Molecular Properties of the M-phase inducer phosphatase 1 for chemical carcinogenesis - receptor activation

Jeff Cromwell, PhD

The Mathematical Learning Space Research Portfolio

1 Abstract

Maps of biological networks can have many interacting molecules required for study. Because of the magnitude N=548 pathways, 5 percent of these were examined with respect to the receptor categories. For human diseases such as cancer, microRNAs in cancer and chemical carcinogenesis - receptor activation has CDC25A; M-phase inducer phosphatase 1. Here several molecular properties are examined with respect to these two categories and descriptive statistics and maximum likelihood estimation techniques for Pearson distribution identification are provided for comparision and contrast. A cluster dendrogram of Kidra factors is also presented.

2 Introduction

Biological mechanisms involving receptor activation fall into two broad categories: (i) those that involve cell surface receptors and some intracellular receptors that activate signal transduction pathways, resulting in biological responses, including gene transcription, and (ii) those that involve intracellular receptors that translocate into the nucleus and act as transcription factors regulating gene expression. Both classes of receptors can be involved in mechanisms of carcinogenesis. [401]

Table 1 has a list of receptor pathways based on a sample of N=548 pathways. This list of N=27 represents 5 percent with specifically Chemical carcinogenesis - receptor activation, Retinoic acid receptor (RAR) and retinoid X receptor (RXR) agonists/antagonists, T cell receptor signaling pathway, B cell receptor signaling pathway and Cytokine-cytokine receptor interaction. [401]

ID	Receptor
path 04060	Cytokine-cytokine receptor interaction
path 04061	Viral protein interaction with cytokine and cytokine receptor
path 04080	Neuroactive ligand-receptor interaction
path 04512	ECM-receptor interaction
path 04620	Toll-like receptor signaling pathway
path 04621	NOD-like receptor signaling pathway
path 04622	RIG-I-like receptor signaling pathway
path 04625	C-type lectin receptor signaling pathway
path 04660	T cell receptor signaling pathway
path 04662	B cell receptor signaling pathway
path 05207	Chemical carcinogenesis - receptor activation
path 07211	Serotonin receptor agonists/antagonists
path 07212	Histamine H1 receptor antagonists
path 07213	Dopamine receptor agonists/antagonists
path 07214	beta-Adrenergic receptor agonists/antagonists
path 07215	alpha-Adrenergic receptor agonists/antagonists
path 07221	Nicotinic cholinergic receptor antagonists
path 07222	Peroxisome proliferator-activated receptor (PPAR) agonists
path 07223	Retinoic acid receptor (RAR) and retinoid X receptor (RXR) agonists/antagonists
path 07224	Opioid receptor agonists/antagonists
path 07225	Glucocorticoid and mineralocorticoid receptor agonists/antagonists
path 07226	Progesterone, androgen and estrogen receptor agonists/antagonists
path 07227	Histamine H2/H3 receptor agonists/antagonists
path 07228	Eicosanoid receptor agonists/antagonists
path 07229	Angiotensin receptor and endothelin receptor antagonists
path 07230	GABA-A receptor agonists/antagonists
path 07235	N-Methyl-D-aspartic acid (NMDA) receptor antagonists

Carcinogenesis is a multistage process of (1) initiation, (2) promotion, and (3) progression stages. Chemicals or environmental factors can induce and/or enhance the carcinogenic process of carcinogens (a) genotoxic or (b) non-genotoxic agents. Genotoxic agent begin with deoxyribonucleic acid (DNA) generating DNA damage while Non-genotoxic carcinogens are chemicals for tumors through multiple non-genotoxic events and epigenetic alterations without direct interaction with DNA such as receptor carcinogenesis activation (i) cell surface receptors and some intracellular receptors for activation of signal transduction pathways like gene transcription, and (ii) intracellular receptors into the nucleus with transcription factors regulating gene expression. [401]

*

Email address: http://mathlearningspace.weebly.com/ (Jeff Cromwell, PhD)

Figure 1 has the visual representation of Chemical carcinogenesis - receptor activation with ERK, PI3K and JAK-STAT signaling.

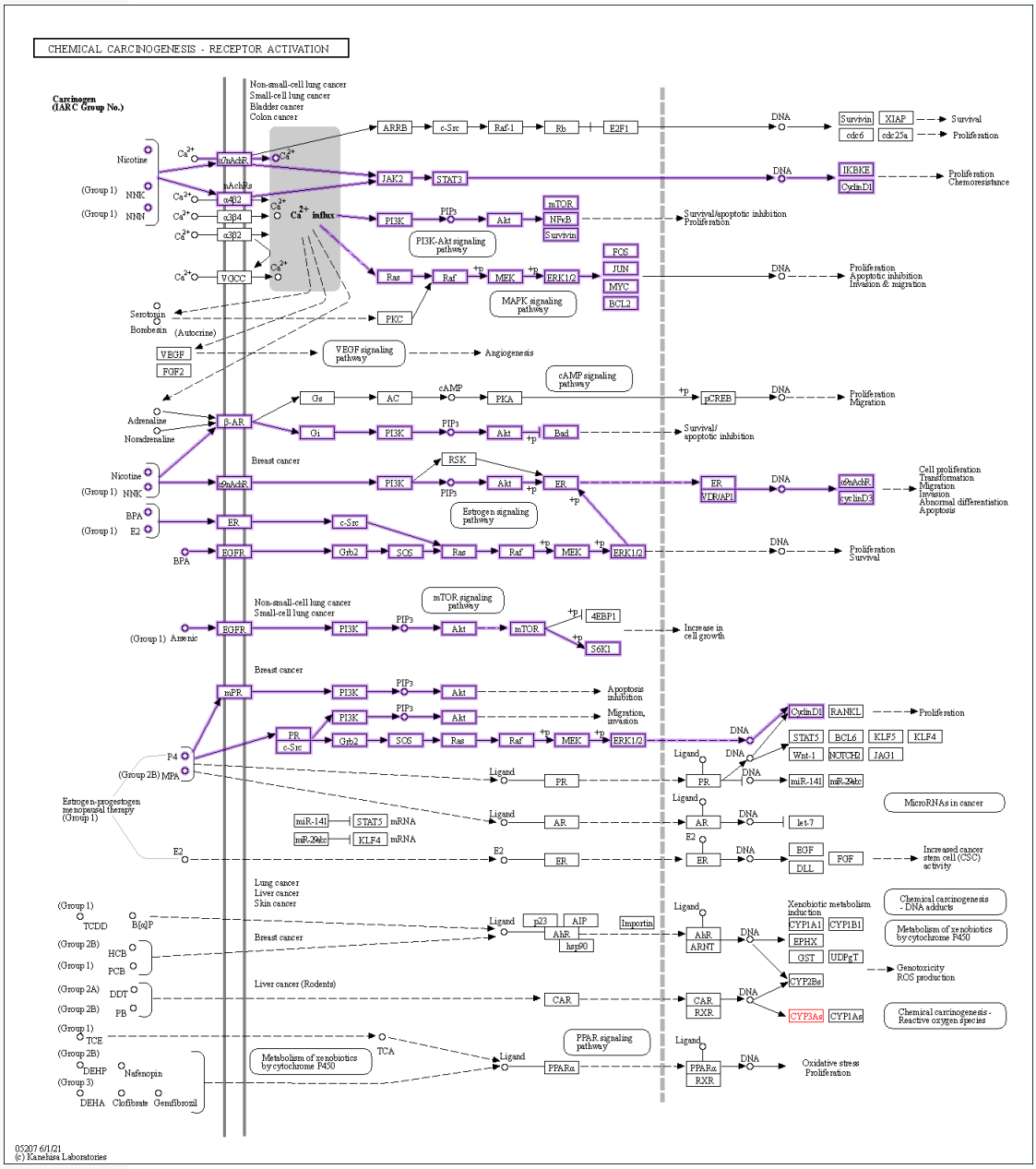

Figure 1: Chemical carcinogenesis - receptor activation [401]

The M-phase inducer phosphatase 1 [EC:3.1.3.48] has pathways such as (a) map04110 Cell cycle, (b) map04218 Cellular senescence, (c) map04914 Progesterone-mediated oocyte maturation (d) map05206 MicroRNAs in cancer and (e) map05207 Chemical carcinogenesis - receptor activation. For Cellular Processes such as Cell growth and death, the cell cycle has CDC25A; M-phase inducer phosphatase 1 as cellular senescence. For Organismal Systems such as the Endocrine system with Progesterone-mediated oocyte maturation CDC25A; M-phase inducer phosphatase 1 is involved. For Human Diseases such as Cancer, MicroRNAs in cancer and Chemical carcinogenesis - receptor activation has CDC25A; M-phase inducer phosphatase 1. [401]

Protein families: metabolism such as protein phosphatases and associated proteins with (a) genetic information processing (b) Chromosome and associated proteins and (c) DNA repair and recombination proteins CDC25A; M-phase inducer phosphatase 1 is present. These hydrolases acting on ester bonds is an Phosphoric-monoester hydrolases with protein-tyrosine-phosphatase such as CDC25A; M-phase inducer phosphatase 1. These Protein phosphatases and associated proteins as Protein tyrosine phosphatases (PTPs) are Class III PTPs CDC25s. The chromosome and associated proteins of the eukaryotic type has Centrosome formation and ciliogenesis proteins specifically Centrosome duplication proteins K06645 CDC25A; M-phase inducer phosphatase 1 and is involved in DNA repair with recombination proteins Eukaryotic type and other check point factors.[401]

Table 2 has the genes for this CDC25A M-phase inducer phosphatase 1 [EC:3.1.3.48].

1	HSA: 993(CDC25A)	ORO: 101369809(CDC25A)	HAI: 109375294(CDC25A)	ETL: 114069284(CDC25A)
2	PTR: 460341(CDC25A)	ELK: 111144080	DRO: 112309765(CDC25A)	FPG: 101911899(CDC25A)
3	PPS: 100985899(CDC25A)	MPUF: 101680806(CDC25A)	SHON: 118976202(CDC25A)	FCH: 102054923(CDC25A)
4	GGO: 101142222(CDC25A)	EJU: 114225128(CDC25A)	AJM: 119038075(CDC25A)	CLV: 102084553(CDC25A)
5	PON: 100460228(CDC25A)	MLX: 118005263(CDC25A)	PDIC: 114501673(CDC25A)	EGZ: 104121808(CDC25A)
6	NLE: 100588003(CDC25A)	FCA: 101098445(CDC25A)	MMF: 118633065(CDC25A)	NNI: 104019237(CDC25A)
7	MCC: 710858(CDC25A)	PYU: 121011146(CDC25A)	RFQ: 117037274(CDC25A)	ACUN: 113476811(CDC25A)
8	MCF: 102137696(CDC25A)	PBG: 122487215(CDC25A)	PALE: 102897196(CDC25A)	PADL: 103916196(CDC25A)
9	CSAB: 103228635(CDC25A)	PTG: 102957057(CDC25A)	PGIG: 120582298(CDC25A)	AAM: 106499367
10	CATY: 105573640(CDC25A)	PPAD: 109247747(CDC25A)	RAY: 107513304(CDC25A)	AROW: 112977410(CDC25A)
11	PANU: 101017438(CDC25A)	AJU: 106979176(CDC25A)	MJV: 108402658(CDC25A)	NPD: 112951558(CDC25A)
12	RRO: 104671299(CDC25A)	HHV: 120229362(CDC25A)	TOD: 119259478(CDC25A)	DNE: 112993598(CDC25A)
13	RBB: 108517663(CDC25A)	BTA: 520188(CDC25A)	LAV: 100659991(CDC25A)	ASN: 102384816(CDC25A)
14	TFN: 117080687(CDC25A)	BOM: 102285046(CDC25A)	TMU: 101346532	AMJ: 102569214(CDC25A)
15	PTEH: 111555493(CDC25A)	BIU: 109576463(CDC25A)	MDO: 100018150(CDC25A)	CPOO: 109305787(CDC25A)
16	CJC: 100408832(CDC25A)	BBUB: 102415490(CDC25A)	GAS: 123230824(CDC25A)	GGN: 109291133(CDC25A)
17	SBQ: 101041208(CDC25A)	CHX: 102181731(CDC25A)	SHR: 100924467(CDC25A)	PSS: 102458852(CDC25A)
18	MMUR: 105874913(CDC25A)	OAS: 101113178(CDC25A)	PCW: 110217106(CDC25A)	CMY: 102931776(CDC25A)
19	MMU: 12530(Cdc25a)	ODA: 120868201(CDC25A)	OAA: 100085311(CDC25A)	CPIC: 101944298(CDC25A)
20	MCAL: 110301556(Cdc25a)	CCAD: 122424249(CDC25A)	GGA: 420375(CDC25A)	TST: 117871895(CDC25A)
21	MPAH: 110327415(Cdc25a)	SSC: 100737380(CDC25A)	PCOC: 116238507(CDC25A)	CABI: 116828404(CDC25A)
22	RNO: 171102(Cdc25a)	CFR: 102509600(CDC25A)	MGP: 100546502(CDC25A)	ACS: 100558641(cdc25a) 103280245
23	MCOC: 116072801(Cdc25a)	CBAI: 105083373(CDC25A)	CJO: 107308656(CDC25A)	PVT: 110082540(CDC25A)
24	MUN: 110565748(Cdc25a)	CDK: 105103857(CDC25A)	NMEL: 110390717(CDC25A)	SUND: 121935243(CDC25A)
25	CGE: 100767961(Cdc25a)	BACU: 103016547(CDC25A)	APLA: 101800901(CDC25A)	PBI: 103059140(CDC25A)
26	PLEU: 114705236(Cdc25a)	LVE: 103085456(CDC25A)	ACYG: 106046801(CDC25A)	TSR: 106549893(CDC25A)
27	NGI: 103750230(Cdc25a)	OOR: 101282274(CDC25A)	TGU: 100231363(CDC25A)	PGUT: 117673238
28	HGL: 101699133(Cdc25a)	DLE: 111174440(CDC25A)	LSR: 110475093 110476932	VKO: 123017889(CDC25A)
29	CCAN: 109699005(Cdc25a)	PCAD: 102992519(CDC25A)	SCAN: 103813021(CDC25A)	PMUA: 114607182(CDC25A)
30	OCU: 100354280(CDC25A)	PSIU: 116762728(CDC25A)	PMOA: 120510615(CDC25A)	ZVI: 118092194
31	OPI: 101519009(CDC25A)	ECB: 100064223(CDC25A)	OTC: 121337999(CDC25A)	GJA: 107124887(CDC25A)
32	TUP: 102482369(CDC25A)	EPZ: 103567910(CDC25A)	PRUF: 121357168(CDC25A)	XLA: 398141(cdc25a.L) 734734(cdc25a.S)
33	CFA: 484780(CDC25A)	EAI: 106821782(CDC25A)	GFR: 102040683(CDC25A)	XTR: 733851(cdc25a)
34	VVP: 112912171(CDC25A)	MYB: 102240322(CDC25A) 102252775 102252873	FAB: 101816817(CDC25A)	NPR: 108800705
35	VLG: 121494228(CDC25A)	MYD: 102760761(CDC25A)	PHI: 102104005(CDC25A)	TPRE: 106651124
36	AML: 100477247(CDC25A)	MMYO: 118668522(CDC25A)	PMAJ: 107201060(CDC25A)	HHAL: 106685042
37	UMR: 103675112(CDC25A)	MNA: 107529418(CDC25A)	CCAE: 111925581(CDC25A)	VDE: 111245511
38	UAH: 113256823(CDC25A)	PKL: 118715710(CDC25A)	CCW: 104686374(CDC25A)	VJA: 111260799

Table 3 has the genes for enzyme 3.1.3.48. [401]

1		DUSP9	PTPN3	PTPRS
2	NEDD8-MDP1	EYA4	PTPN4	PTPRZ1
3	PTPRU	EYA1	PTPN6	DUSP21
4	CDKN3	EYA2	PTPN7	PTP4A1
5	DUSP14	EYA3	PTPN9	DUSP26
6	PTPN21	PTPN23	PTPN11	EPM2A
7	PTPRT	PTPN20	PTPN12	PTP4A2
8	PTP4A3	PTPN22	PTPN14	DUSP16
9	DUSP10	PTPN18	PTPRA	DUSP11
10	DUSP12	PGP	PTPRB	PTPN5
11	PTPMT1	DUSP28	PTPRC	UBASH3B
12	DUSP15	DUSP29	PTPRD	SSH2
13	DUSP19	PTPRQ	PTPRE	CDC14B
14	MDP1	DUSP13	PTPRF	CDC14A
15	DUSP18	ACP1	PTPRG	MTMR3
16	DUSP1	UBASH3A	PTPRH	MTMR4
17	DUSP2	SSH1	PTPRJ	STYXL2
18	DUSP3	DUSP23	PTPRK	CDC25A
19	DUSP4	SSH3	PTPRM	CDC25B
20	DUSP5	DUSP22	PTPRN	CDC25C
21	DUSP6	PTEN	PTPRN2	
22	DUSP7	PTPN1	PTPRO	NEDD8-MDP1
23	DUSP8	PTPN2	PTPRR	PTPRU

3 Results

Molecular properties are abundant in dimensional reduction for collections of sequences. Examples such as stability, binding potential aliphatic and hydrophobicity along with the charge at different pH is a few of available molecular properties to examine based on the gene ontology ids in Table 4. The net charge of a protein sequence based on the Henderson-Hasselbalch equation based on pH 5, 7 and 9. The aliphatic index is the relative volume occupied by aliphatic side chains (Alanine, Valine, Isoleucine, and Leucine) and is a positive factor for the increase of thermostability of globular proteins. The potential protein interaction index proposed by Boman (2003) based in the amino acid sequence of a protein and provides an overall estimate of the potential of a peptide to bind to membranes or other proteins as receptors. A protein have high binding potential if the index value is higher than 2.48. This index predicts the stability of a protein based on its amino acid composition, a protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. Hydrophobicity is an important stabilization force in protein folding; this force changes depending on the solvent in which the protein is found. [1001]

	Protein	Stability Index	Binding Potential	ALiphatic	f.1	CpH5	CpH7	CpH9
1	NEDD8-MDP1	60.7	2.138	90.4	-0.4917	3.844	-1.307	-3.952
2	PTPRU	45.5	1.718	78.7	-0.3351	47.091	-2.142	-43.863
3	CDKN3	59.5	1.893	87.4	-0.3127	5.303	-2.690	-14.485
4	DUSP14	37.8	1.189	92.1	-0.0121	18.016	10.781	4.553
5	PTPN21	59.3	2.087	77.4	-0.5717	51.151	11.233	-14.317
6	PTPRT	37.3	1.656	81.6	-0.3348	39.583	-2.163	-38.394
7	PTP4A3	39.5	1.638	81.7	-0.2676	15.551	10.005	3.063
8	DUSP10	57.2	1.563	84.6	-0.2878	16.963	4.222	-10.701
9	DUSP12	45.6	1.387	81.5	-0.2159	8.645	-0.648	-12.103
10	PTPMT1	42.4	1.596	97.5	-0.1607	18.486	12.335	7.259
11	DUSP15	45.4	1.569	80.1	-0.2437	18.990	8.969	-4.792
12	DUSP19	51.8	1.657	86.7	-0.2530	4.438	-1.226	-7.939
13	MDP1	60.0	1.853	84.7	-0.3426	4.036	-1.834	-4.867
14	DUSP18	46.6	0.971	93.8	0.1037	8.120	1.512	-5.118
15	DUSP1	47.2	1.160	87.8	0.0981	11.732	1.035	-11.625
16	DUSP2	59.4	1.508	92.6	0.0111	12.824	4.288	-8.010
17	DUSP3	26.0	1.727	86.5	-0.3081	5.649	1.586	-3.419
18	DUSP4	61.3	1.557	83.9	-0.1297	13.849	1.953	-13.175
19	DUSP5	62.0	1.423	86.9	-0.0974	20.182	8.973	-5.681
20	DUSP6	52.2	1.674	86.2	-0.2635	-7.564	-15.299	-24.515
21	DUSP7	51.5	1.483	76.5	-0.3010	5.848	-5.895	-17.435
22	DUSP8	67.1	1.563	74.6	-0.3274	17.559	7.040	-6.383
23	DUSP9	67.9	1.585	87.9	-0.2255	5.086	-3.821	-13.404
24	EYA4	51.8	1.692	66.9	-0.4629	-2.600	-15.833	-27.481
25	EYA1	54.4	1.761	64.1	-0.5157	8.428	-6.756	-16.175
26	EYA2	49.1	1.679	71.3	-0.4245	11.431	-4.637	-17.111
27	EYA3	54.3	1.713	73.4	-0.4492	-2.330	-16.523	-26.431
28	PTPN23	60.4	1.525	80.8	-0.4290	52.137	-2.154	-31.872
29	PTPN20	38.8	2.010	78.7	-0.4660	4.439	-8.061	-18.398
30	PTPN22	48.0	2.020	68.9	-0.6089	25.114	4.221	-13.656
31	PTPN18	48.9	1.763	76.6	-0.3641	17.354	6.597	-5.397
32	PGP	35.6	1.182	93.7	0.0234	3.235	-1.761	-9.569
33	DUSP28	66.7	1.198	81.1	-0.0625	6.537	3.287	-2.555
34	DUSP29	50.7	2.310	75.8	-0.6214	4.048	-4.409	-9.901
35	PTPRQ	36.0	1.429	85.4	-0.2428	16.137	-31.681	-67.750
36	DUSP13	38.0	1.240	95.5	-0.0117	9.348	1.757	-4.938
37	ACP1	51.5	2.158	72.8	-0.4918	3.871	-0.534	-8.458
38	UBASH3A	43.8	1.612	80.6	-0.2775	26.272	5.712	-16.465
39	SSH1	67.7	2.028	72.8	-0.5931	20.609	-12.965	-35.703
40	DUSP23	41.8	1.463	91.1	-0.1293	8.293	3.066	-2.084
41	SSH3	69.6	2.098	77.4	-0.5815	-0.558	-22.794	-33.185
42	DUSP22	57.7	1.699	87.0	-0.3114	11.872	3.762	-3.032
43	PTEN	44.2	2.242	67.0	-0.6896	10.307	-5.168	-18.500
44	PTPN1	44.6	2.029	71.7	-0.6021	9.913	-6.153	-18.719
45	PTPN2	52.2	2.256	74.9	-0.6405	18.964	6.206	-3.493
46	PTPN3	47.9	1.989	78.6	-0.4860	36.113	0.609	-26.027
47	PTPN4	48.3	2.065	74.7	-0.5600	37.196	5.436	-19.431
48	PTPN6	44.9	2.147	74.2	-0.6871	25.797	5.140	-7.497
49	PTPN7	54.9	1.827	80.4	-0.4825	12.017	-1.349	-11.791
50	PTPN9	40.8	1.695	85.4	-0.3383	25.118	6.565	-7.333
51	PTPN11	43.1	2.330	71.1	-0.7354	25.553	2.539	-12.655
52	PTPN12	49.0	2.357	67.0	-0.7655	6.425	-21.859	-41.271
53	PTPN14	54.7	2.020	78.2	-0.5395	67.907	19.153	-8.100
54	PTPRA	35.8	1.781	79.1	-0.3796	16.550	-3.082	-21.691
55	PTPRB	34.5	1.767	82.4	-0.3708	69.982	11.841	-26.850
56	PTPRC	41.7	2.012	71.9	-0.5932	24.955	-21.845	-56.406
57	PTPRD	39.1	1.760	79.1	-0.4187	41.566	-10.560	-41.415
58	PTPRE	35.2	1.672	84.9	-0.3419	22.544	0.206	-16.949
59	PTPRF	43.9	1.717	78.8	-0.3682	31.488	-16.091	-47.169
60	PTPRG	43.7	1.893	72.3	-0.4886	33.787	-15.168	-45.604
61	PTPRH	34.4	1.719	72.2	-0.4279	-1.438	-28.501	-51.900
62	PTPRJ	41.2	1.535	78.2	-0.3168	6.739	-26.325	-53.558
63	PTPRK	44.7	1.595	80.6	-0.3157	17.865	-28.370	-69.860
64	PTPRM	38.1	1.729	76.8	-0.3834	36.189	-7.737	-45.969
65	PTPRN	56.0	1.542	85.4	-0.3029	32.722	1.507	-20.572
66	PTPRN2	56.6	1.889	81.3	-0.4545	11.822	-20.855	-40.925
67	PTPRO	48.0	1.303	83.3	-0.1735	15.882	-18.521	-44.357
68	PTPRR	49.0	1.472	97.4	-0.1791	32.695	11.432	-5.472
69	PTPRS	42.3	1.742	77.5	-0.3982	35.937	-12.116	-43.888
70	PTPRZ1	45.5	1.659	78.8	-0.3664	-47.877	-124.079	-163.683
71	DUSP21	39.0	1.486	92.4	-0.0189	8.210	3.883	0.165
72	PTP4A1	39.1	1.766	86.8	-0.3006	12.732	7.766	1.091
73	DUSP26	52.4	1.745	89.7	-0.2223	19.519	9.541	3.514
74	EPM2A	47.1	1.345	79.8	-0.2350	10.384	-2.387	-13.576
75	PTP4A2	42.6	1.779	82.9	-0.3497	9.805	4.737	-2.653
76	DUSP16	59.4	1.772	79.0	-0.3707	24.423	3.836	-15.489
77	DUSP11	61.5	2.670	60.2	-1.0185	34.860	18.138	7.269
78	PTPN5	55.2	1.608	80.9	-0.3642	-11.590	-30.840	-48.348
79	UBASH3B	45.4	1.456	81.0	-0.2744	20.840	-0.782	-22.827
80	SSH2	63.9	2.164	71.1	-0.6370	4.759	-54.985	-92.621
81	CDC14B	43.7	2.175	72.7	-0.5131	34.452	19.699	3.412
82	CDC14A	49.2	2.037	69.4	-0.5342	38.035	20.516	5.387
83	MTMR3	59.8	2.024	75.4	-0.4927	15.106	-32.199	-78.993
84	MTMR4	55.1	1.884	71.5	-0.4786	19.775	-24.155	-75.780
85	STYXL2	71.3	2.933	59.7	-0.9433	-9.366	-44.562	-60.285
86	CDC25A	59.8	2.305	66.8	-0.7277	18.783	-0.485	-15.414
87	CDC25B	71.1	2.316	71.7	-0.6047	13.741	-5.712	-20.119
88	CDC25C	49.4	2.104	75.8	-0.5907	13.751	-1.552	-17.789

Table 5 has the descriptive statistics for Table 4.

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Stability Index	1	88.00	49.63	9.77	48.62	49.27	9.91	26.00	71.30	45.30	0.26	-0.56	1.04
Binding Potential	2	88.00	1.78	0.34	1.73	1.77	0.31	0.97	2.93	1.96	0.47	0.65	0.04
ALiphatic	3	88.00	79.60	8.10	79.11	79.57	8.93	59.75	97.51	37.76	0.00	-0.27	0.86
f.1	4	88.00	-0.38	0.21	-0.37	-0.38	0.18	-1.02	0.10	1.12	-0.13	0.34	0.02
CpH5	5	88.00	17.09	16.85	15.33	16.39	13.77	-47.88	69.98	117.86	0.22	2.66	1.80
CpH7	6	88.00	-5.22	18.95	-1.27	-2.90	9.77	-124.08	20.52	144.60	-3.12	15.79	2.02
CpH9	7	88.00	-22.88	26.14	-15.45	-19.22	15.75	-163.68	7.27	170.95	-2.28	8.15	2.79

Based on the maximum likelihood estimates for Table 4, the following Pearson types with moments are Stability Index has type 1 a=4.923427, b=6.628805, location=20.12304 and scale=69.28884. Binding Potential has type=4, m=7.137992, nu=-5.157446, location=1.336375, and scale=1.051895. ALiphatic has type 1, a=7.820113, b=7.317297, location=46.12254, scale=64.81443. f.1 has type=4, m=6.609686, nu=0.5796581, location=-0.3484698, scale=0.6799085. CpH5 has type 4, m=1.987773, nu=-1.178801, location=7.396149, and scale=17.46834. CpH7 has type 4, m=1.847313, nu=1.826933, location=8.64192 and scale=12.959 CpH9 has type 5 with shape=4.602605, location=21.80742 and scale=-161.5185.

Figure 2 has Cluster Dendrogram for Table 4.

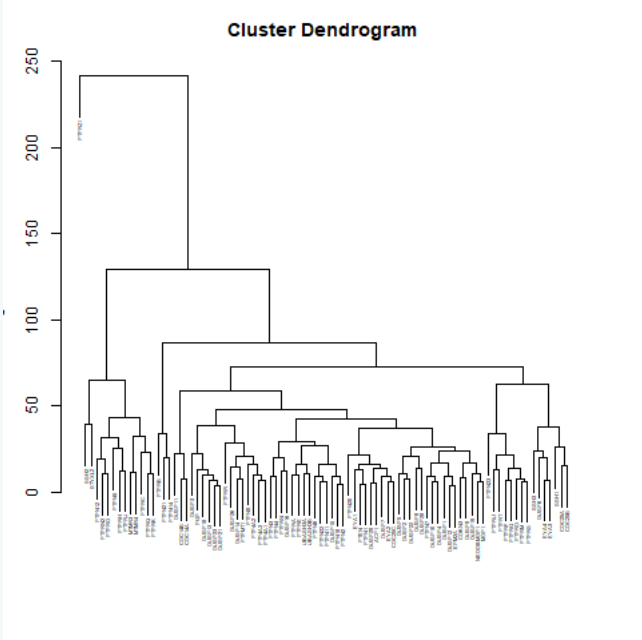

Figure 2: Cluster Dendrogram for Table 4. [401]

Table 6 has the k cluster groupings.

	Stability Index	Binding Potential	ALiphatic	f.1	CpH5	CpH7	CpH9
1	48.92	1.73	81.59	-0.34	15.85	2.94	-8.89
2	43.81	1.72	78.20	-0.39	39.72	-8.52	-42.27
3	49.50	1.96	79.31	-0.49	63.01	14.08	-16.42
4	60.94	1.98	73.59	-0.53	5.73	-12.80	-26.12
5	47.44	1.76	77.06	-0.42	11.06	-25.92	-57.20
6	45.51	1.66	78.85	-0.37	-47.88	-124.08	-163.68
7	67.62	2.55	65.44	-0.79	-2.30	-49.77	-76.45

Figure 3 has the original tree with the 7 cluster tree for comparison.

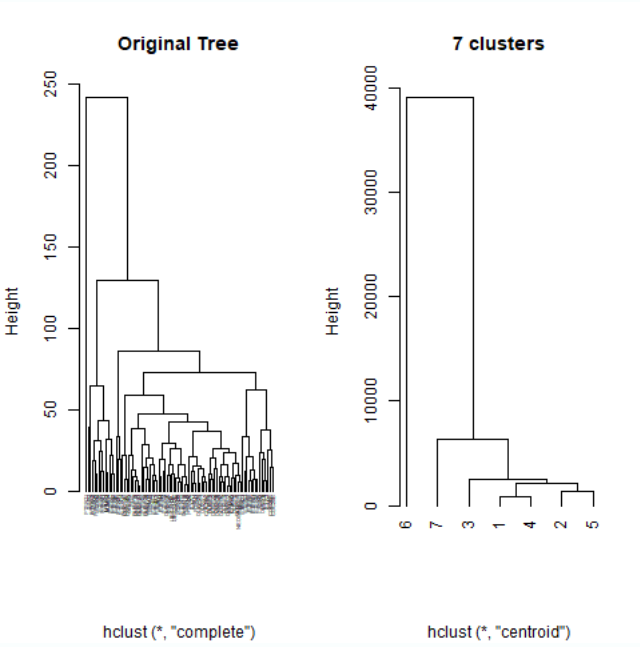

Figure 3: Cluster Dendrogram for Table 4. [401]

In the design of molecular scales for comparision, consider (a) crucianiProperties [3] (b) kideraFactors [4] (c) zScales [5] (d) FASGAI [6] (e) tScales [7] (f) VHSE [8] (g) protFP [9] (h) stScales [10] (i) BLOSUM [11] and (j) MSWHIM [1001]. The Kidera Factors are from multivariate analysis to 188 physical properties of the 20 amino acids with dimensionaltiy reduction techniques. A 10-dimensional vector of orthogonal factors where the first four factors are essentially pure physical properties; the remaining six factors are superpositions of several physical properties are presented in Table 7. [1001]

Table 7 has the kidera factors for the genes.

	Names	HBF	SCS	ESP	H	DBP	PSV	FEP	OAR	PKC	SH
1	NEDD8-MDP1	-133	-121	65	224	-132	-277	108	-29	-3	-71
2	PTPRU	-34	-200	30	75	-144	-219	25	-64	2	39
3	CDKN3	-22	-205	48	103	-190	-184	-15	-138	67	54
4	DUSP14	-34	-192	163	-7	-62	-227	-41	-36	-23	32
5	PTPN21	-12	-144	1	147	-99	-267	-1	-174	-3	24
6	PTPRT	6	-187	88	89	-89	-236	-16	-53	-13	-16
7	PTP4A3	-90	-162	84	111	-34	-279	75	-109	11	175
8	DUSP10	18	-283	44	100	-30	-260	5	-127	24	-53
9	DUSP12	-122	-241	-63	56	-91	-282	92	-104	-60	48
10	PTPMT1	-173	-149	155	112	-139	-333	91	-92	-33	91
11	DUSP15	12	-335	26	55	-138	-179	77	-214	30	25
12	DUSP19	-87	-221	69	136	-107	-288	-33	8	-2	-50
13	MDP1	-61	-123	-4	91	-123	-315	30	-83	54	34
14	DUSP18	-92	-214	110	-35	36	-215	-138	-127	20	-46
15	DUSP1	-96	-379	-9	-13	-184	-299	-27	-52	29	38
16	DUSP2	-114	-295	-26	-8	-294	-286	111	-133	82	135
17	DUSP3	-44	-241	1	114	-80	-307	33	27	48	-57
18	DUSP4	-35	-308	12	46	-188	-261	-5	-128	-14	115
19	DUSP5	-84	-291	52	52	-144	-294	9	-131	-17	90
20	DUSP6	32	-227	-45	60	-146	-314	-26	-77	39	-19
21	DUSP7	38	-350	-110	47	-135	-315	16	-83	-23	3
22	DUSP8	117	-395	-166	28	-162	-345	68	-223	-28	42
23	DUSP9	4	-288	-120	19	-243	-367	97	-143	4	90
24	EYA4	117	-313	-1	105	-53	-275	-75	-151	-116	-48
25	EYA1	182	-297	15	91	-64	-246	-91	-120	-87	-85
26	EYA2	113	-264	5	73	-85	-275	-55	-59	-46	5
27	EYA3	32	-292	17	139	-46	-277	-103	-167	-99	-82
28	PTPN23	3	-257	-163	48	-11	-355	43	-306	38	9
29	PTPN20	-48	-93	11	136	-115	-234	-49	-23	46	-51
30	PTPN22	34	-126	-33	166	-21	-306	-93	-200	-11	-50
31	PTPN18	-8	-280	-5	91	-172	-242	115	-76	-76	58
32	PGP	-93	-401	-59	14	-226	-373	160	6	-37	133
33	DUSP28	-151	-466	-175	1	-208	-325	186	-183	-31	63
34	DUSP29	-123	-107	-73	184	-164	-269	11	-2	1	125
35	PTPRQ	59	-195	136	78	-65	-291	-97	-101	-11	-98
36	DUSP13	-86	-279	-5	-17	-166	-336	8	-53	28	118
37	ACP1	-62	-148	40	166	-126	-134	10	-21	84	-34
38	UBASH3A	-66	-195	-46	54	-152	-294	51	-162	30	53
39	SSH1	-9	-247	-156	149	-105	-315	-29	-222	-31	26
40	DUSP23	-40	-245	-17	-5	-189	-289	168	-75	80	128
41	SSH3	-90	-251	-92	164	-152	-240	26	-185	-88	-37
42	DUSP22	-165	-150	-18	97	-121	-256	52	19	38	40
43	PTEN	15	-22	-38	167	-27	-252	-25	21	140	48
44	PTPN1	-65	-98	-104	141	-73	-238	17	-94	3	28
45	PTPN2	-78	-40	17	194	-85	-216	41	-45	-19	-20
46	PTPN3	-20	-140	45	142	-43	-228	-51	-74	20	21
47	PTPN4	19	-117	38	146	-29	-216	-14	-84	49	-31
48	PTPN6	-27	-143	26	216	-52	-249	36	31	-76	-14
49	PTPN7	-67	-184	1	121	-92	-192	88	-154	-27	-12
50	PTPN9	-80	-105	47	98	-74	-281	66	-49	26	-1
51	PTPN11	-38	-109	45	243	-39	-212	28	24	-53	47
52	PTPN12	47	-161	-42	207	-78	-191	-90	-171	35	-57
53	PTPN14	-58	-133	33	159	-34	-233	-28	-119	-7	6
54	PTPRA	-27	-168	103	136	-77	-217	-25	-55	-20	-71
55	PTPRB	35	-188	119	123	-77	-287	-89	-64	-52	-1
56	PTPRC	-10	-166	-1	174	-45	-242	-58	-83	40	-45
57	PTPRD	6	-186	77	129	-62	-248	-22	-94	-48	-24
58	PTPRE	-109	-94	47	102	-46	-230	59	-40	29	-50
59	PTPRF	-16	-215	52	108	-100	-250	-1	-125	-75	15
60	PTPRG	10	-198	29	138	-67	-214	-82	-56	-42	-33
61	PTPRH	78	-295	85	105	-142	-195	-29	-93	-165	-10
62	PTPRJ	82	-296	108	103	-84	-258	-91	-60	-71	-49
63	PTPRK	-9	-186	38	62	-94	-201	-18	-49	27	-23
64	PTPRM	24	-182	89	101	-74	-207	-19	-43	-11	-16
65	PTPRN	-24	-320	-74	61	-112	-317	52	-212	-35	45
66	PTPRN2	-43	-274	-119	114	-159	-326	7	-173	-35	84
67	PTPRO	-26	-143	81	35	-35	-287	-100	-113	-23	-33
68	PTPRR	-99	-181	102	93	-25	-287	-39	-103	20	-53
69	PTPRS	14	-217	42	106	-96	-250	14	-107	-48	-4
70	PTPRZ1	7	-271	4	119	-62	-292	-178	-114	-45	-44
71	DUSP21	-34	-170	153	31	-84	-244	-174	-53	46	-112
72	PTP4A1	-102	-111	81	115	-91	-240	131	20	72	63
73	DUSP26	-118	-163	35	51	-187	-256	48	-32	1	166
74	EPM2A	-86	-207	-30	8	-125	-188	98	42	-73	83
75	PTP4A2	-81	-110	15	95	-94	-250	166	-30	106	121
76	DUSP16	-13	-262	-17	128	-73	-325	-98	-207	-45	16
77	DUSP11	140	19	-14	193	-35	-187	21	-54	77	24
78	PTPN5	-20	-191	-70	44	-119	-208	7	-200	-20	33
79	UBASH3B	-22	-199	-6	41	-82	-256	40	-125	8	2
80	SSH2	-37	-230	-94	183	-77	-220	-79	-194	-27	17
81	CDC14B	-4	-134	77	167	-136	-220	-8	-7	57	-19
82	CDC14A	67	-184	-0	140	-113	-304	-10	-10	49	-12
83	MTMR3	-37	-227	-74	121	-154	-209	-57	-180	0	42
84	MTMR4	23	-219	-74	85	-97	-195	-16	-170	24	49
85	STYXL2	-97	-200	-127	332	-256	-237	-105	-98	-172	34
86	CDC25A	44	-153	-135	176	-67	-288	18	-136	14	108
87	CDC25B	-30	-193	-124	163	-170	-264	-10	-203	5	92
88	CDC25C	1	-159	-70	165	-59	-282	79	-89	7	88

Figure 4 has the cluster dendrogram from Table 7 of kidera factors.

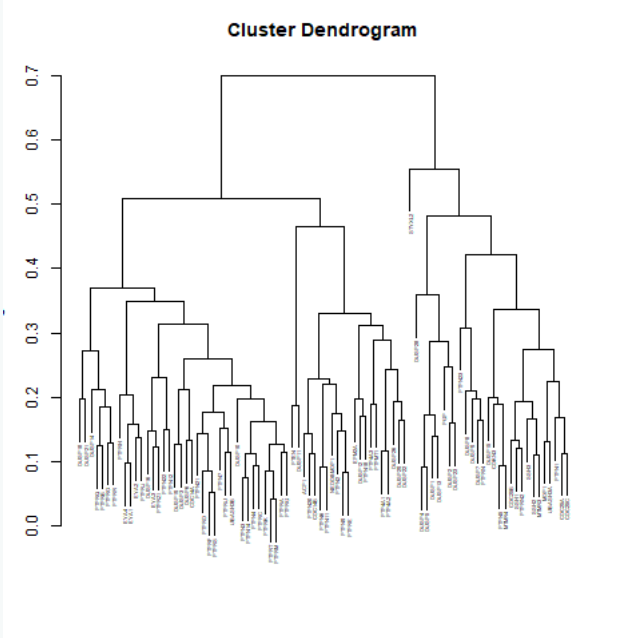

Figure 4: Cluster Dendrogram for Table 7 of kidra factors. [1001]

4 Conclusions

Several molecular properties are examined with respect to these two categories in carcinogenesis. Specifically, multi-species genes were examined for Hydrolases acting on ester bonds as an Phosphoric-monoester hydrolases with protein-tyrosine-phosphatase such as CDC25A; M-phase inducer phosphatase 1. These Protein phosphatases and associated proteins of Protein tyrosine phosphatases (PTPs) are Class III PTPs CDC25s were examined. In addition, genes for enzyme 3.1.3.48 were presented and sequences analyzed with both molecular properties and descriptive statistics with maximum likelihood estimation techniques for Pearson distribution identification. A cluster dendrogram of Kidra factors was also presented.

5 Bibliography

[2] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. (2000) The Protein Data Bank Nucleic Acids Research, 28: 235-242.

[3] Cruciani, G., Baroni, M., Carosati, E., Clementi, M., Valigi, R., and Clementi, S. (2004) Peptide studies by means of principal properties of amino acids derived from MIF descriptors. J. Chemom. 18, 146-155.

[4] Kidera, A., Konishi, Y., Oka, M., Ooi, T., and Scheraga, H. A. (1985). Statistical analysis of the physical properties of the 20 naturally occurring amino acids. Journal of Protein Chemistry, 4(1), 23-55.

[5] Sandberg M, Eriksson L, Jonsson J, Sjostrom M, Wold S: New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. J Med Chem 1998, 41:2481-2491.

[6] Liang, G., and Li, Z. (2007). Factor analysis scale of generalized amino acid information as the source of a new set of descriptors for elucidating the structure and activity relationships of cationic antimicrobial peptides. Molecular Informatics, 26(6), 754-763.

[7] Tian F, Zhou P, Li Z: T-scale as a novel vector of topological descriptors for amino acids and its application in QSARs of peptides. J Mol Struct. 2007, 830: 106-115. 10.1016/j.molstruc.2006.07.004.

[8] van Westen, G. J., Swier, R. F., Wegner, J. K., IJzerman, A. P., van Vlijmen, H. W., and Bender, A. (2013). Benchmarking of protein descriptor sets in proteochemometric modeling (part 1): comparative study of 13 amino acid descriptor sets. Journal of cheminformatics, 5(1), 41.

[9] Yang, L., Shu, M., Ma, K., Mei, H., Jiang, Y., and Li, Z. (2010). ST-scale as a novel amino acid descriptor and its application in QSAM of peptides and analogues. Amino acids, 38(3), 805-816.

[10] Georgiev, A. G. (2009). Interpretable numerical descriptors of amino acid space. Journal of Computational Biology, 16(5), 703-723.

[11] Zaliani, A., and Gancia, E. (1999). MS-WHIM scores for amino acids: a new 3D-description for peptide QSAR and QSPR studies. Journal of chemical information and computer sciences, 39(3), 525-533.

[100] Martin Becker and Stefan Klößner (2017). PearsonDS: Pearson Distribution System. R package version 1.1. https://CRAN.R-project.org/package=PearsonDS

[101] Guha, R. (2007). 'Chemical Informatics Functionality in R'. Journal of Statistical Software 6(18)

[400] Kanehisa, Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K.; KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45, D353-D361 (2017).

[401] Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M.; KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44, D457-D462 (2016).

[402] Kanehisa, M. and Goto, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27-30 (2000).

[403] Petri, V., Jayaraman, P., Tutaj, M., Hayman, G. T., Smith, J. R., De Pons, J., ... Jacob, H. J. (2014). The pathway ontology – updates and applications. Journal of Biomedical Semantics, 5, 7. http://doi.org/10.1186/2041-1480-5-7

[601] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, von Mering C. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019 Jan; 47:D607-613.PubMed

[602] Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017 Jan; 45:D362-68.PubMed

[603] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015 Jan; 43:D447-52.PubMed

[604] Franceschini A, Lin J, von Mering C, Jensen LJ. SVD-phy: improved prediction of protein functional associations through singular value decomposition of phylogenetic profiles. Bioinformatics. 2015 Nov; btv696.PubMed

[605] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013 Jan; 41:D808-15.PubMed

[606] Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 2011 Jan; 39:D561-8.PubMed

[607] Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C. STRING 8–a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 2009 Jan; 37:D412-6.PubMed

[608] von Mering C, Jensen LJ, Kuhn M, Chaffron S, Doerks T, Krueger B, Snel B, Bork P. STRING 7–recent developments in the integration and prediction of protein interactions. Nucleic Acids Res. 2007 Jan; 35:D358-62.PubMed

[609] von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen MA, Bork P. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic Acids Res. 2005 Jan; 33:D433-7.PubMed

[610] von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. 2003 Jan; 31:258-61.PubMed

[611] Snel B, Lehmann G, Bork P, Huynen MA. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res. 2000 Sep 15;28(18):3442-4.PubMed

[700] National Cancer Institute Annual Plan and Budget Proposal FY 2019 https://www.cancer.gov/about-nci/budget/plan

[800] Kim, Sunghwan et al. “PubChem in 2021: new data content and improved web interfaces.” Nucleic Acids Res. vol. 49,D1 (2021): D1388-D1395. doi:10.1093/nar/gkaa971

[1000] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[1001] Osorio, D., Rondon-Villarreal, P. and Torres, R. Peptides: A package for data mining of antimicrobial peptides. The R Journal. 7(1), 4-14 (2015).