Istituto di analisi dei sistemi ed informatica "Antonio Ruberti" (IASI)

Bioinformatics and Computational Biology

Daniele Santoni

June 2020

Research areas

- **Epigenetics – DNA methylation and Cancer: Genomic differentially methylated Micro and Macro windows in human**
- **Immunology – Immunoinformatics: Peptides, distance from human self and MHC class I binding**
- **Nucleosome occupancy - Chromatin structure: Analysis of nucleotide sequences by tetranucleotide helical rise profiles**
- **Co-occurring Mutations in HIV Integrase: designing a map of mutations in treated and untreated patients based on mutual information**
- **Codon usage and the role of rare codon patterns in protein folding**
- **Protein contact networks - An Emerging Paradigm in Chemistry**
- **Nullomers and high order nullomers in genomic sequences**

Daniele Santoni - June 2020

DNA Methylation and Cancer

DNA Methylation → It is a molecular process consisting in the reversible addition of a -CH3 methyl group in the Cytosine occurring in the dinucleotide CpG

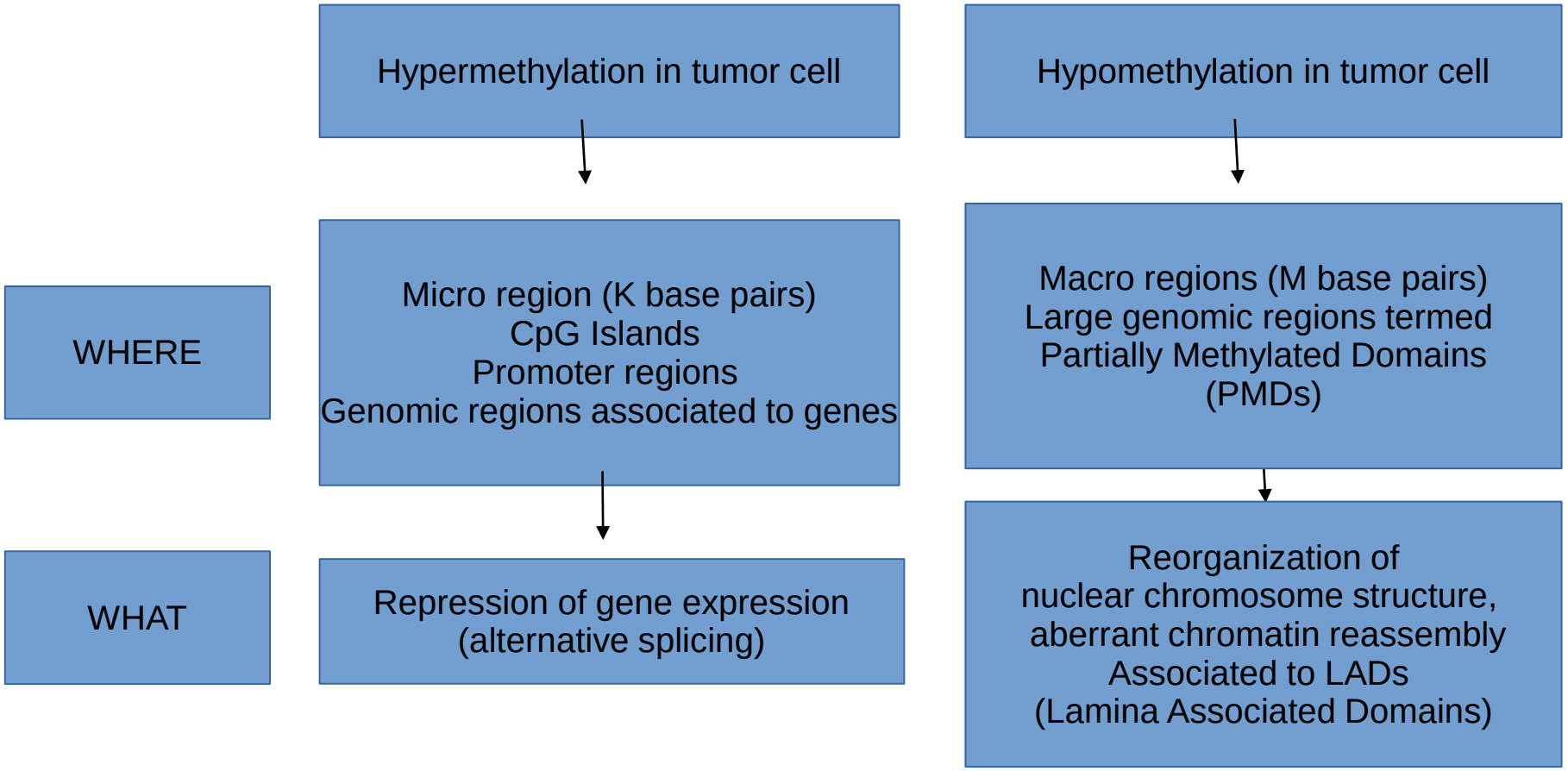
Methylation of the Cytosines occurring in the dinucleotide CpG is the most relevant mechanism that allows epigenetics to provide a further level of gene expression regulation (it has for instance a major role in cellular differentiation).

Since the end of ’70s it was clear that methylation events played a role in cancer [Holliday, 1979] and later in the ’80s several studies and models were proposed [Frost and Kerbel, 1983, Jones, 1986] in this context.

Later starting from 2000, with the availability of new platforms (Illumina) able to perform analysis on the whole human methylome, it was observed that both hypomethylation and hypermethylation occur in tumor cells.
It was also observed that in several cancer type and chromosome regions, the genome is globally subject more to hypomethylation rather than hypermethylation
(Lister et al., 2009, Hon et al., 2012, Madakashira and Sadler 2017).

Hypo/Hypermethylation in Cancer

Hypermethylation and Hypomethylation of tumor cells are typically associated to different genomic regions and are responsible for different functional and structural biological alteration



TCGA – The Cancer Genome Atlas Dataset

Illumina Infinium Human Methylation 450 K array about - 450k probes.
For each sample and for each probe a beta value
(0≤beta≤1) accounting for the methylation probability of the cytosine associated to the probe is available
(there are anyway missing values).

Tumor type	TCGA Abbreviation	Number of samples (available couples healthy/disease)
Breast Invasive Carcinoma	BRCA	96
Head and Neck Squamous Carcinoma	HNSC	50
Kidney Renal Clear Carcinoma	KIRC	160
Liver Hepatocellular Carcinoma	LIHC	50
Prostate Adenocarcinoma	PRAD	50
Thyroid Carcinoma	THCA	56

Statistical evaluation of differential methylation between Tumor and Healthy samples

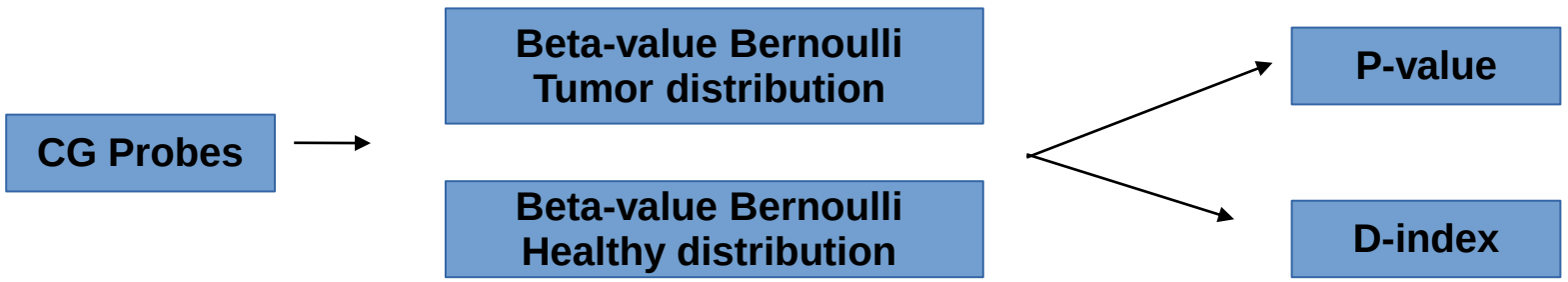
Mann-Whitney U test

Providing a P-value related to each given probe and each cancer type related to differential methylation between helathy and disease

D-index

$$d(k) = \frac{\mu_C(k) - \mu_H(k)}{(\sigma_C(k) + \sigma_H(k))/2}$$

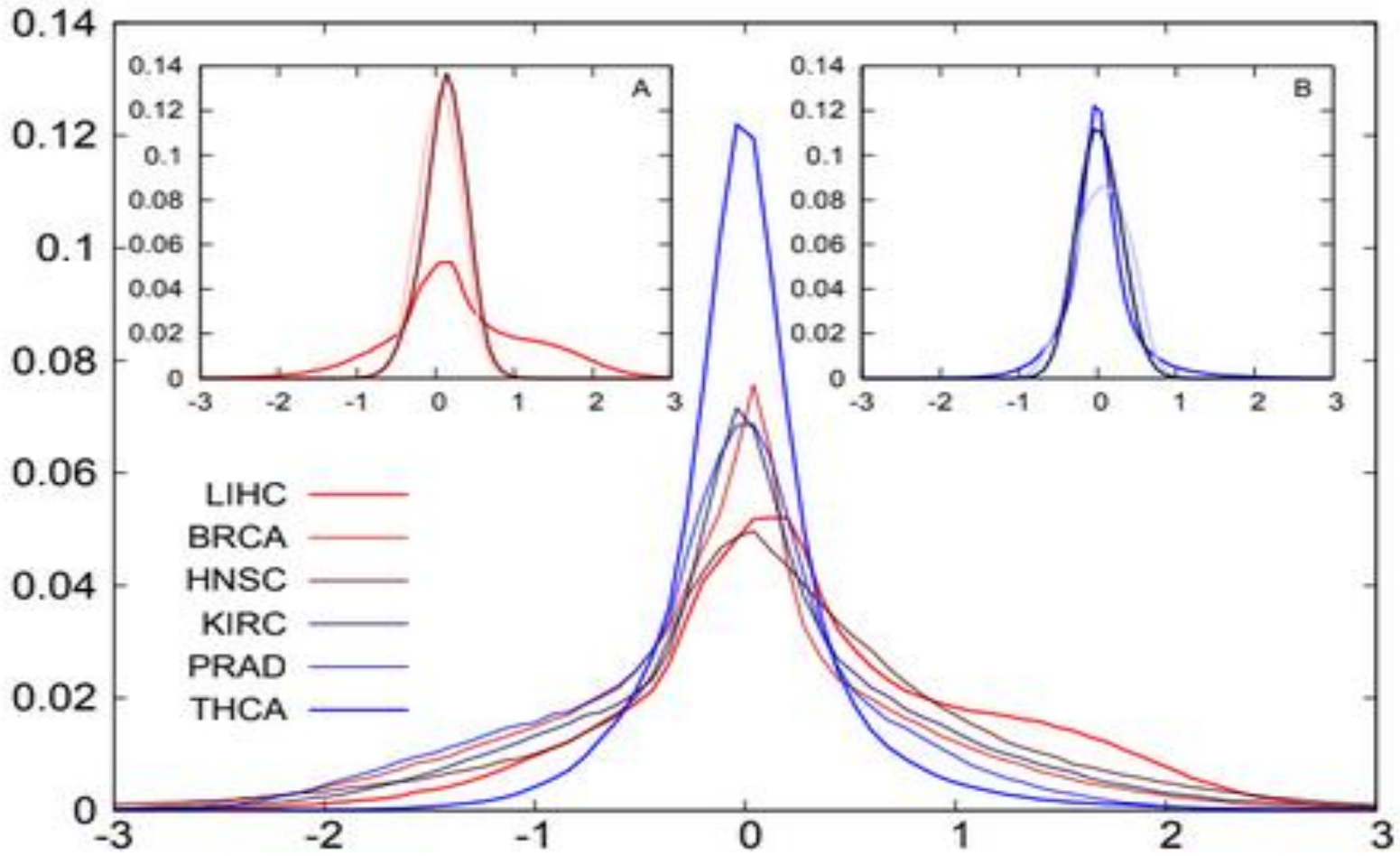
where μC(k) and μH(k) are the average values of the cancer C(k) and healthy samples H(k), respectively, and σC (k) and σH(k) their standard deviations.



Statistics on differential methylation between Tumor and Healthy samples

Disease	<d>	# Hyper	# Hypo	Unreliable	# Hyper $P < 10^{-2}$	# Hypo $P < 10^{-2}$	# Hyper $P < 10^{-3}$	# Hypo $P < 10^{-3}$
BRCA	-0.09	208,478	187,048	90,052	123,057	85,642	104,531	69,995
HNSC	0.11	172,433	222,777	90,638	77,761	114,681	61,337	90,324
KIRC	0.05	192,517	203,173	89,888	117,804	124,424	102,897	109,835
LIHC	0.30	145,561	249,874	90,143	45,889	134,515	30,674	106,312
PRAD	-0.17	224,561	171,303	89,714	106,992	51,071	88,314	36,566
THCA	0.02	199,687	195,970	89,921	54,218	46,823	35,343	31,760

Distribution of d-values in different cancer types



Statistics on 2K and 1M base pairs windows

Disease	Win#	N_L	%S	%Hypo	%Hyper	$\langle d_L \rangle$ Hypo	$\langle d_L \rangle$ Hyper
BRCA	36	~66	~93	~67	~26	~1.70	~-1.50
HNSC	28	~102	~94	~50	~44	~2.03	~-2.19
KIRC	20	~61	~91	~53	~38	~1.37	~-1.37
LIHC	37	~185	~94	~93	~1	~2.04	~-1.00
PRAD	3	~67	~92	~61	~31	~1.19	~-1.13
THCA	0	0	0	0	0	0	0

1M bp window features - minimum of 50 probes of which 90% significative with a P-value < 10^{-3}

Disease	Win#	N_L	%S	%Hypo	%Hyper	$\langle d_L \rangle$ Hypo	$\langle d_L \rangle$ Hyper
BRCA	65	~27	~98	~4	~94	~1.02	~-1.52
HNSC	69	~27	~99	~15	~84	~1.58	~-1.98
KIRC	54	~28	~98	~23	~75	~1.11	~-1.50
LIHC	29	~27	~98	~58	~40	~1.60	~-1.83
PRAD	46	~28	~98	0	~98	-	~-1.46
THCA	2	~32	~97	0	~97	0	~-1.3

2k bp window features - minimum of 20 probes of which 95% significative with a P-value < 10^{-3}

Note: Hypermethylation is associated to negative D values while Hypomethylation is associated to positive D values, since D is based on differences between Healthy - Disease

Best 1M base pairs windows

Id	Disease	Chr	Region	N_L	%S	%Hypo	%Hyper	$\langle d_L \rangle$ Hypo	$\langle d_L \rangle$ Hyper
1	BRCA	chr7	153.400.000-154.400.000	58	∼93	∼43	∼50	∼1.59	∼-1.94
2	BRCA	chr12	126.400.000-127.600.000	54	∼92	∼61	∼31	∼2.12	∼-1.87
3	BRCA	chr16	6.800.000-7.800.000	51	∼90	∼74	∼16	∼2.12	∼-1.87
4	HNSC	chr2	78.800.000-79.800.000	52	∼94	∼61	∼33	∼2.75	∼-2.72
5	HNSC	chr8	63.400.000-64.400.000	64	∼92	∼11	∼81	∼1.53	∼-2.86
6	HNSC	chr19	20.000.000-21.000.000	53	100	∼13	∼87	∼2.47	∼-2.46
7	KIRC	chr5	63.600.000-64.600.000	54	∼92	∼9	∼83	∼1.91	∼-1.43
8	LIHC	chr5	3800000-4800000	72	∼97	∼97	0	∼2.32	-
9	LIHC	chr21	30.000.000-31.000.000	91	100	100	0	∼2.36	-

A relevant overlap between those regions and LADs studied and experimentally identified by Guelen and colleagues [Guelen et al., 2008] was observed.
4 windows out of 9 are totally contained in LADs (show a relevant hypomethylated pattern) while other 3 windows show a significant overlap with LADs of Guelen study.

In particular window 1 is contained in LAD chr7:152172250-154954755, as well as window 3 LAD chr16:5153535-8604427, window 4 contained in LAD chr2:75937802-80210697, window 5 contained and window number 8 contained in LAD chr5:1662775-5410587. A significant overlap was observed for window 5 and 7 (with LADS chr8:62727813-64076406 and chr5:61782370-64054900).

Best 2K base pairs windows

Id	Disease	Chr	Region	$\#w$	$\langle \mathcal{N}_L \rangle$	%S	Methyl	$\langle d_L \rangle$	Annotation
1	BRCA	chr6	29552000-29555500	4	∼31	100%	Hyper	∼-2.19	OR2I1P unproc. pseudogene
2	BRCA	chr6	30126000-30129000	3	∼29	98%	Hyper	∼-2.40	NONE
3	BRCA	chr7	27129500-27131500	1	21	100%	Hyper	∼-2.29	HOXA-AS2 antisense HOXA-AS3 antisense HOXA3 protein coding HOXA4 protein coding RP1-170O19.22 proc. transcript
4	LIHC	chr10	132785000-132788500	4	∼28	98%	Hyper	∼-2.35	NKX6-2 protein coding
5	LIHC	chr14	101022000-101024000	1	20	100%	Hypo	∼1.94	AL132709.2 MIR299 MIR411 MIR379
6	LIHC	chr14	101065000-101067000	1	22	95%	Hypo	∼1.82	MIR369 MIR409 MIR410 MIR412 MIR656
7	KIRC	chr7	27142500-27146500	5	∼36	100%	Hyper	∼-2.40	HOXA-AS3 antisense HOXA3 protein coding HOXA5 protein coding RP1-170O19.22 proc. transcript
8	KIRC	chr6	30683500-30686500	3	∼22	100%	Hypo	∼2.25	PPP1R18 protein coding AL662797.1 miRNA
9	KIRC	chr1	24931000-24933000	1	∼22	100%	Hyper	∼-2.00	RUNX3 protein coding
10	PRAD	chr6	30006000-30008500	2	∼45	98%	Hyper	∼-2.38	HCG4P3 unproc. pseudogene HLA-J transcr. unproc. pseudogene ZNRD1-AS1 antisense
11	PRAD	chr6	30128000-30129000	3	∼29	96%	Hyper	∼-1.89	NONE
12	PRAD	chr7	95395000-95397000	1	22	95%	Hyper	∼-2.10	PON3 protein coding
13	HNSC	chr7	27185500-27187500	1	21	100%	Hyper	∼-2.83	HOXA11 protein coding HOXA11-AS antisense HOXA11-AS1-4 misc-RNA HOXA11-AS1-5 misc-RNA RP1-170O19.14 linc-RNA
14	HNSC	chr6	28634000-28637000	3	∼32	100%	Hyper	∼-2.52	RP11-373N24.2 proc. pseudogene
15	HNSC	chr1	24931000-24933000	1	22	95%	Hyper	∼-2.50	RUNX3 protein coding

Best 2K base pairs windows features and related functional elements

Pan-cancer windows

Id	Disease	Chr	Region	$\#w$	$\langle \mathcal{N}_L \rangle$	$\langle \mathcal{N}_L^+ \rangle$	$\langle \mathcal{N}_L^- \rangle$	$\langle d_L^+ \rangle$	$\langle d_L^- \rangle$	Annotation
1	BRCA HNSC KIRC LIHC PRAD THCA	chr14	101022000-101024000	1	20	18.33	0	1.50	0	AL132709.2 MIR299 MIR411 MIR379
2	BRCA HNSC KIRC LIHC PRAD	chr6	29827000-29829000	1	30	1.8	~21.8	1.21	-1.43	HLA-G
3	BRCA HNSC KIRC LIHC PRAD	chr6	30126000-30129000	3	29	0.73	28.47	0.81	-1.91	NONE

Id	Disease	Chr	Region	$\#w$	$\langle \mathcal{N}_L \rangle$	$\langle \mathcal{N}_L^+ \rangle$	$\langle \mathcal{N}_L^- \rangle$	$\langle d_L^+ \rangle$	$\langle d_L^- \rangle$
1	BRCA HNSC KIRC LIHC PRAD	chr1	157400000-159000000	4	132.75	85.95	18.3	1.89	-1.10
2	BRCA HNSC KIRC LIHC	chr11	4200000-5600000	3	133.67	105.33	6.83	1.85	-0.81
3	BRCA KIRC LIHC PRAD	chr1	152200000-153200000	1	190	149.75	10	1.66	-1.29
4	BRCA HNSC KIRC LIHC	chr1	247400000-248400000	1	167	111	28.75	1.87	-1.48
5	BRCA HNSC KIRC PRAD	chr6	99600000-100600000	1	195	20.25	138.5	1.19	-1.59
6	BRCA HNSC KIRC LIHC	chr10	130400000-131400000	1	176	124.25	20.75	1.56	-1.27

All pan-cancer windows reported in the table result to be associated to one or more LADs [Guelen et al., 2008] : window 1 with two LADs 157145909- 157930561 and 158098845-158966229, window 2 with 4169582-5213736, window 3 with 152025174-153450845, window 4 with 247581217-249099782, window 5 with 97742337-99777339 and 100057867-100545737, window 6 with 129251344-131662247.

The farther the better: investigating how distance from human self affects the propensity peptides to be presented on cell surface by MHC class I complex

Our analysis is based on two major *steps* that are pivotal in the ability of immune system to be effective in fighting pathogens:

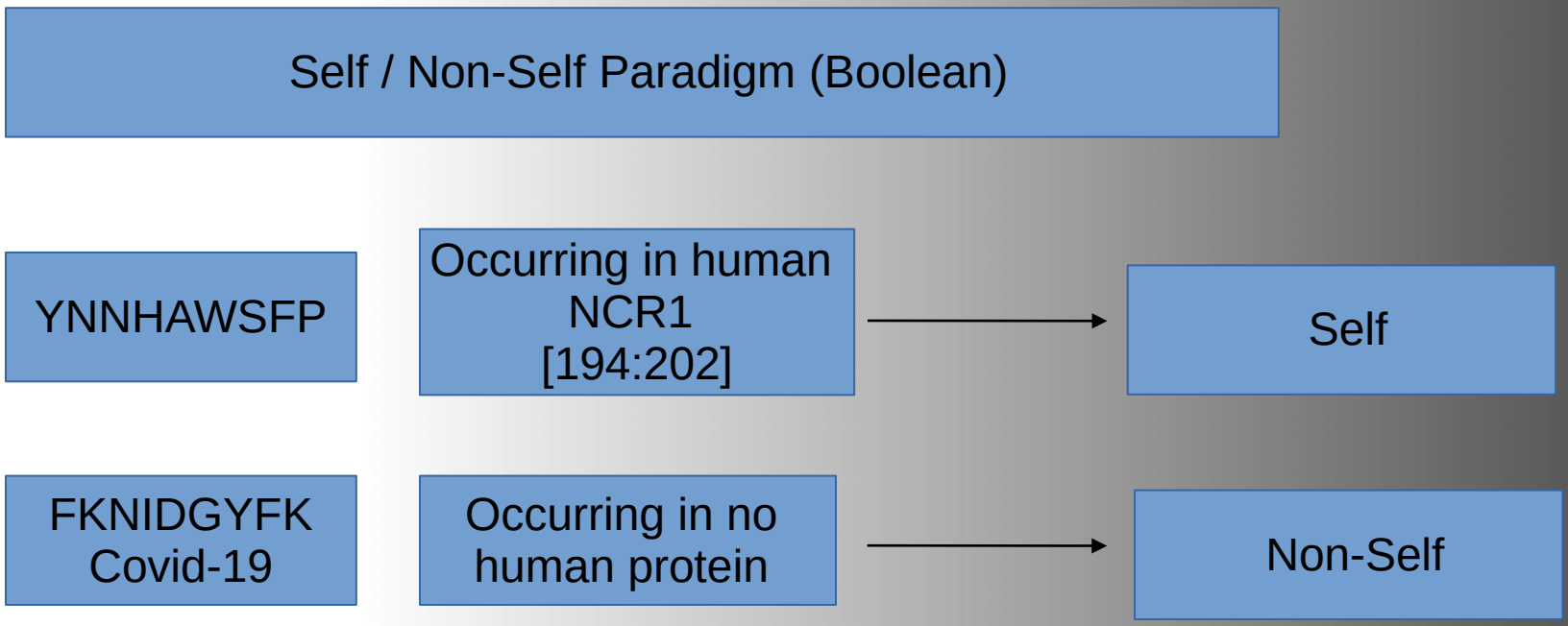
- **Peptides presentation on cell surface through MHC class-I.**

Most of human cells (in particular APC cells Antigen Presenting Cells) expose on their surface some portion (peptides) of internal proteins as a view of what they are doing and who they are. Also potential pathogen peptides (for example those peptides of viruses that invade the cell) are exposed through MHC (Major Histocompatibility Complex) class I and II.

- **Self/non-self paradigm**

Immune system is able to discriminate between presented peptides coming from endogenous proteins or exogenous (potentially harmful) ones.

Boolean and Liquid Self / Non-Self



In this work the analysis was performed on the set of contiguous 9-mers (since it is known from literature that 85-90% of epitopes are contiguous 9-mers) and we only focused on Class I MHC.

Human self (GRCh38) consists of ~ 11.000.000 unique 9-mers contiguous peptides (in the following we will refer to human self as HSA).

Boolean and Liquid Self / Non-Self

Non-Self peptides (everything excepting human peptides) are very different from each other in terms of aminoacid composition and structure

Some of them are absent from self but very close to many human peptides (one mutation step for example)

Some others are very far from human self peptides

Let's formally define a distance between a given peptide and human self (HSA)

Boolean and Liquid Self / Non-Self

Given a 9-mer p and a 9-mer q belonging to human self we define their distance as the Hamming distance D(p,q)

$$B(i) = \begin{cases} 0 & p_i = q_i \\ 1 & p_i \neq q_i \end{cases}$$

$$D(p,q) = \sum_{i=1}^{9} B(i)$$

Hamming distance D(p,q)

Distance from Human self M(p)

$$M(p) = \min_{q \in HSA} \{D(p,q)\}$$

Distance M ranges from 0 to 9

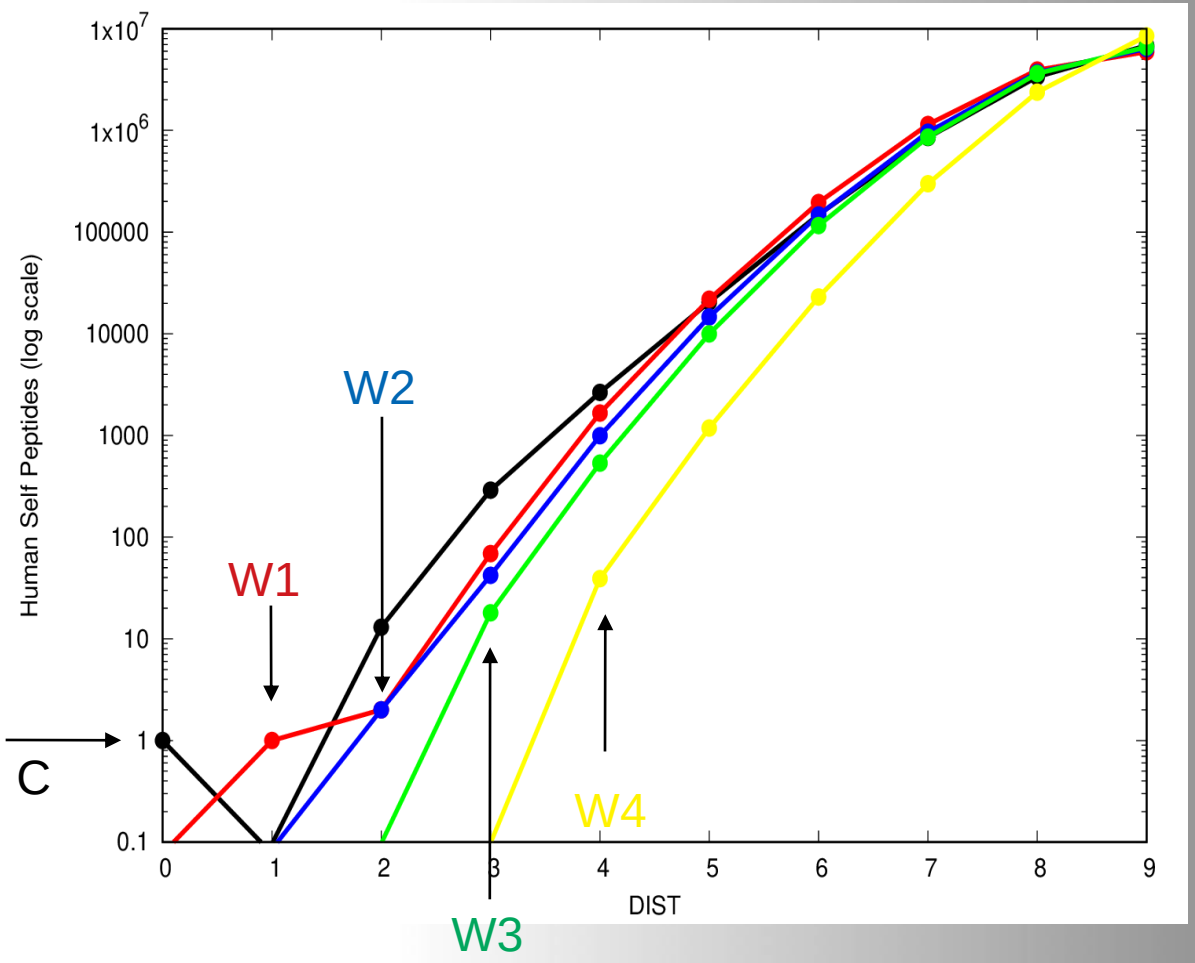
Self/Non-Self Classes (Liquid)

Partition of 9-mers in classes
C, W1, W2 … W4
for Trypanosoma c, HIV-1, HHV-1 and Covid-19

Distance M naturally defines a partition of peptides of a given organism in classes C (Common) W1, W2, W3 and W4

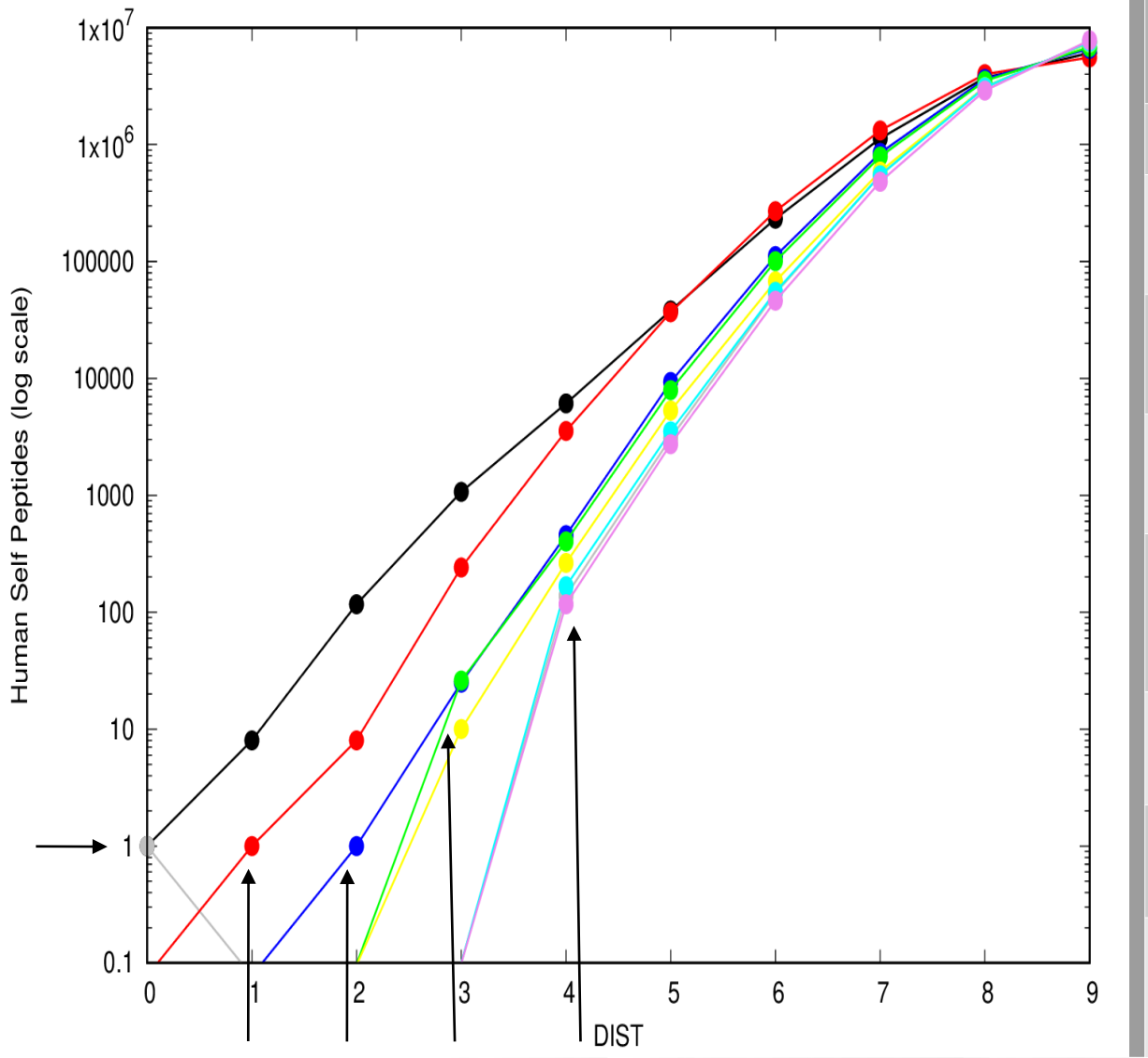
Organism	P	C (%)	W1 (%)	W2 (%)	W3 (%)	W4 (%)
Tc	8,937,165	19,966 (0.22)	303,662 (3.40)	4,663,994 (52.19)	3,939,218 (44.08)	10,325 (0.11)
HIV1	4,530,559	868 (0.02)	63,530 (1.40)	4,466,161 (43.09)	2,503,467 (55.25)	10,366 (0.23)
HHV1	77,173	298 (0.39)	4,131 (5.35)	44,542 (57.71)	28,202 (36.54)	70 (0.09)
Covid-19	9,591	2 (0.00)	156 (1.62)	4,104 (42.79)	5,302 (55.28)	27 (0.28)

Distribution of peptides distances from Human Self



Peptide	Organism Class	Color
KKDKKKKAD	Covid-19 C	Black
LSEARQHLK	Covid-19 W1	Red
DSVTVKNGS	Covid-19 W2	Blue
TAESHVDTD	Covid-19 W3	Green
CWHTNCYDY	Covid-19 W4	Yellow

Distribution of peptides distances from Human Self

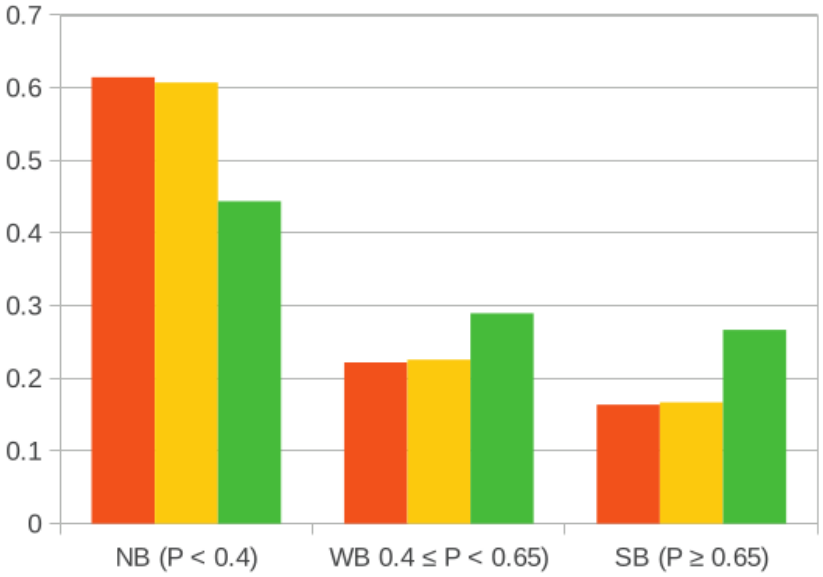


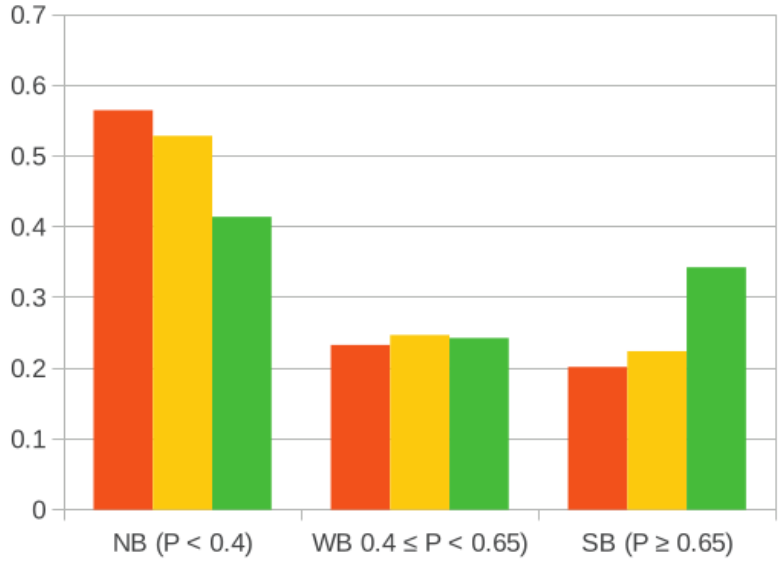
Peptide	Organism Class	Color
PTGPLGPPG	Human	Black
YNNHAWSFP	Human (W4)	Gray
LGGGSASTS	Tc W1	Red
PEFVLTDYT	Tc W2	Blue
RSVQGTWNP	Tc W3	Green
FKNIDGYFK	Covid19 W3	Yellow
NWQVNNRNE	Tc W4	Cyan
YYHKNNKSW	Covid19 W4	Violet

Human Immunodeficiency Virus type 1 - HIV1 and Human herpes simplex virus 1 – HHV1 Binding classes MHC class-I

NetMHC 4.0 (81 different HLAs were considered)

We define a peptide as a strong binder (SB) if it strongly binds at least one out of the 81 considered HLAs. In the same way a peptide is defined as a weak binder (WB) if it is not a SB and it weakly binds al least one out of the 81 considered HLAs. No bind (NB) otherwise.

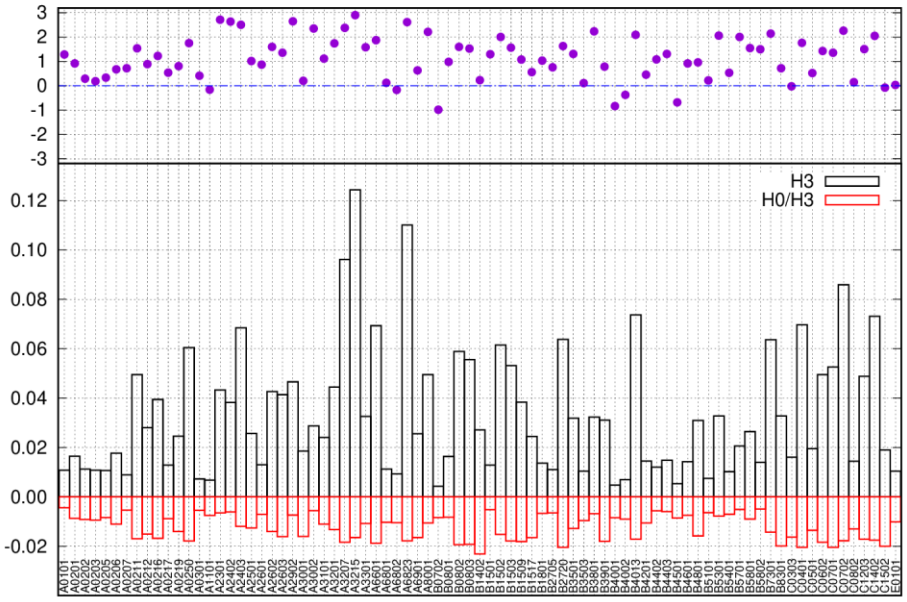




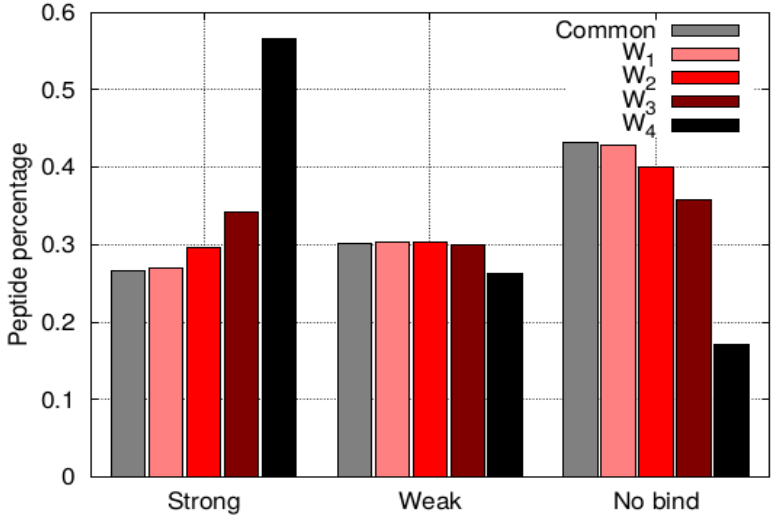
Green Bar is related to W4
Yellow Bar is related to W3
Red Bar is related to all peptides

Trypanosoma cruzi dsitribution of MHC-I binding of W classes

NetMHC 4.0 (81 different HLAs were considered)

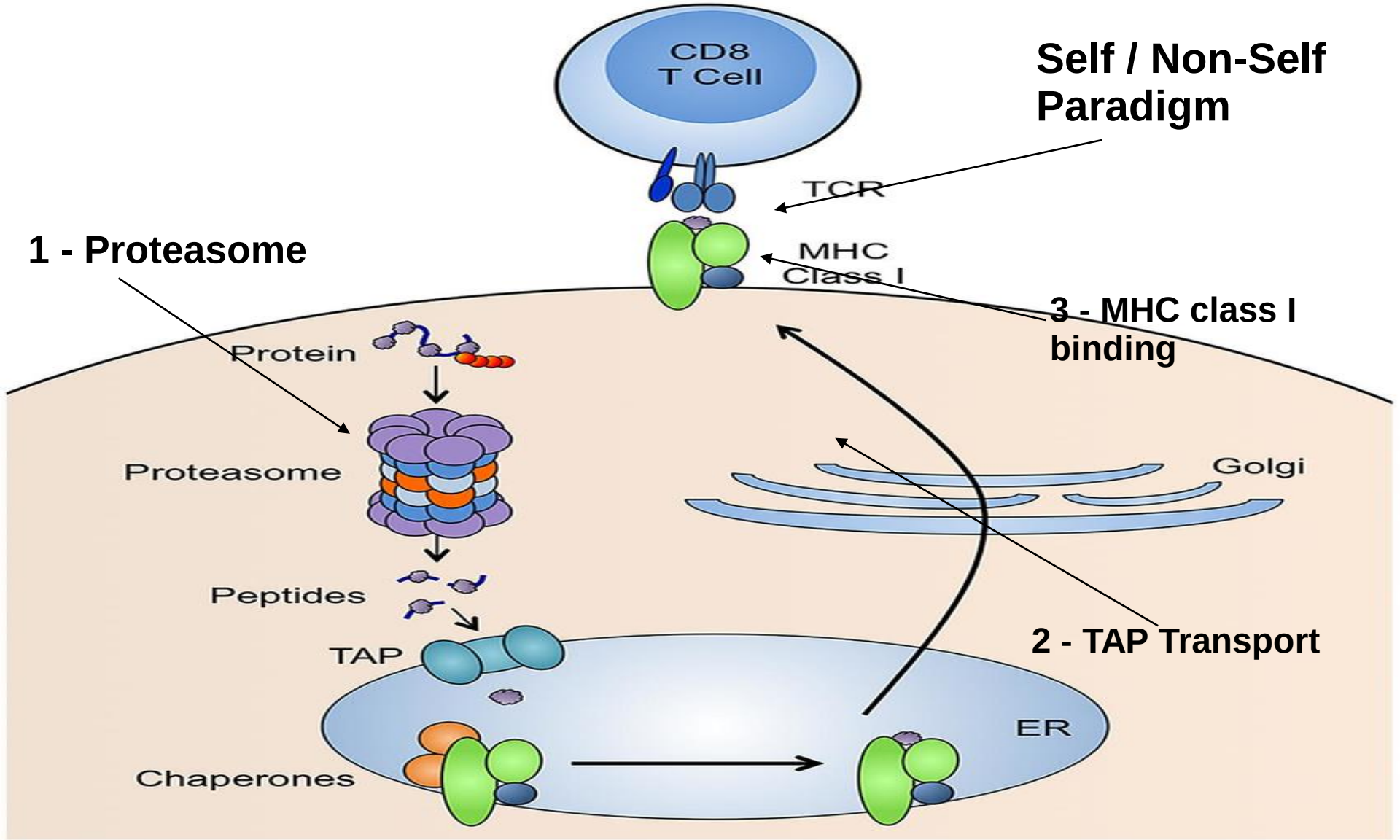


Percentage of Strong bind peptides for each HLAs for W4 (black) vs other classes (red).



We define a peptide as a strong binder (SB) if it strongly binds at least one out of the 81 considered HLAs. In the same way a peptide is defined as a weak binder (WB) if it is not a SB and it weakly binds al least one out of the 81 considered HLAs. No bind (NB) otherwise.

Immunology: proteasome cleveage, TAP-transport and MHC class-I binding



Further data

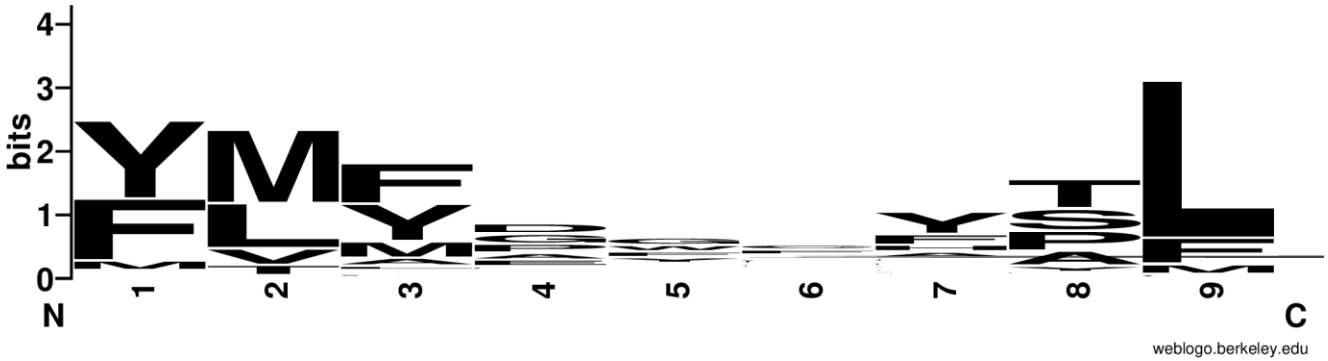
Co-evolution in the presentation pipeline of petides
Differences between whole set of and W4 peptides

W	CLE (%)	TAP (%)	MHC-I SB (%)
8,917,199	2,903,390 (32)	2,388,082 (82)	1,603,821 (67)
8,917,199	—	3,985,392 (45)	2,171,288 (54)
8,917,199	—	—	2,819,532 (32)

All peptides presentation pipeline

W4 presentation pipeline

W_4	CLE (%)	TAP (%)	MHC-I SB (%)
10,325	3,217 (31)	2,903 (90)	2,563 (88)
10,325	—	5,994 (58)	4,526 (76)
10,325	—	—	5,843 (57)



Aminoacid patterns of presented peptides (Logoplot)

Further data

Covid-19 potential epitopes

Peptide	#SB-HLA	#WB-HLA	ORF(nsp)	Position	CLES-core	TAP-Score
YQCGHYKHI	7	9	ORF1(nsp3)	1831-1839	0.1801	0.7400
CMMCYKRNR	3	2	ORF1(nsp3)	2409-2417	0.0427	1.6400
HNWNCVNCD	1	0	ORF1(nsp3)	2448-2456	0.0541	-1.8390
HIQWMVMFT	1	5	ORF1(nsp4)	3125-3133	0.0236	-0.5460
CISTKHFYW	**3**	**8**	**ORF1(nsp4)**	**3147-3155**	**0.7680**	**0.9770**
YQCAMRPNF	9	15	ORF1(nsp5)	3389-3397	0.0536	2.6460
SWVMRIMTW	3	7	ORF1(nsp6)	3658-3666	**0.6403**	**1.4160**
CKCCYDHVI	2	2	ORF1(nsp13)	5351-5359	0.2798	0.5450
MMGFKMNYQ	2	2	ORF1(nsp14)	5982-5990	0.0269	-0.0460
WHHSIGFDY	**4**	**7**	**ORF1(nsp14)**	**6152-6160**	**0.6175**	**2.9280**
KLMGHFAWW	**8**	**13**	**ORF1(nsp16)**	**6980-6988**	**0.4431**	**1.0240**
YVMHANYIF	**27**	**17**	**ORF1(nsp16)**	**7020-7028**	**0.5218**	**2.6880**
YYHKNNKSW*	**8**	**12**	**ORF2**	**144-152**	**0.9329**	**1.2060**
MQMAYRFNG	3	5	ORF2	900-908	0.0234	-1.0540
YIKWPWYIW	**10**	**14**	**ORF2**	**1209-1217**	**0.8066**	**0.9420**
IMRLWLCWK	**4**	**5**	**ORF3**	**124-132**	**0.8803**	**0.6670**
MRLWLCWKC	1	5	ORF3	125-133	0.0894	0.2950
FLCWHTNCY	**11**	**14**	**ORF3**	**146-154**	**0.9264**	**2.806**
CWHTNCYDY	**3**	**6**	**ORF3**	**148-156**	**0.9621**	**3.2080**

Best ranking W4 peptides. Second and third columns reported the numbers of HLAs the peptide strongly and weakly, respectively, bind. In the fourth column the ORF in which the peptide occur is reported. Fifth column indicates the position of peptide in the correspondent ORF. Proteasome Cleavage and TAP-transport predicted scores are reported in sixth and seventh column, respectively. The peptide YYHKNNKSW, whose aminoacids result to be belong to the solvent-accessible surface area of the spike protein, has been labeled with “*”

Nucleosome occupancy - Chromatin structure: Analysis of nucleotide sequences by tetranucleotide helical rise profiles

Nucleosome is the basic structural unit of DNA packaging in eukaryotes and it typically consists of 146 DNA base pairs wrapped around a protein complex (histone octamer).

Nucleosome occupancy is driven by several factors that synergystically play in order to determine nucleosome landscape.

Nucleome mapping (how the nucleosomes are located along the chromosomes) play a fundamental role in may aspects of cell life.

We distinguish three categories of sequences according to their ability to form stable nucleosome (Segal and Widom).

- **Stable** (always forming strong nucleosome)
- **Facultative** (they can form nucleosome or not depending on cell state)
- **Nucleosome free** (never forming nucleosome)

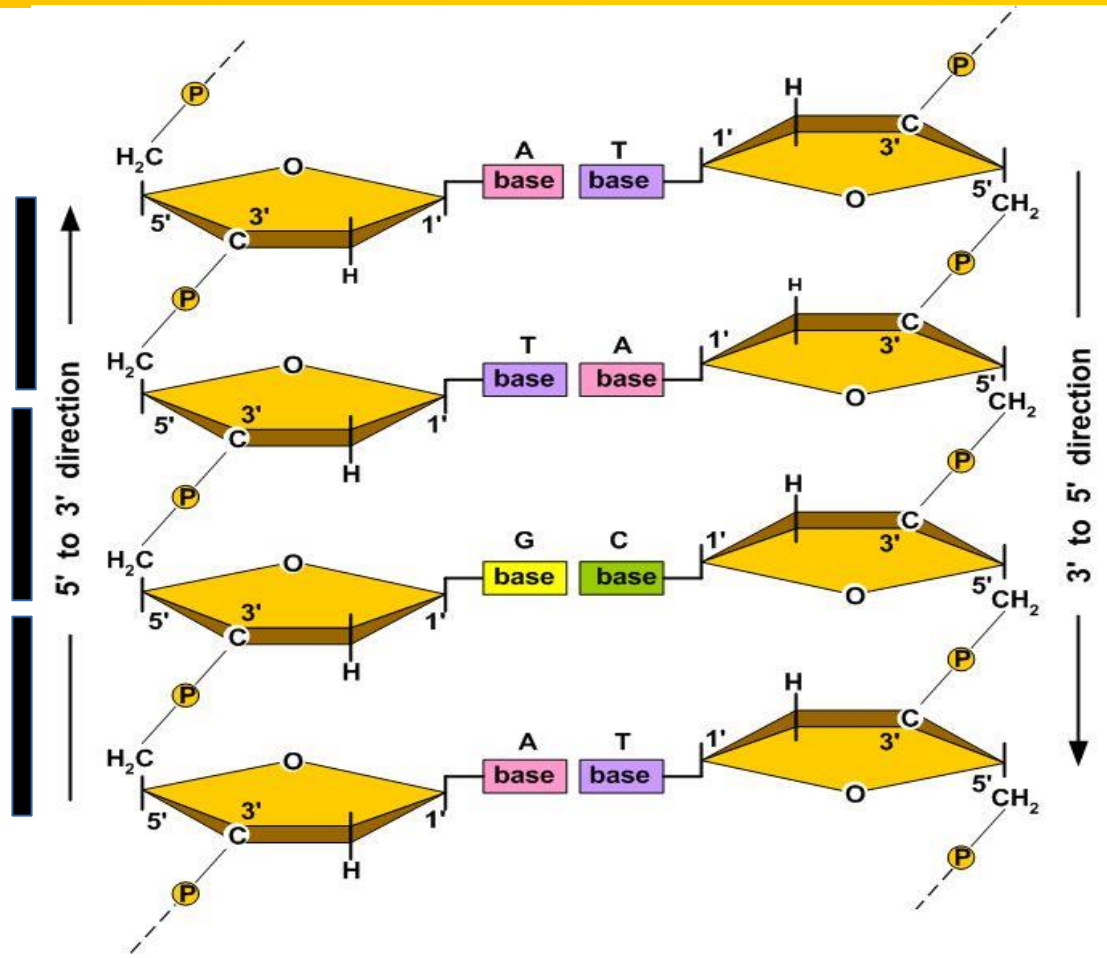
The first and central player that impacts on nucleosome formation is the sequence. The aim of our work is to develop a sequence-dependent algorithm to identify those genomic sequence-regions that are likely to form stable nucleosomes.

What is DNA Helical Rise ?

Helical rise

Helical rise of dinucleotide GT/AC derived from tetrad AGTA/TACT

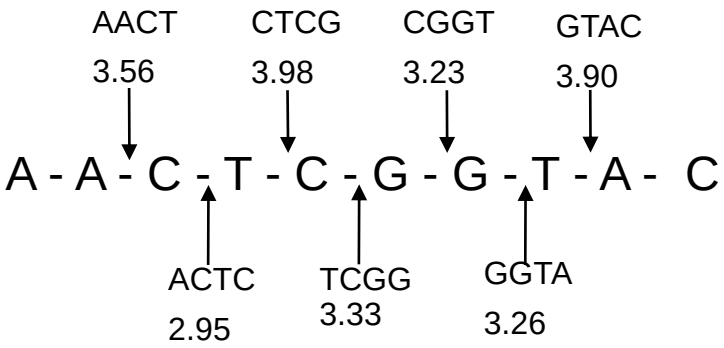
Helical rise



Preferential Nucleosome Occupancy and DNA Helical Rise

Average 3.2 Å
Maximal value 4.46 Å (CGCA/TGCG)
Minimal value 2.36 Å (ATGA/TCAT)

Bistable dinucleotides
CG - TA - CC/GG and AG/CT

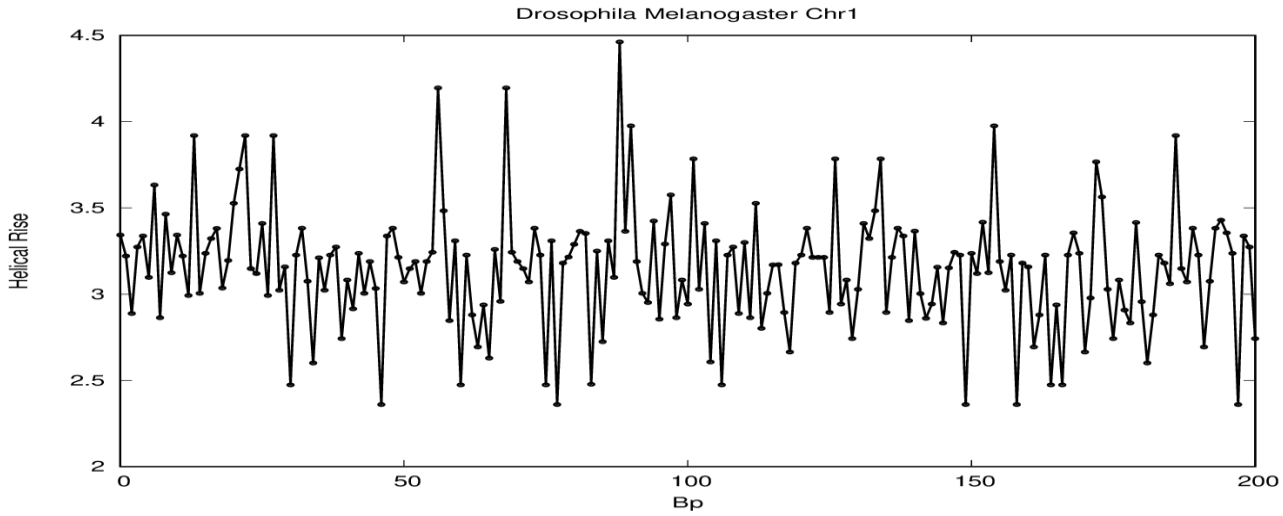


Tetranucleotide helical rise values for dinucleotide CG

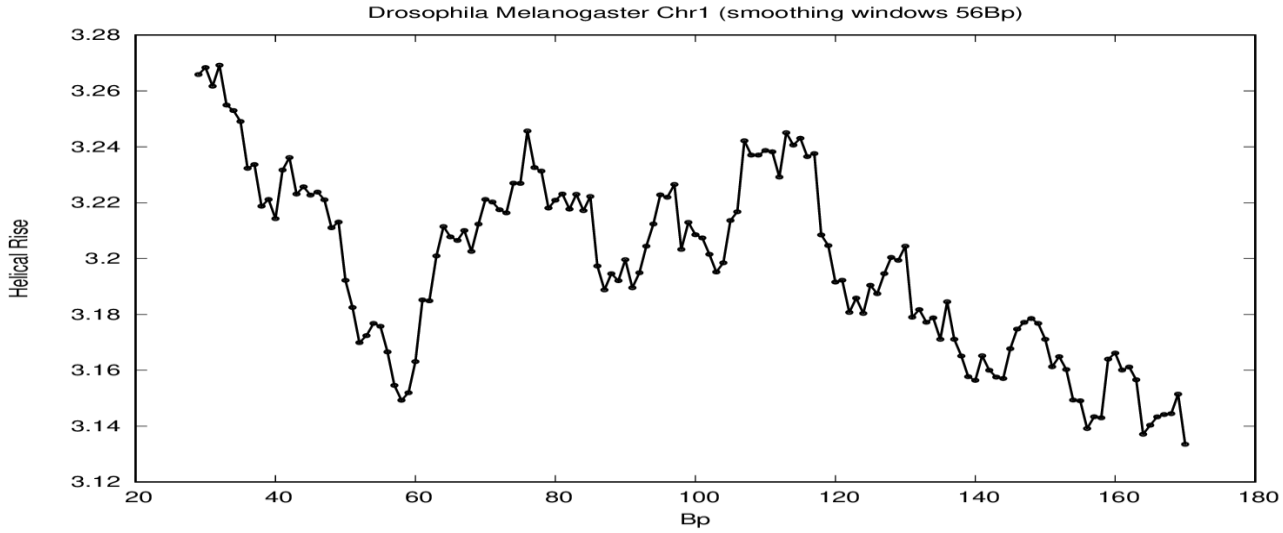
117	ACGA\TCGT	3.12
118	ACGG\CCGT	2.81
119	ACGC\GCGT	2.48
120	ACGT	4.02
121	TCGC\GCGA	3.36
122	CCGC\GCGG	2.83
123	GCGC	2.61
124	CCGA\TCGG	3.33
125	CCGG	2.85
126	TCGA	3.44

Range 2.48 – 4.02

DNA Helical Rise profiles



Plain helical rise tetranucleotide profile



Smoothing windows (56 base pairs) tetranucleotide helical rise profile

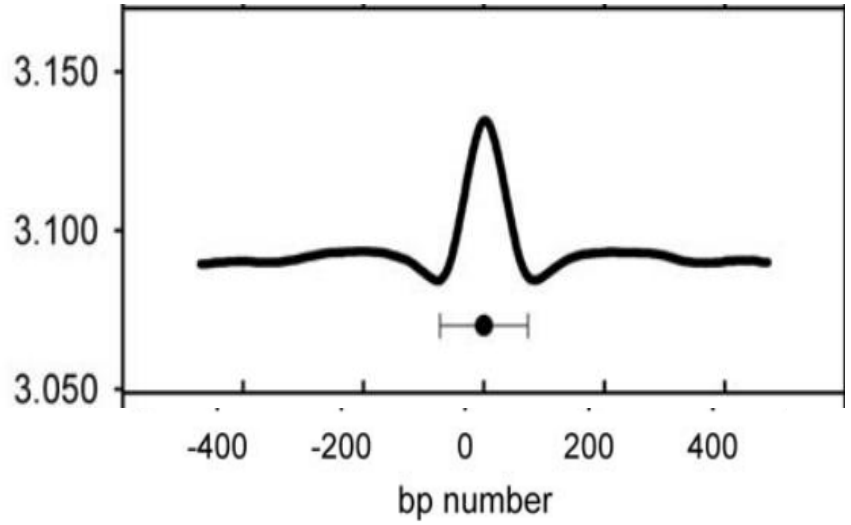
Preferential Nucleosome Occupancy at High Values of DNA Helical Rise

DNA Research **19**, 81–90, (2012)
Advance Access Publication on 9 January 2012
doi:10.1093/dnares/dsr043

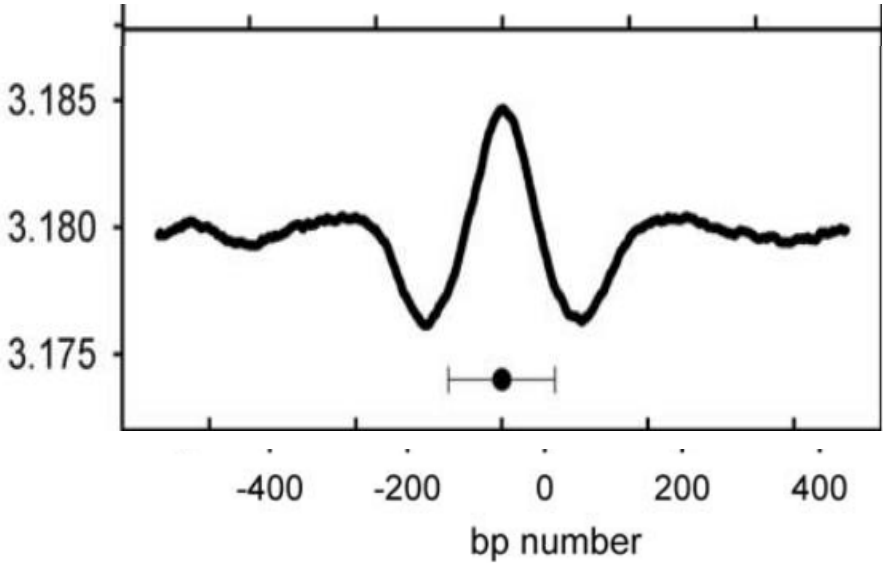
Preferential Nucleosome Occupancy at High Values of DNA Helical Rise

Francesco Pedone[1] and Daniele Santoni[2,3,*]

Department of Biology and Biotechnology, 'Sapienza' University, P.le A. Moro 5, 00185 Rome, Italy[1]; Barcelona Institute for Research in Biomedicine (IRB), Barcelona Science Park, C/Samitier 1-5, 08015 Barcelona, Spain[2] and Istituto di Analisi dei Sistemi e Informatica Antonio Ruberti, Consiglio Nazionale delle Ricerche, Viale Manzoni 30, 00185 Rome, Italy[3]



P. falciparum average of 89 115 points
(www.plasmoDB.org)



Human chromosome 20 average of 65 379 points
Schones, D.E., Cui, K., Cuddapah, S., et al. 2008, Dynamic regulation of nucleosome positioning in the human genome, Cell, 132, 887– 98.

The impact of DNA Helical rise on DNA binding protein

Contents lists available at ScienceDirect

Genomics

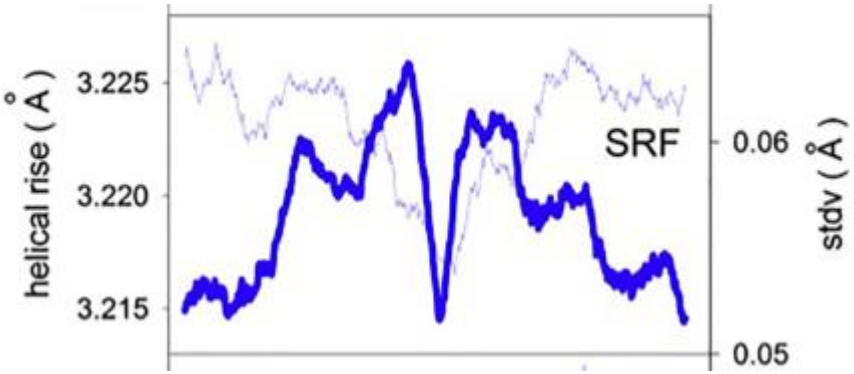
journal homepage: www.elsevier.com/locate/ygeno

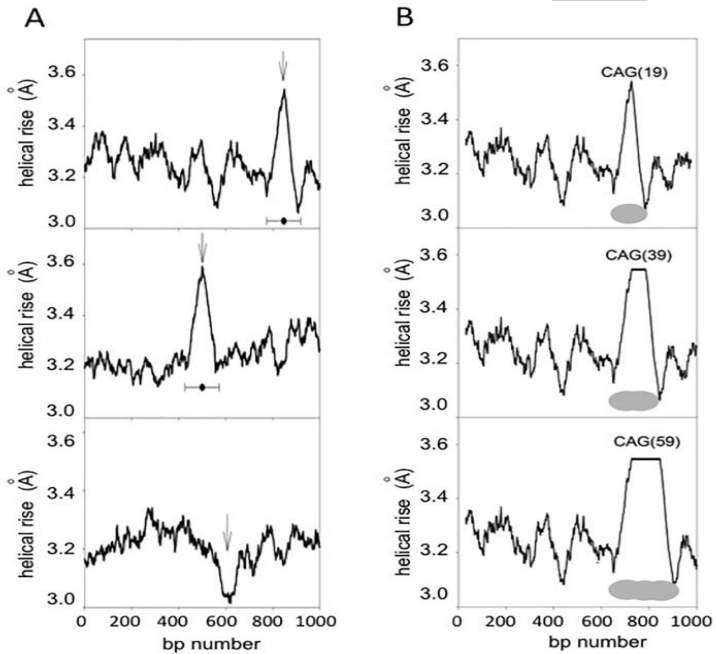
Original Article

A study of the impact of DNA helical rise on protein-DNA interaction

Francesco Pedone[a], Filomena Mazzei[b], Daniele Santoni[c,*]

[a] Department of Biology and Biotechnology, "Sapienza" University, P.le A. Moro 5, Rome 00185, Italy
[b] Environment and Health Department, Istituto Superiore di Sanità, V.le Regina Elena 299, Rome 00161, Italy
[c] National Research Council of Italy, Institute for System Analysis and Computer Science "Antonio Ruberti", Via





Thanks a lot for your attention

And I wish to thank:

DNA Methylation and Cancer:

Danilo Pignotti - Department of Biology, University Tor Vergata, Laurea magistrale in Bioinformatica

Davide Vergni - Istituto per le Applicazioni del Calcolo “Mauro Picone” - CNR

In the search of epitopes using High order Nullomers

Rosanna Gaudio - Department of Biology, University Tor Vergata, Laurea magistrale in Bioinformatica

Davide Vergni - Istituto per le Applicazioni del Calcolo “Mauro Picone” - CNR

Sequence-dependent algorithms for nucleosome mapping

Francesco Pedone - Department of Biology and Biotechnology, ‘Sapienza’ University

for the time we shared working together

Others …

DNA Research, 2018, 25(1), 103–112
doi: 10.1093/dnares/dsx041
Advance Access Publication Date: 23 October 2017
Full Paper

OXFORD

Full Paper

Investigating transcription factor synergism in humans

Fabio Cumbo[1,2,3], Davide Vergni[4], and Daniele Santoni[1,*]

[1]Institute for Systems Analysis and Computer Science 'Antonio Ruberti', National Research Council of Italy, 00185 Rome, Italy, [2]Department of Engineering, Third University of Rome, 00146 Rome, Italy, [3]SYSBIO.IT–Centre of Systems Biology, 20126 Milan, Italy, and [4]Institute for Applied Mathematics 'Mauro Picone', National Research Council of Italy, 00185 Rome, Italy

PLOS ONE

RESEARCH ARTICLE

Nullomers and High Order Nullomers in Genomic Sequences

Davide Vergni[1], Daniele Santoni[2]*

1 Istituto per le Applicazioni del Calcolo "Mauro Picone" - CNR, Via dei Taurini 19, 00185, Rome, Italy,
2 Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti" - CNR, Via dei Taurini 19, 00185, Rome, Italy

* daniele.santoni@iasi.cnr.it

Contents lists available at ScienceDirect

Journal of Immunological Methods

journal homepage: www.elsevier.com/locate/jim

ELSEVIER

Research paper

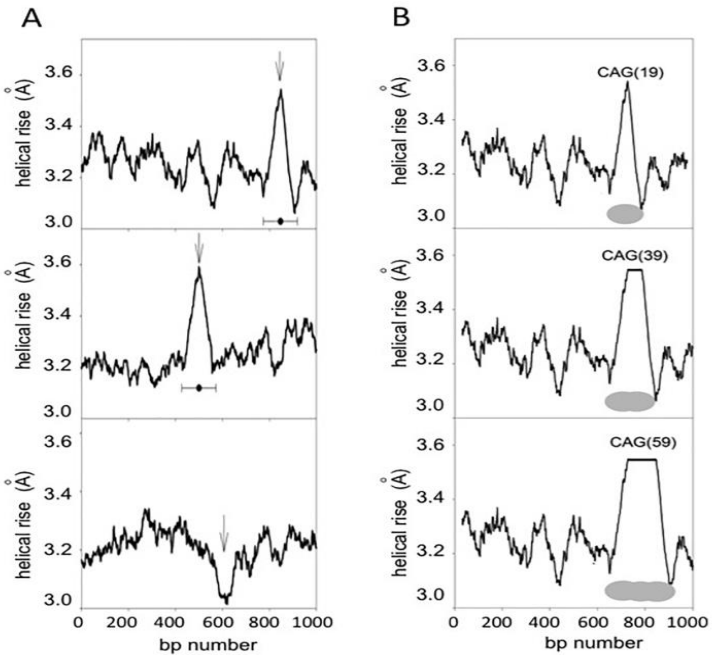
Viral peptides-MHC interaction: Binding probability and distance from human peptides

Daniele Santoni*

Institute for System Analysis and Computer Science "Antonio Ruberti", National Research Council of Italy, Via dei Taurini 19, Rome 00185, Italy

Check for updates

Further data



HTT, DMPK1 and FMR-2 genes whose mutations are associated with the most common neurodegenerative disorders (Hungtinton, Myotonic dystrophy and Fragile X). The helical rise profiles of 1kbp long sequences of the HTT, DMPK1 and FMR-2 genes carrying 19 CAG, 20 CTG and 15 CGG repeats, corresponding to the sequences found in “healthy” subjects, have been evaluated.