels of 64 nucleotide triplets computed for the genomic regions characterized by high TFIIB binding intensity, after removal of sequences with Kozak translation start site consensus sequence. (A) This figure is complementary to Fig. 2A of the main text. After removing DNA sequences containing the Kozak translation start site (see Fig. S3A), we are left with 4360 sequences (out of 6097 sequences used in Fig. 2A of the main text). The definition of the triplet enrichment Δn is identical to the one used in Fig. 2 of the main text. (B) The plot containing 64 points, representing the linear correlation between the triplet enrichment Δn with and without removing the Kozak sequences, respectively.

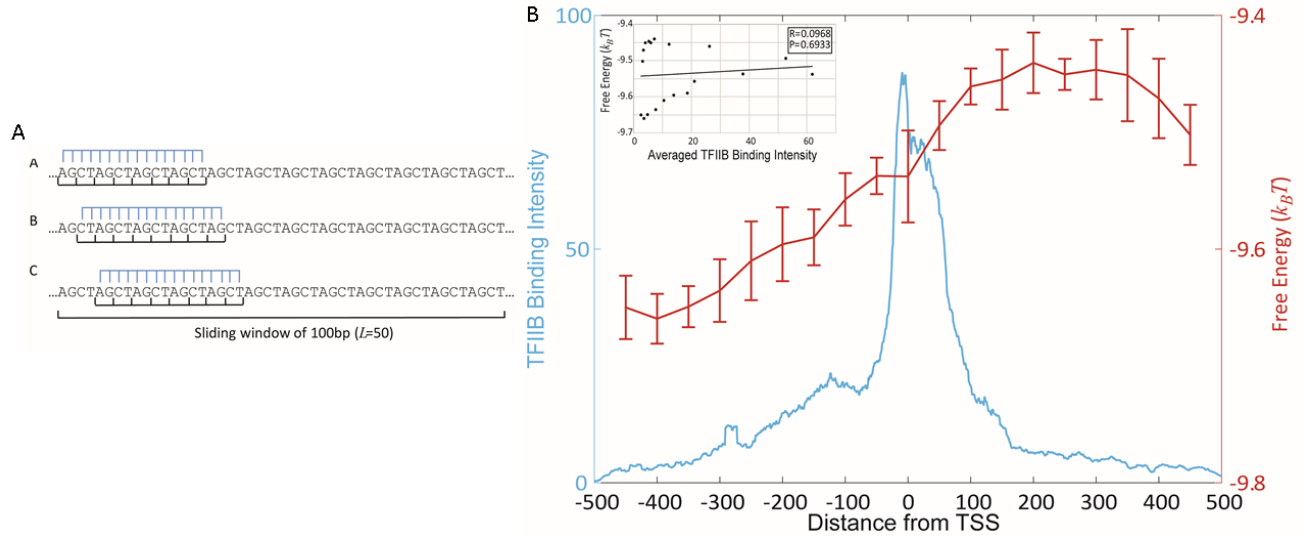


FIGURE S5 Free energy of nonconsensus TFIIB-DNA binding based on doublets does not correlate with the TFIIB binding intensity. The computed average free energy of nonconsensus TFIIB-DNA binding (using the model random binding interaction potential with all 16 possible nucleotide doublets) and the profile of the average TFIIB binding intensity measured around the TSSs of 6097 genes (see Fig. 1 of the main text). In order to compute the free energy, we used Eqs. (1-3), as described in the main text, with the index α running over 16 possible nucleotide doublets (instead of 64 possible triplets used in the main text in order to compute the free energy). (A) Cartoon illustrating the calculation of the nonconsensus protein-DNA binding energies, U , as a model random binder slides along the sliding window for doublets. For example, the model protein-DNA binding energy for the first, top position, $U(1)=4K_{AG}+4K_{CT}$. Here all 16 possible energy parameters (for each model TF), K_α , are drawn from the Gaussian distribution with the zero mean and the standard deviation, $\sigma=2k_B T$. (B) The average free energy was calculated every 50 bp, within the interval (-450 bp; 450 bp). In order to compute the free energy, we used a sliding window of 100 bp. To compute error bars, we calculated the mean free energy for each chromosome and divided the results into five randomly chosen subgroups and computed the mean for each subgroup. The error bars are defined as one standard deviation of mean free energy between the subgroups. (Inset) The correlation between the free energy and the TFIIB binding intensity with the Pearson correlation coefficient and the p -value.

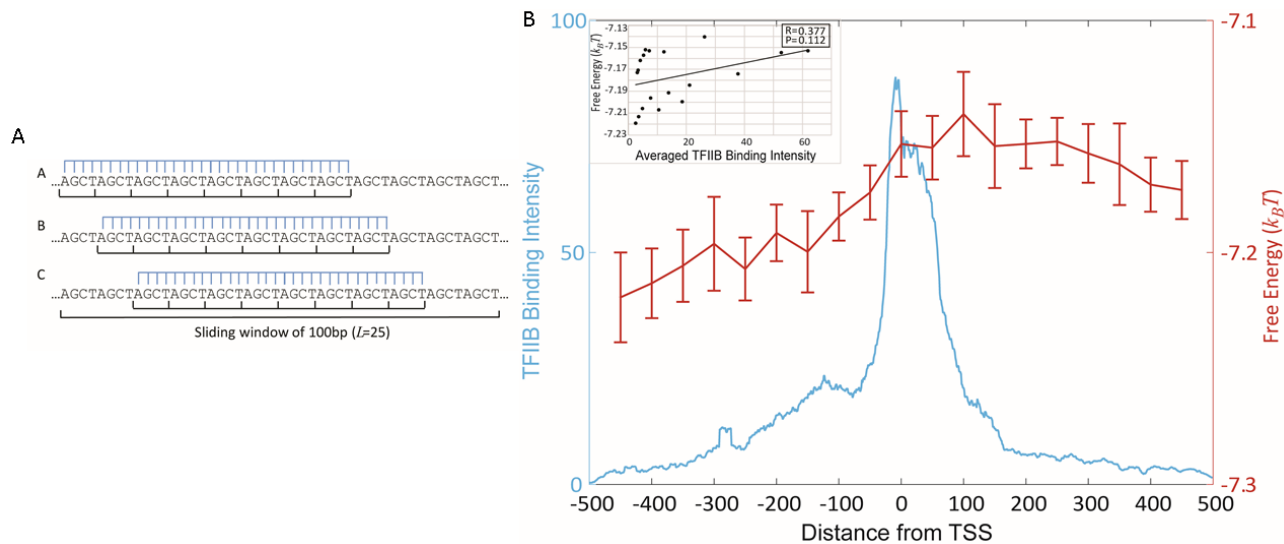


FIGURE S6 Free energy of nonconsensus TFIIB-DNA binding based on quadruplets does not correlate with the TFIIB binding intensity. The computed average free energy of nonconsensus TFIIB-DNA binding (using the model random binding interaction potential with all 256 possible nucleotide quadruplets) and the profile of the average TFIIB binding intensity measured around 6097 genes (see Fig. 1 of the main text). In order to compute the free energy, we used Eqs. (1-3), as described in the main text, with the index α running over 256 possible nucleotide quadruplets (instead of 64 possible triplets used in the main text). (A) Cartoon illustrating the calculation of the nonconsensus protein-DNA binding energies, U , as a model random binder slides along the sliding window for doublets. For example, the model protein-DNA binding energy for the first, top position, $U(1)=8K_{AGCT}$. Here all 256 possible energy parameters (for each model TF), K_α , are drawn from the Gaussian distribution with the zero mean and the standard deviation, $\sigma=2k_B T$. (B) The average free energy was calculated every 50 bp, within the interval (-450 bp; 450 bp). In order to compute the free energy, we used a sliding window of 100 bp. To compute error bars, we calculated the mean free energy for each chromosome and divided the results into five randomly chosen subgroups and computed the mean for each subgroup. The error bars are defined as one standard deviation of mean free energy between the subgroups. (Inset) The correlation between the free energy and the TFIIB binding intensity with the Pearson correlation coefficient and the p -value.

Supporting Tables

	Mean [0;100]	Mean [0;100] Rand	KS-test	P-Value	Δ (Mean [0;100]-Mean [0;100] rand)	Mean [-450;-350]	Mean [-450;-350] Rand	KS-test	P-Value	Δ (Mean [-450;-350]-Mean[-450;-350] rand)
AAA	0.818	0.547	1	3.194E-44	0.272	2.607	1.689	1	1.06E-66	0.918
AAC	0.670	0.743	1	5.95E-11	-0.072	1.258	1.417	1	1.46E-14	-0.159
AAG	1.292	0.951	1	2.048E-65	0.341	1.744	1.476	1	1.55E-30	0.268
AAT	0.439	0.516	1	3.259E-12	-0.077	1.357	1.431	1	0.01854	-0.074
ACA	0.642	0.743	1	3.675E-23	-0.102	1.469	1.423	0	0.387044	0.046
ACC	1.062	1.358	1	3.673E-57	-0.296	1.379	1.610	1	7.52E-30	-0.231

ACG	0.903	1.490	1	1.66E-268	-0.588	0.768	1.515	1	0	-0.747
ACT	0.822	0.825	0	0.4344062	-0.003	1.325	1.315	0	0.291719	0.010
AGA	1.287	0.952	1	1.194E-64	0.334	1.789	1.478	1	7.42E-36	0.312
AGC	2.165	1.618	1	2.61E-123	0.547	1.793	1.524	1	2.01E-32	0.269
AGG	2.311	1.979	1	9.101E-42	0.333	2.260	1.626	1	3.3E-122	0.634
AGT	1.100	0.970	1	2.385E-12	0.131	1.252	1.281	0	0.502634	-0.029
ATA	0.258	0.494	1	9.72E-125	-0.236	0.860	1.433	1	4.4E-193	-0.574
ATC	0.557	0.803	1	1.631E-97	-0.246	0.994	1.314	1	9.88E-59	-0.320
ATG	0.728	0.936	1	1.745E-39	-0.207	1.021	1.287	1	1.49E-57	-0.266
ATT	0.516	0.590	1	6.57E-14	-0.073	1.332	1.378	0	0.231387	-0.047
CAA	0.730	0.755	1	4.105E-05	-0.026	1.470	1.427	0	0.118211	0.044
CAC	1.074	1.356	1	1.165E-59	-0.282	1.676	1.611	0	0.059437	0.065
CAG	2.155	1.623	1	8.87E-117	0.532	2.284	1.518	1	2.1E-226	0.766
CAT	0.651	0.813	1	2.143E-36	-0.162	1.024	1.314	1	1.38E-61	-0.290
CCA	1.473	1.401	1	0.0024857	0.071	1.995	1.616	1	3.09E-62	0.379
CCC	2.778	3.157	1	1.899E-12	-0.378	2.719	2.497	1	3.48E-21	0.222
CCG	3.131	3.293	1	5.144E-16	-0.162	1.618	2.138	1	1.1E-96	-0.520
CCT	1.976	1.722	1	1.337E-24	0.254	2.176	1.622	1	7.19E-91	0.554
CGA	1.078	1.510	1	1.27E-139	-0.432	0.827	1.513	1	5.1E-285	-0.687
CGC	3.362	3.311	1	3.612E-09	0.051	1.641	2.139	1	1.68E-91	-0.498
CGG	3.790	3.808	1	5.945E-11	-0.018	1.558	2.070	1	3.3E-101	-0.512
CGT	1.088	1.736	1	9.85E-294	-0.647	0.708	1.445	1	0	-0.737
CTA	0.580	0.798	1	4.591E-72	-0.218	0.992	1.309	1	2.15E-59	-0.318
CTC	2.177	1.746	1	2.384E-59	0.431	2.207	1.622	1	6.22E-94	0.585
CTG	2.579	1.890	1	1.34E-192	0.689	2.149	1.453	1	5E-199	0.697
CTT	1.331	1.074	1	2.199E-35	0.257	1.612	1.374	1	5.86E-29	0.238
GAA	1.230	0.941	1	1.228E-54	0.289	1.671	1.473	1	2.72E-17	0.197
GAC	1.253	1.534	1	1.396E-51	-0.281	1.177	1.516	1	4.38E-71	-0.340
GAG	2.750	2.021	1	9.79E-139	0.729	2.154	1.632	1	1.14E-58	0.522
GAT	0.659	0.927	1	1.163E-84	-0.268	0.943	1.279	1	7.86E-70	-0.336
GCA	1.602	1.564	0	0.6949699	0.037	1.579	1.514	0	0.11215	0.064
GCC	3.419	3.327	1	0.0223591	0.092	2.345	2.142	1	5.91E-14	0.203
GCG	4.002	3.836	1	4.071E-11	0.166	1.560	2.078	1	2E-117	-0.518
GCT	2.477	1.868	1	1.51E-146	0.609	1.731	1.449	1	1.67E-32	0.283
GGA	2.427	1.987	1	1.482E-63	0.440	1.976	1.618	1	9.82E-56	0.358
GGC	4.173	3.860	1	7.107E-15	0.314	2.307	2.079	1	1.86E-13	0.228
GGG	3.781	4.712	1	2.321E-57	-0.931	2.646	2.293	1	1.25E-35	0.352
GGT	1.729	2.073	1	5.782E-64	-0.344	1.325	1.446	1	9.42E-09	-0.122
GTA	0.561	0.916	1	1.87E-163	-0.354	0.789	1.281	1	8.8E-171	-0.492
GTC	1.465	1.766	1	5.475E-64	-0.301	1.172	1.439	1	6.5E-44	-0.267
GTG	1.912	2.097	1	5.428E-22	-0.186	1.454	1.442	0	0.143382	0.012
GTT	1.093	1.153	1	0.0061886	-0.060	1.132	1.284	1	2.49E-15	-0.152
TAA	0.443	0.514	1	4.087E-09	-0.070	1.209	1.423	1	1.38E-21	-0.214
TAC	0.436	0.786	1	7.4E-203	-0.350	0.834	1.318	1	1.4E-142	-0.484

TAG	0.670	0.924	1	1.065E-76	-0.254	0.911	1.284	1	4.71E-92	-0.373
TAT	0.309	0.567	1	6.44E-127	-0.258	0.876	1.386	1	5.8E-150	-0.509
TCA	0.830	0.834	0	0.2012681	-0.004	1.409	1.315	1	1.28E-05	0.094
TCC	2.117	1.731	1	8.48E-56	0.386	2.077	1.624	1	1.74E-71	0.453
TCG	1.254	1.743	1	1.08E-164	-0.490	0.795	1.438	1	5E-259	-0.644
TCT	1.402	1.094	1	2.304E-45	0.307	1.726	1.372	1	2.02E-50	0.354
TGA	1.106	0.974	1	2.102E-13	0.133	1.355	1.290	1	0.007369	0.065
TGC	1.836	1.809	0	0.2064999	0.027	1.475	1.446	0	0.184563	0.029
TGG	2.255	2.134	1	4.946E-08	0.121	1.795	1.448	1	4.98E-51	0.347
TGT	1.138	1.156	1	0.0225222	-0.018	1.262	1.274	1	0.004641	-0.012
TTA	0.438	0.580	1	1.268E-36	-0.143	1.187	1.387	1	1.09E-15	-0.200
TTC	1.403	1.087	1	4.787E-59	0.316	1.636	1.374	1	8.9E-32	0.262
TTG	1.104	1.151	1	0.000423	-0.047	1.256	1.276	0	0.848099	-0.020
TTT	1.203	0.824	1	4.301E-54	0.378	2.344	1.496	1	6.85E-71	0.848

TABLE S1 This table is complementary to Fig. 2 of the main text. It provides the two-sample Kolmogorov–Smirnov *p*-values for the statistical significance of the enrichment levels for all 64 nucleotide triplets. Triplets colored in yellow did not show significant enrichment or depletion at [0;100], triplets colored in blue did not show significant enrichment or depletion at [-450;-350], triplets colored in green did not show significant enrichment or depletion at both [0;100] and [-450;-350].

	Mean [0;100]	Mean [0;100] Rand	KS-test	P-Value	Δ (Mean [0;100]-Mean [0;100] rand)	Mean [-450;-350]	Mean [-450;-350] Rand	KS-test	P-Value	Δ (Mean [-450;-350]-Mean[-450;-350] rand)
AA	3.254	2.703	1	1.72E-35	0.551	6.796	5.728	1	1.19E-38	1.068
AC	3.465	4.501	1	8.25E-197	-1.04	4.945	5.825	1	8.82E-125	-0.88
AG	6.934	5.332	1	5.40E-260	1.602	7.18	5.815	1	1.69E-189	1.365
AT	2.079	2.886	1	2.58E-128	-0.81	4.045	5.13	1	5.94E-108	-1.084
CA	4.657	4.484	1	8.91E-05	0.173	6.462	5.833	1	1.24E-49	0.629
CC	9.46	9.476	0	0.904168	-0.02	8.984	8.125	1	5.79E-31	0.859
CG	9.408	10.35	1	3.33E-55	-0.94	5.106	7.374	1	0	-2.268
CT	6.731	5.332	1	1.81E-189	1.4	7.004	5.669	1	5.97E-185	1.335
GA	5.958	5.335	1	1.61E-40	0.622	5.993	5.816	1	0.040686	0.177
GC	11.61	10.34	1	1.07E-92	1.27	7.576	7.383	0	0.090044	0.193
GG	12.23	12.53	1	0.00061	-0.3	8.52	7.594	1	5.66E-40	0.925
GT	5.087	5.97	1	6.45E-102	-0.88	4.58	5.343	1	2.09E-100	-0.763
TA	1.877	2.9	1	8.01E-208	-1.02	3.703	5.117	1	1.09E-186	-1.414
TC	5.655	5.325	1	3.19E-10	0.33	6.075	5.688	1	5.34E-15	0.388
TG	6.402	5.96	1	1.95E-19	0.442	5.875	5.35	1	9.20E-36	0.525
TT	4.19	3.569	1	4.36E-28	0.621	6.157	5.212	1	5.49E-33	0.944

TABLE S2 This table is complementary to Fig. S5 of Supplemental Information. It provides the two-sample Kolmogorov–Smirnov *p*-values for the statistical significance of the enrichment levels for all 16 nucleotide doublets. Doubles colored in yellow did not show significant enrichment or depletion at [0;100], doublets colored in blue did not show significant enrichment or depletion at [-450;-350], doublets colored in green did not show significant enrichment or depletion at both [0;100] and [-450;-350]. In particular, in the interval [0;100], 15 out of 16 doublets showed a significant enrichment/depletion. In the interval [-450; -350], 15 out of 16 doublets showed a significant enrichment/depletion.

	Mean [0;100]	Mean [0;100] Rand	KS-test	P-Value	Δ (Mean [0;100]-Mean [0;100] rand)	Mean [-450;-350]	Mean [-450;-350] Rand	KS-test	P-Value	Δ (Mean [-450;-350]-Mean[-450;-350] rand)
AAAA	0.221	0.119	1	2.32E-16	0.102	0.987	0.473	1	4.37E-55	0.514
AAAC	0.173	0.143	1	0.0004303	0.030	0.419	0.359	1	6.36E-06	0.060
AAAG	0.281	0.183	1	2.43E-34	0.098	0.559	0.381	1	6.25E-52	0.178
AAAT	0.134	0.111	1	0.0010087	0.023	0.5	0.373	1	3.38E-19	0.127
AACA	0.146	0.142	0	1	0.004	0.367	0.358	0	0.985359	0.009
AACC	0.203	0.226	0	0.3874321	-0.023	0.348	0.367	0	0.361213	-0.019
AACG	0.154	0.261	1	9.20E-32	-0.107	0.183	0.348	1	8.77E-78	-0.165
AACT	0.16	0.153	0	0.2207591	0.008	0.333	0.311	0	0.079623	0.021
AAGA	0.292	0.18	1	2.83E-40	0.112	0.452	0.377	1	8.77E-13	0.075
AAGC	0.351	0.254	1	1.21E-26	0.097	0.408	0.352	1	1.10E-08	0.055
AAGG	0.408	0.327	1	1.28E-14	0.081	0.492	0.385	1	7.30E-21	0.106
AAGT	0.228	0.177	1	3.38E-12	0.050	0.349	0.313	1	0.009745	0.036
AATA	0.075	0.109	1	0.0028256	-0.035	0.311	0.372	1	2.41E-10	-0.061
AATC	0.121	0.143	0	0.0680461	-0.023	0.281	0.315	1	7.23E-05	-0.034
AATG	0.143	0.176	1	0.0426714	-0.033	0.313	0.312	0	0.765907	0.000
AATT	0.097	0.128	0	0.2032548	-0.031	0.355	0.34	0	0.530452	0.015
ACAA	0.127	0.15	0	0.3516819	-0.023	0.342	0.355	0	0.219633	-0.012
ACAC	0.153	0.224	1	9.48E-21	-0.071	0.337	0.363	1	4.96E-10	-0.026
ACAG	0.258	0.265	0	1	-0.008	0.463	0.35	1	1.03E-32	0.113
ACAT	0.098	0.15	1	3.33E-07	-0.051	0.242	0.31	1	5.13E-15	-0.068
ACCA	0.185	0.223	1	0.0002353	-0.038	0.334	0.369	1	0.000256	-0.035
ACCC	0.34	0.433	1	6.80E-21	-0.094	0.456	0.495	1	0.002679	-0.039
ACCG	0.292	0.478	1	9.18E-82	-0.186	0.243	0.435	1	2.06E-92	-0.192
ACCT	0.232	0.248	0	0.6791332	-0.016	0.368	0.342	1	0.021324	0.026
ACGA	0.121	0.26	1	2.89E-57	-0.139	0.152	0.354	1	2.59E-113	-0.202
ACGC	0.32	0.473	1	5.71E-63	-0.153	0.248	0.433	1	3.39E-90	-0.184
ACGG	0.295	0.555	1	8.88E-143	-0.259	0.222	0.424	1	5.63E-103	-0.202
ACGT	0.158	0.275	1	1.18E-41	-0.117	0.155	0.318	1	8.22E-84	-0.164
ACTA	0.089	0.147	1	1.30E-10	-0.058	0.202	0.311	1	2.28E-31	-0.109

ACTC	0.235	0.245	0	0.3580176	-0.011	0.382	0.341	1	8.22E-06	0.041
ACTG	0.298	0.267	0	0.1819288	0.031	0.381	0.32	1	2.33E-09	0.061
ACTT	0.191	0.171	1	0.0107202	0.020	0.319	0.306	0	0.296602	0.013
AGAA	0.29	0.18	1	2.67E-34	0.110	0.518	0.381	1	3.97E-29	0.137
AGAC	0.262	0.254	0	0.9948072	0.009	0.357	0.355	0	0.792859	0.002
AGAG	0.531	0.342	1	3.08E-64	0.189	0.591	0.379	1	4.68E-64	0.212
AGAT	0.188	0.186	1	0.0035506	0.002	0.28	0.307	1	0.035946	-0.028
AGCA	0.343	0.259	1	9.92E-16	0.084	0.415	0.351	1	3.03E-07	0.064
AGCC	0.656	0.467	1	3.90E-45	0.190	0.614	0.436	1	4.41E-50	0.177
AGCG	0.669	0.553	1	5.64E-10	0.116	0.341	0.431	1	1.10E-25	-0.090
AGCT	0.476	0.268	1	3.77E-92	0.208	0.448	0.313	1	5.55E-43	0.135
AGGA	0.527	0.338	1	5.25E-58	0.189	0.577	0.382	1	3.49E-66	0.195
AGGC	0.723	0.547	1	8.60E-33	0.176	0.648	0.433	1	1.53E-72	0.216
AGGG	0.681	0.736	0	0.1305816	-0.055	0.645	0.485	1	1.06E-26	0.160
AGGT	0.355	0.328	0	0.3108584	0.027	0.371	0.326	1	8.33E-07	0.045
AGTA	0.118	0.168	1	1.26E-07	-0.050	0.233	0.313	1	2.32E-16	-0.080
AGTC	0.316	0.28	1	5.83E-06	0.036	0.302	0.313	0	0.846948	-0.012
AGTG	0.397	0.345	1	8.16E-09	0.052	0.381	0.322	1	1.95E-06	0.059
AGTT	0.256	0.197	1	1.04E-12	0.059	0.312	0.29	0	0.257892	0.022
ATAA	0.081	0.11	0	0.1218812	-0.029	0.281	0.373	1	2.80E-20	-0.092
ATAC	0.048	0.149	1	1.03E-32	-0.101	0.16	0.311	1	1.35E-71	-0.150
ATAG	0.067	0.176	1	1.44E-34	-0.109	0.159	0.309	1	6.77E-65	-0.150
ATAT	0.058	0.118	1	1.01E-08	-0.060	0.212	0.34	1	1.26E-42	-0.128
ATCA	0.096	0.149	1	1.16E-08	-0.053	0.227	0.311	1	1.48E-20	-0.084
ATCC	0.191	0.238	1	2.13E-11	-0.047	0.314	0.342	0	0.28965	-0.028
ATCG	0.114	0.275	1	2.14E-83	-0.161	0.131	0.312	1	4.29E-108	-0.181
ATCT	0.15	0.168	0	0.2358064	-0.018	0.282	0.305	0	0.112337	-0.023
ATGA	0.118	0.173	1	3.73E-08	-0.055	0.24	0.313	1	2.16E-15	-0.074
ATGC	0.153	0.279	1	1.30E-46	-0.127	0.22	0.313	1	1.96E-26	-0.093
ATGG	0.329	0.322	0	0.9905053	0.006	0.272	0.317	1	6.28E-06	-0.045
ATGT	0.123	0.196	1	3.85E-14	-0.074	0.23	0.29	1	1.13E-13	-0.059
ATTA	0.066	0.117	1	1.40E-08	-0.051	0.278	0.336	1	2.35E-05	-0.057
ATTC	0.13	0.178	1	9.28E-06	-0.048	0.278	0.304	1	0.024503	-0.026
ATTG	0.145	0.188	1	1.52E-06	-0.043	0.198	0.293	1	9.73E-29	-0.094
ATTT	0.17	0.142	1	3.40E-05	0.027	0.482	0.345	1	6.38E-25	0.137
CAAA	0.167	0.145	1	0.0108053	0.023	0.491	0.358	1	3.00E-30	0.133
CAAC	0.17	0.228	1	5.68E-08	-0.058	0.286	0.363	1	2.39E-15	-0.077
CAAG	0.279	0.257	0	0.2952025	0.023	0.412	0.352	1	1.28E-08	0.060
CAAT	0.106	0.155	1	2.83E-05	-0.049	0.226	0.311	1	1.56E-22	-0.085
CACA	0.198	0.237	1	0.0001909	-0.038	0.429	0.362	1	0.046583	0.067
CACC	0.357	0.439	1	3.05E-16	-0.082	0.522	0.489	0	0.109726	0.033

CACG	0.255	0.464	1	8.85E-119	-0.209	0.272	0.436	1	4.60E-68	-0.164
CACT	0.255	0.239	0	0.9999999	0.016	0.408	0.343	1	7.45E-09	0.065
CAGA	0.378	0.265	1	2.80E-36	0.113	0.508	0.352	1	1.28E-44	0.157
CAGC	0.773	0.46	1	3.91E-99	0.312	0.671	0.436	1	2.06E-74	0.235
CAGG	0.647	0.553	1	8.66E-07	0.094	0.739	0.434	1	3.29E-145	0.306
CAGT	0.336	0.287	1	1.13E-09	0.048	0.366	0.313	1	1.35E-07	0.053
CATA	0.058	0.146	1	6.68E-26	-0.088	0.17	0.311	1	9.09E-65	-0.142
CATC	0.184	0.246	1	3.29E-14	-0.062	0.281	0.337	1	3.63E-10	-0.056
CATG	0.24	0.276	1	0.0086397	-0.035	0.237	0.313	1	1.02E-20	-0.076
CATT	0.163	0.168	0	1	-0.006	0.309	0.304	0	1	0.005
CCAA	0.229	0.215	0	0.9974366	0.014	0.438	0.368	1	1.76E-08	0.069
CCAC	0.359	0.44	1	6.23E-19	-0.081	0.56	0.492	1	4.57E-08	0.069
CCAG	0.631	0.461	1	2.75E-33	0.170	0.738	0.437	1	1.72E-127	0.301
CCAT	0.237	0.239	0	0.8971656	-0.002	0.282	0.342	1	4.49E-09	-0.060
CCCA	0.458	0.431	0	0.7678608	0.026	0.691	0.494	1	4.91E-55	0.197
CCCC	0.804	1.097	1	4.34E-18	-0.294	0.855	0.912	1	0.046896	-0.058
CCCG	0.922	1.035	1	5.00E-24	-0.112	0.681	0.724	1	1.36E-05	-0.043
CCCT	0.562	0.532	0	0.9997722	0.029	0.625	0.507	1	1.60E-18	0.117
CCGA	0.365	0.464	1	4.46E-27	-0.099	0.268	0.437	1	5.57E-66	-0.168
CCGC	1.272	1.053	1	1.66E-08	0.218	0.688	0.726	1	1.88E-16	-0.038
CCGG	1.149	1.115	0	0.5744381	0.034	0.563	0.657	1	1.09E-12	-0.094
CCGT	0.311	0.558	1	2.69E-120	-0.247	0.223	0.42	1	9.11E-95	-0.197
CCTA	0.157	0.244	1	1.84E-24	-0.087	0.278	0.346	1	1.28E-11	-0.068
CCTC	0.742	0.55	1	6.96E-24	0.191	0.783	0.5	1	5.24E-73	0.282
CCTG	0.681	0.53	1	4.93E-22	0.151	0.669	0.422	1	8.70E-108	0.247
CCTT	0.372	0.326	1	0.0005468	0.047	0.458	0.363	1	3.77E-12	0.095
CGAA	0.172	0.265	1	3.66E-21	-0.093	0.199	0.353	1	1.54E-67	-0.154
CGAC	0.258	0.471	1	9.18E-108	-0.213	0.196	0.437	1	8.90E-147	-0.242
CGAG	0.525	0.555	1	5.47E-05	-0.029	0.333	0.432	1	7.06E-30	-0.098
CGAT	0.11	0.267	1	4.32E-87	-0.157	0.114	0.314	1	2.74E-132	-0.201
CGCA	0.455	0.486	1	0.0005802	-0.031	0.308	0.434	1	5.52E-48	-0.127
CGCC	1.142	1.048	1	0.0076493	0.094	0.692	0.722	1	2.10E-06	-0.030
CGCG	1.098	1.104	1	1.59E-19	-0.006	0.451	0.655	1	6.07E-72	-0.205
CGCT	0.638	0.537	1	3.54E-06	0.101	0.326	0.421	1	4.87E-24	-0.094
CGGA	0.652	0.554	1	3.91E-09	0.098	0.321	0.428	1	2.12E-30	-0.107
CGGC	1.425	1.117	1	2.04E-18	0.308	0.527	0.657	1	7.79E-39	-0.129
CGGG	1.205	1.358	1	1.19E-22	-0.153	0.596	0.671	1	4.35E-10	-0.075
CGGT	0.47	0.61	1	1.20E-50	-0.140	0.202	0.393	1	4.06E-98	-0.191
CGTA	0.108	0.27	1	7.41E-88	-0.162	0.088	0.316	1	1.31E-173	-0.227
CGTC	0.371	0.551	1	1.12E-70	-0.180	0.234	0.425	1	1.66E-89	-0.191
CGTG	0.387	0.613	1	1.71E-117	-0.226	0.259	0.398	1	1.08E-48	-0.138

CGTT	0.211	0.333	1	9.28E-44	-0.123	0.155	0.319	1	1.12E-80	-0.163
CTAA	0.107	0.144	1	2.93E-05	-0.037	0.28	0.316	1	0.001088	-0.037
CTAC	0.161	0.253	1	2.34E-23	-0.092	0.245	0.341	1	9.69E-24	-0.096
CTAG	0.22	0.269	1	1.45E-08	-0.049	0.259	0.317	1	1.62E-09	-0.058
CTAT	0.085	0.173	1	7.33E-24	-0.087	0.171	0.306	1	4.13E-57	-0.135
CTCA	0.287	0.245	1	0.0001026	0.042	0.479	0.34	1	6.58E-38	0.140
CTCC	0.853	0.539	1	1.31E-74	0.314	0.852	0.498	1	1.52E-125	0.354
CTCG	0.503	0.53	1	7.25E-08	-0.027	0.344	0.422	1	1.59E-19	-0.078
CTCT	0.513	0.33	1	1.60E-59	0.183	0.548	0.359	1	5.59E-59	0.189
CTGA	0.397	0.268	1	2.98E-36	0.129	0.438	0.312	1	2.55E-36	0.126
CTGC	0.897	0.529	1	6.68E-104	0.368	0.603	0.423	1	1.31E-47	0.181
CTGG	0.829	0.604	1	3.23E-40	0.225	0.695	0.399	1	1.78E-141	0.296
CTGT	0.427	0.328	1	3.40E-13	0.099	0.393	0.315	1	1.26E-11	0.078
CTTA	0.116	0.17	1	4.88E-07	-0.054	0.231	0.308	1	1.56E-15	-0.077
CTTC	0.543	0.322	1	1.37E-85	0.221	0.489	0.359	1	6.01E-27	0.129
CTTG	0.307	0.338	1	0.014088	-0.031	0.358	0.317	1	0.000418	0.041
CTTT	0.351	0.229	1	1.59E-36	0.122	0.477	0.334	1	1.47E-37	0.143
GAAA	0.288	0.185	1	4.07E-37	0.103	0.552	0.383	1	9.92E-43	0.169
GAAC	0.229	0.265	0	0.0667583	-0.036	0.314	0.349	1	0.040134	-0.035
GAAG	0.568	0.331	1	1.67E-111	0.237	0.489	0.383	1	1.39E-19	0.106
GAAT	0.131	0.172	1	1.41E-05	-0.041	0.277	0.309	1	0.004406	-0.032
GACA	0.209	0.263	1	4.03E-07	-0.054	0.311	0.346	1	0.008365	-0.035
GACC	0.359	0.474	1	1.36E-29	-0.116	0.343	0.44	1	6.14E-22	-0.097
GACG	0.382	0.558	1	3.39E-57	-0.175	0.228	0.431	1	2.58E-100	-0.202
GACT	0.288	0.266	0	0.9109212	0.022	0.31	0.318	0	0.295203	-0.008
GAGA	0.477	0.34	1	1.85E-34	0.137	0.557	0.381	1	3.74E-42	0.176
GAGC	0.83	0.555	1	3.11E-62	0.276	0.508	0.431	1	4.01E-08	0.077
GAGG	1.013	0.736	1	4.92E-51	0.277	0.758	0.484	1	5.95E-70	0.274
GAGT	0.401	0.336	1	3.24E-10	0.065	0.327	0.32	0	0.999997	0.007
GATA	0.072	0.175	1	1.51E-31	-0.103	0.153	0.309	1	6.62E-71	-0.156
GATC	0.17	0.27	1	6.07E-34	-0.100	0.248	0.328	1	1.23E-20	-0.080
GATG	0.258	0.331	1	4.58E-14	-0.073	0.246	0.319	1	7.05E-18	-0.073
GATT	0.153	0.189	1	0.0006216	-0.036	0.258	0.291	1	0.002102	-0.032
GCAA	0.232	0.272	0	0.0836628	-0.040	0.333	0.351	0	0.264289	-0.018
GCAC	0.315	0.466	1	9.53E-58	-0.151	0.372	0.434	1	2.49E-10	-0.063
GCAG	0.87	0.564	1	1.74E-93	0.306	0.645	0.432	1	6.02E-62	0.213
GCAT	0.167	0.275	1	1.96E-37	-0.108	0.228	0.311	1	8.11E-24	-0.083
GCCA	0.518	0.465	1	2.67E-05	0.054	0.514	0.436	1	1.54E-13	0.078
GCCC	0.938	1.018	1	7.28E-22	-0.080	0.803	0.723	1	3.81E-05	0.080
GCCG	1.287	1.149	1	2.35E-05	0.139	0.5	0.658	1	3.07E-46	-0.158
GCCT	0.64	0.551	1	2.55E-08	0.090	0.626	0.42	1	2.46E-65	0.206

GCGA	0.455	0.55	1	3.59E-24	-0.095	0.278	0.433	1	1.90E-67	-0.155
GCGC	1.258	1.107	1	5.72E-05	0.151	0.573	0.652	1	5.03E-15	-0.079
GCGG	1.821	1.356	1	7.17E-36	0.465	0.588	0.673	1	2.63E-26	-0.085
GCGT	0.433	0.601	1	1.21E-76	-0.168	0.23	0.4	1	7.29E-77	-0.170
GCTA	0.197	0.275	1	5.66E-18	-0.077	0.235	0.318	1	6.70E-23	-0.083
GCTC	0.721	0.516	1	7.06E-32	0.205	0.511	0.416	1	1.01E-13	0.095
GCTG	1.104	0.595	1	2.50E-150	0.509	0.615	0.395	1	4.38E-78	0.220
GCTT	0.434	0.335	1	1.22E-22	0.099	0.37	0.318	1	1.30E-06	0.053
GGAA	0.546	0.339	1	5.79E-87	0.207	0.545	0.381	1	2.42E-51	0.164
GGAC	0.474	0.56	1	1.05E-12	-0.086	0.384	0.428	1	4.94E-05	-0.044
GGAG	1.156	0.727	1	2.03E-117	0.429	0.767	0.483	1	1.03E-75	0.285
GGAT	0.225	0.324	1	1.03E-28	-0.099	0.298	0.318	0	0.387432	-0.020
GGCA	0.496	0.555	1	5.68E-08	-0.059	0.464	0.43	1	0.004637	0.035
GGCC	1.118	1.133	1	0.0042573	-0.014	0.71	0.656	1	0.000292	0.055
GGCG	1.645	1.349	1	1.06E-11	0.295	0.615	0.673	1	3.35E-10	-0.058
GGCT	0.87	0.595	1	2.56E-65	0.275	0.574	0.394	1	8.19E-56	0.180
GGGA	0.807	0.727	1	2.72E-12	0.081	0.704	0.479	1	2.00E-76	0.225
GGGC	1.24	1.349	1	1.61E-17	-0.109	0.744	0.669	1	1.62E-07	0.075
GGGG	1.148	1.801	1	2.86E-68	-0.654	0.8	0.785	1	1.18E-07	0.015
GGGT	0.548	0.784	1	1.25E-62	-0.236	0.434	0.427	0	1	0.007
GGTA	0.208	0.343	1	2.64E-40	-0.134	0.189	0.326	1	1.06E-54	-0.137
GGTC	0.449	0.605	1	6.23E-56	-0.156	0.36	0.4	1	0.000185	-0.041
GGTG	0.694	0.763	1	1.54E-07	-0.069	0.447	0.423	0	0.19287	0.024
GGTT	0.361	0.398	1	0.0185169	-0.037	0.318	0.31	0	0.598857	0.008
GTAA	0.147	0.172	0	0.3036719	-0.024	0.225	0.314	1	3.57E-21	-0.089
GTAC	0.11	0.272	1	1.04E-88	-0.162	0.141	0.318	1	2.92E-102	-0.177
GTAG	0.228	0.337	1	6.87E-27	-0.109	0.223	0.324	1	1.86E-26	-0.101
GTAT	0.071	0.19	1	4.65E-45	-0.119	0.161	0.289	1	2.51E-47	-0.127
GTCA	0.245	0.281	1	0.0002985	-0.036	0.268	0.314	1	7.68E-08	-0.046
GTCC	0.469	0.535	1	6.93E-15	-0.066	0.39	0.427	1	0.004482	-0.037
GTCG	0.378	0.612	1	1.29E-117	-0.234	0.169	0.397	1	2.37E-138	-0.229
GTCT	0.359	0.337	0	0.3857596	0.022	0.354	0.314	1	3.13E-05	0.040
GTGA	0.398	0.339	1	3.69E-10	0.059	0.358	0.324	1	0.003059	0.034
GTGC	0.456	0.605	1	2.25E-56	-0.148	0.342	0.399	1	4.57E-08	-0.057
GTGG	0.701	0.775	1	2.71E-07	-0.074	0.454	0.421	0	0.214065	0.033
GTGT	0.335	0.397	1	1.56E-11	-0.062	0.299	0.31	1	4.72E-06	-0.010
GTTA	0.124	0.196	1	4.00E-14	-0.071	0.187	0.291	1	7.51E-34	-0.104
GTTC	0.313	0.346	0	0.1486678	-0.033	0.28	0.319	1	0.005446	-0.039
GTTG	0.337	0.393	1	9.94E-09	-0.056	0.249	0.309	1	4.85E-09	-0.059
GTTT	0.31	0.248	1	6.93E-15	0.062	0.39	0.306	1	9.72E-15	0.084
TAAA	0.134	0.107	1	0.0003774	0.027	0.436	0.371	1	0.000134	0.065

TAAC	0.093	0.145	1	6.64E-10	-0.052	0.21	0.314	1	1.13E-31	-0.103
TAAG	0.15	0.172	0	0.3791144	-0.022	0.242	0.312	1	8.02E-14	-0.070
TAAT	0.062	0.121	1	1.78E-08	-0.059	0.255	0.346	1	1.23E-14	-0.091
TACA	0.083	0.149	1	1.79E-13	-0.066	0.275	0.313	1	4.94E-05	-0.038
TACC	0.134	0.244	1	4.38E-41	-0.110	0.185	0.346	1	6.65E-75	-0.161
TACG	0.103	0.278	1	8.84E-98	-0.174	0.096	0.313	1	1.22E-153	-0.217
TACT	0.111	0.174	1	1.53E-09	-0.063	0.227	0.305	1	1.49E-15	-0.077
TAGA	0.125	0.179	1	3.36E-05	-0.054	0.227	0.313	1	7.98E-19	-0.086
TAGC	0.192	0.274	1	1.05E-18	-0.082	0.229	0.313	1	1.23E-19	-0.084
TAGG	0.223	0.337	1	6.85E-32	-0.114	0.253	0.322	1	2.98E-13	-0.069
TAGT	0.123	0.189	1	1.73E-12	-0.066	0.185	0.292	1	1.80E-29	-0.107
TATA	0.051	0.123	1	3.97E-12	-0.071	0.177	0.34	1	1.75E-62	-0.164
TATC	0.078	0.171	1	3.04E-28	-0.093	0.157	0.304	1	1.94E-63	-0.147
TATG	0.081	0.192	1	2.06E-38	-0.111	0.168	0.29	1	8.52E-48	-0.122
TATT	0.096	0.139	1	2.52E-05	-0.044	0.316	0.343	1	0.007712	-0.027
TCAA	0.123	0.148	1	0.0255913	-0.026	0.299	0.311	0	0.940762	-0.012
TCAC	0.22	0.251	1	0.0015397	-0.031	0.362	0.342	0	0.476476	0.020
TCAG	0.354	0.276	1	8.59E-16	0.079	0.445	0.314	1	6.17E-38	0.131
TCAT	0.126	0.173	1	2.65E-06	-0.047	0.246	0.303	1	5.68E-11	-0.057
TCCA	0.297	0.235	1	1.85E-06	0.062	0.479	0.34	1	6.99E-41	0.138
TCCC	0.675	0.549	1	6.01E-14	0.126	0.743	0.507	1	8.75E-77	0.237
TCCG	0.608	0.531	1	0.0037732	0.077	0.32	0.421	1	1.12E-28	-0.101
TCCT	0.517	0.319	1	1.25E-58	0.198	0.571	0.361	1	9.18E-78	0.210
TCGA	0.123	0.283	1	2.94E-75	-0.160	0.143	0.312	1	3.94E-90	-0.170
TCGC	0.469	0.546	1	6.60E-18	-0.077	0.271	0.422	1	1.42E-57	-0.151
TCGG	0.483	0.607	1	2.00E-35	-0.124	0.272	0.395	1	2.25E-36	-0.123
TCGT	0.165	0.334	1	2.06E-80	-0.169	0.129	0.319	1	4.53E-108	-0.189
TCTA	0.132	0.174	1	0.0001156	-0.041	0.241	0.305	1	5.59E-11	-0.064
TCTC	0.463	0.332	1	1.95E-30	0.130	0.546	0.362	1	1.26E-59	0.185
TCTG	0.477	0.339	1	1.23E-38	0.138	0.462	0.313	1	8.20E-47	0.150
TCTT	0.318	0.227	1	3.75E-21	0.091	0.409	0.333	1	2.06E-14	0.076
TGAA	0.208	0.168	1	0.0001207	0.040	0.371	0.309	1	3.05E-09	0.062
TGAC	0.245	0.276	1	0.0058242	-0.031	0.262	0.315	1	3.09E-09	-0.052
TGAG	0.513	0.341	1	1.85E-69	0.172	0.462	0.323	1	2.20E-38	0.139
TGAT	0.128	0.187	1	4.58E-14	-0.059	0.2	0.292	1	2.15E-24	-0.092
TGCA	0.297	0.276	0	0.1071655	0.020	0.394	0.314	1	7.11E-09	0.080
TGCC	0.481	0.537	1	1.98E-08	-0.056	0.433	0.424	0	0.857022	0.009
TGCG	0.562	0.611	1	1.24E-16	-0.050	0.262	0.397	1	8.41E-57	-0.135
TGCT	0.478	0.321	1	1.46E-26	0.157	0.383	0.317	1	7.33E-09	0.065
TGGA	0.419	0.331	1	1.63E-16	0.088	0.393	0.321	1	5.48E-09	0.072
TGGC	0.755	0.613	1	1.48E-17	0.142	0.443	0.395	1	0.00061	0.047

TGGG	0.717	0.764	1	5.98E-05	-0.048	0.637	0.42	1	6.99E-81	0.217
TGGT	0.344	0.384	1	1.17E-05	-0.039	0.308	0.313	0	0.981663	-0.006
TGTA	0.123	0.192	1	1.28E-14	-0.069	0.237	0.289	1	1.14E-07	-0.052
TGTC	0.318	0.327	0	0.064245	-0.009	0.285	0.314	1	8.57E-04	-0.028
TGTG	0.424	0.397	0	0.5423292	0.027	0.367	0.311	0	0.155412	0.057
TGTT	0.258	0.253	0	0.2910308	0.005	0.319	0.305	0	0.999772	0.014
TTAA	0.097	0.118	0	0.2924167	-0.021	0.355	0.339	0	0.850337	0.016
TTAC	0.112	0.171	1	9.30E-10	-0.059	0.235	0.308	1	8.66E-14	-0.073
TTAG	0.139	0.193	1	5.08E-07	-0.054	0.25	0.289	1	0.000541	-0.040
TTAT	0.085	0.146	1	2.57E-07	-0.060	0.275	0.343	1	1.08E-11	-0.069
TTCA	0.192	0.171	0	0.0720371	0.021	0.376	0.305	1	9.29E-11	0.071
TTCC	0.584	0.321	1	5.96E-127	0.263	0.552	0.363	1	7.51E-58	0.190
TTCG	0.25	0.353	1	1.51E-21	-0.103	0.171	0.316	1	5.25E-63	-0.145
TTCT	0.364	0.236	1	1.74E-50	0.129	0.477	0.333	1	4.58E-42	0.144
TTGA	0.178	0.192	0	0.7541037	-0.014	0.259	0.29	1	0.01464	-0.031
TTGC	0.306	0.333	1	0.0036129	-0.027	0.307	0.317	0	0.351682	-0.010
TTGG	0.373	0.388	0	0.2655821	-0.015	0.36	0.313	1	8.07E-05	0.047
TTGT	0.236	0.247	0	1	-0.012	0.284	0.305	0	0.065911	-0.022
TTTA	0.128	0.147	0	0.9905053	-0.019	0.413	0.345	1	3.33E-07	0.068
TTTC	0.399	0.232	1	5.11E-67	0.168	0.533	0.335	1	1.12E-72	0.198
TTTG	0.306	0.242	1	1.93E-12	0.064	0.403	0.305	1	1.04E-17	0.098
TTTT	0.357	0.198	1	2.61E-22	0.159	0.784	0.393	1	9.00E-61	0.392

TABLE S3 This table is complementary to Fig. S6 of Supplemental Information. It provides the two-sample Kolmogorov–Smirnov p -values for the statistical significance of the enrichment levels for all 256 nucleotide quadruplets. Quadruplets colored in yellow did not show significant enrichment or depletion at [0;100], quadruplets colored in blue did not show significant enrichment or depletion at [-450;-350], quadruplets colored in green did not show significant enrichment or depletion at both [0;100] and [-450;-350]. In particular, in the interval [0;100], 215 out of 256 quadruplets showed a significant enrichment/depletion. In the interval [-450; -350], 225 out of 256 quadruplets showed a significant enrichment/depletion.