Supplementary Material (ESI) for Molecular BioSystems

SUPPLEMENTARY INFORMATION

for

Genome-wide Analysis Predicts DNA Structural Motifs as Nucleosome Exclusion Signals

Kangkan Halder[a$], Rashi Halder[b$], and Shantanu Chowdhury[a,b,*]

[$] these authors contributed equally

[a] Proteomics and Structural Biology Unit, [b] G.N.R. Knowledge Centre for Genome Informatics,
Institute of Genomics and Integrative Biology, CSIR, Mall Road, Delhi 110 007, India

*Corresponding author: shantanuc@igib.res.in
Fax: 91 11 2766 7471

Supplementary Table ST1

Distribution of PG4 motifs and control sequences in S. cerevisiae and human genomes

Whole Genome	Region Length (bp)	no .of PG4 motifs	motif per kb	no .of control motif	Control motif per kb
S. cerevisiae (PG4 stem size n=2)	12,070,821	5,308	0.44	14,894	1.23
S. cerevisiae (PG4 stem size n=3)	12,070,821	39	0.003	129	0.01
Human (PG4 stem size n=2)*	257,060,000	1,430,026	5.56	739,956	2.88
Human (PG4 stem size n=3)*	257,060,000	89,455	0.35	25135	0.098
Human (PG4 stem size n=3)#	~3,000,000,000	~376,000	0.13	-	-

* ±5kb region of 25,706 gene transcription start sites
\# whole genome (*Nucleic Acids Res.*, 2005, **33**, 2901-2907, *Nucleic Acids Res.*, 2005, **33**, 2908-2916)

Supplementary Table ST2

PG4 distribution across different NOP regions within *S. cerevisiae*

Nucleosome Occupancy Probability (NOP)[a]	Length of the region (bp)	no .of PG4 motifs	motif per kb	no .of control sequences	Control sequences per kb
0.0 - 0.2	1,936,414[b]	1,122	0.58	2,506	1.29
0.2 - 0.4	8,027,448[c]	2,650	0.33	7,533	0.94
0.4 - 0.6	2,826,252[c]	878	0.31	2,609	0.92
0.6 - 0.8	1,128,192[c]	341	0.30	1,063	0.94
0.8 - 1.0	249,288[c]	85	0.34	223	0.89
Whole Genome[d]	12,070,821	5,308[e]	0.44	14,894	1.23
Nucleosome / non-nucleosome regions as reported in Lee *et al.* [f]	Length (Mb)	No. of PG4 motifs (observed)	PG4 motifs/kb (observed)	PG4 motifs (expected)	p value
Nucleosome regions	4.97	1772	0.35	2187	< 6.03E-18
Fuzzy or delocalized nucleosomes	4.8	2305	0.48	2112	< 0.003
Linker or non-nucleosomal regions	2.29	1281	0.56	1010	<5.39 E-14

[a] the nucleosome occupancy probability (NOP) data has been detailed elsewhere (*Nature*, 2006, **442**, 772-778).

[b] all sequences ≥150 bases were used for determining the number of PG4 motif and control sequences.

[c] all the regions were 156 bp long (*Nature*, 2006, **442**, 772-778).

[d] we used refseq from NCBI as 15 aug 2005

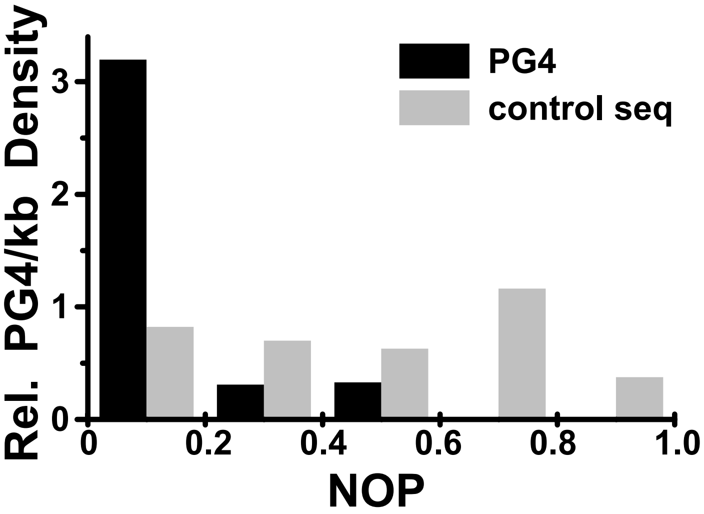
[e] Out of 5308 PG4 motifs (whole genome) we could map 5076 PG4 motifs within various NOP regions. This was due to two reasons: some of the PG4 motifs are expected between two defined NOP regions and stretches below 150 bp were not considered (for ≤ 0.2 NOP).

[f] *Nat. Genet.*, 2007, **39**, 1235-1244

Supplementary Figure S1

Distribution of PG4 motif (red) and NOP (≥ 0.5, blue) (Segal et al., 2006) in 50 bp window across all chromosomes for *S. cerevisiae* (see pdf).

Supplementary Figure S2



Relative motif density (per kb) for PG4 with stem size three (n=3, see methods) and its corresponding control sequence distribution across different NOP regions. No PG4 motif was detected for NOP 0.6 -0.8 and 0.8-1.0.

Supplementary Figure S3

Distribution of PG4 motif (red) and nucleosome occupancy (blue) plotted in a 50 bp window across 3239 human promoters (1.5 kb (-1250 to +250 bp of TSS)) across seven human cell lines (see pdf).

Supplementary Table ST3

PG4 motif density in non-nucleosomal regions[a] within ENCODE [b]

	ENCODE	Non-nucleosomal[a]	Rest of ENCODE
Region Length (bp)	29,955,159	476,804	29,478,355
PG4 (no. of motif)	6,005	170	5,835
PG4 (density, per kb)	0.20	0.36	0.20

a *Genome Res.*, 2007, **17**, 877-885
b *Science*, 2004, **306**, 636-640

Supplementary Table ST4

Biological Process for genes which have ≤ 1 nucleosome and ≥ 3 PG4 motifs in their promoter (Set I) and which have ≥ 3 nucleosomes and ≤ 1 PG4 motif in there promoter (Set II)

Set I		
GO_ID	**TERM**	**P_VALUE** (after Bonferroni correction)
GO:0006468	protein amino acid phosphorylation	2.25E-06
GO:0016310	Phosphorylation	1.42E-05
GO:0006796	phosphate metabolism	1.43E-05
GO:0006793	phosphorus metabolism	1.43E-05
GO:0007242	intracellular signaling cascade	4.27E-05
GO:0008219	cell death	4.45E-05
GO:0016265	Death	5.03E-05
GO:0006366	transcription from RNA polymerase II promoter	5.23E-05
GO:0006915	Apoptosis	2.68E-04
GO:0012501	programmed cell death	2.92E-04
GO:0007399	nervous system development	1.21E-03
GO:0048731	System development	1.33E-03
GO:0007166	cell surface receptor linked signal transduction	2.05E-03
GO:0042981	regulation of apoptosis	2.26E-03
GO:0043067	regulation of programmed cell death	2.51E-03
GO:0030154	cell differentiation	3.15E-03
GO:0006464	protein modification	9.70E-03

Set II

GO_ID	TERM	P_VALUE (after Bonferroni correction)
GO:0050896	response to stimulus	1.32E-09
GO:0006950	response to stress	1.03E-06
GO:0006935	Chemotaxis	2.17E-06
GO:0042330	Taxis	2.17E-06
GO:0009628	response to abiotic stimulus	4.29E-06
GO:0009605	response to external stimulus	4.53E-06
GO:0007155	cell adhesion	1.05E-05
GO:0050874	organismal physiological process	1.75E-05
GO:0042221	response to chemical stimulus	2.13E-05
GO:0048522	Positive regulation of cellular process	6.85E-05
GO:0043207	response to external biotic stimulus	9.53E-05
GO:0045859	regulation of protein kinase activity	1.93E-04
GO:0007156	homophilic cell adhesion	2.13E-04
GO:0051338	regulation of transferase activity	2.13E-04
GO:0009613	response to pest, pathogen or parasite	2.18E-04
GO:0051242	Positive regulation of cellular physiological process	2.89E-04
GO:0042127	regulation of cell proliferation	3.38E-04
GO:0043119	Positive regulation of physiological process	5.49E-04
GO:0001525	Angiogenesis	6.10E-04
GO:0050877	neurophysiological process	6.62E-04
GO:0048514	blood vessel morphogenesis	8.42E-04
GO:0001568	blood vessel development	8.42E-04
GO:0001944	vasculature development	8.42E-04
GO:0007600	sensory perception	8.80E-04
GO:0009607	response to biotic stimulus	1.82E-03
GO:0006954	inflammatory response	1.98E-03
GO:0006952	defense response	2.15E-03
GO:0016337	cell-cell adhesion	2.48E-03
GO:0006955	immune response	2.54E-03
GO:0009611	response to wounding	4.77E-03

GO classes common to gene sets I and II

GO_ID	TERM	P_VALUE (after Bonferroni correction)
GO:0000074	regulation of progression through cell cycle	2.15792E-07
GO:0007049	cell cycle	3.21191E-06
GO:0007154	cell communication	1.96302E-07
GO:0007165	signal transduction	0.0002247

GO:0007167	enzyme linked receptor protein signaling pathway	0.0092893
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	0.0051971
GO:0007243	protein kinase cascade	0.0006798
GO:0007267	cell-cell signaling	1.43889E-06
GO:0007275	Development	1.20889E-06
GO:0008283	cell proliferation	4.49668E-09
GO:0043118	negative regulation of physiological process	0.0017391
GO:0048518	positive regulation of biological process	1.09656E-05
GO:0048519	negative regulation of biological process	0.0027829
GO:0048523	negative regulation of cellular process	0.001527
GO:0050789	regulation of biological process	0.0001944
GO:0050790	regulation of enzyme activity	0.0063527
GO:0050791	regulation of physiological process	0.0005825
GO:0050794	regulation of cellular process	9.98901E-05
GO:0051243	negative regulation of cellular physiological process	0.0010439
GO:0051244	regulation of cellular physiological process	0.0003193

Supplementary Table ST5

Yeast heat shock response gene list. Fourteen (components of ribosomal unit) of 38 genes downregulated (Cluster I) and 10 (UPR) of 29 genes upregulated are in bold.

Downregulated genes (Cluster I)	**Upregulated genes (Cluster II)**
YBR181C	**YCL044C**
YBR191W	**YDL022W**
YDL075W	**YEL060C**
YFR031C-A	**YMR251W-A**
YGR214W	**YNL241C**
YHL001W	**YBL064C**
YIL148W	**YFL016C**
YJL177W	**YLL039C**
YJR145C	**YBR126C**
YPL079W	**YFL014W**
YDR447C	YBR006W
YEL054C	YBR105C
YOL120C	YBR132C
YPL081W	YBR230C
YAL035W	YDL110C
YBL076C	YDL222C
YBR106W	YGR268C
YBR263W	YHR209W
YBR288C	YIL124W
YCL029C	YJL164C

YCR047C	YJR019C
YCR054C	YKL085W
YDL051W	YKL148C
YDR385W	YLR327C
YFL023W	YML100W
YGR221C	YMR181C
YHR081W	YNL015W
YIR035C	YOR065W
YJR016C	YPL230W
YLR336C	
YLR401C	
YMR038C	
YMR078C	
YMR108W	
YNL113W	
YNL149C	
YNL251C	
YPR048W	