

Table S10. Computational search for regulatory elements**A. Supervised search for vgBE-like elements**

CG# from TD list	Synonym	# Clusters	# Different Motifs	Function
CG10121	/	1	3	/
CG10336	CG10336	1	2	/
CG11822	nAcRbeta-21C	1	2	nicotinic acetylcholine-activated cation-selective channel activity; ; GO:0004889; ; GO:0004889; ; GO:0004889
CG1210	Pk61C	1	3	protein serine/threonine kinase activity; ; GO:0004674; receptor signaling protein serine/threonine kinase activity; ; GO:0004702 ; EC:2.7.1.-; ; GO:0004674; protein kinase activity; ; GO:000
CG12534	similar to ALR	1	3	flavin-linked sulfhydryl oxidase activity; ; GO:0016971; ; GO:0016971
CG12840	Tsp42E1	1	2	receptor signaling protein activity; ; GO:0005057; ; GO:0005057
CG13780	Pvf2	1	1	vascular endothelial growth factor receptor binding; ; GO:0005172; ; GO:0005172
CG14059	CG14059	1	2	/
CG14879	CG14879	1	2	/
CG15561	CG15561	1	2	/
CG17108	CG17108	1	3	acetyl-CoA carboxylase activity; ; GO:0003989 ; EC:6.4.1.2; ; GO:0003989 ; EC:6.4.1.2
CG17914	yellow-b	1	2	/
CG18290	Act87E	1	1	structural constituent of cytoskeleton; GO:0005200; ; GO:0005200
CG1874	CG1874	1	1	/
CG1994	l(1)G0020	1	2	N-acetyltransferase activity; GO:0008080 ; EC:2.3.1.-; ; GO:0008080 ; EC:2.3.1.-
CG2028	CkIalpha	1	2	casein kinase I activity; ; GO:0004681 ; EC:2.7.1.-; ; GO:0004681 ; EC:2.7.1.-; ; GO:0004681 ; EC:2.7.1.-; receptor signaling protein serine/threonine kinase activity; ; GO:0004702 ; EC:2.7
CG3198	CG3198	1	2	nuclear mRNA splicing, via spliceosome; ; GO:0000398
CG3200	Reg-2	1	3	wide-spectrum protease inhibitor activity; ; GO:0017114; ; GO:0017114
CG3568	CG3568	1	3	/
CG3830	vg	1	2	wing margin morphogenesis; ; GO:0008587; wing morphogenesis; ; GO:0007476
CG4766	CG4766	1	2	/
CG5355	CG5355	1	2	prolyl oli; GOpeptidase activity; ; GO:0004287 ; EC:3.4.21.26; ; GO:0004287 ; EC:3.4.21.26
CG5518	sda	1	3	membrane alanyl aminopeptidase activity; GO:0004179 ; EC:3.4.11.2; ; GO:0004179 ; EC:3.4.11.2
CG5836	SF1	1	1	pre-mRNA splicing factor activity; GO:0008248; ; GO:0008248; transcription cofactor activity; GO:0003712; ; GO:0008248; ; GO:0003712
CG5840	CG5840	1	1	pyrroline-5-carboxylate reductase activity; GO:0004735 ; EC:1.5.1.2; ; GO:0004735 ; EC:1.5.1.2
CG5913	CG5913	1	2	/
CG6044	CG6044	1	3	/
CG6743	Nup170	1	3	transporter activity; ; GO:0005215; ; GO:0005215
CG7539	Edg91	1	1	structural constituent of pupal cuticle (sensu Insecta); GO:0008011; ; GO:0008011
CG7583	CtBP	1	2	protein C-terminus binding; GO:0008022; ; GO:0008022; transcription corepressor activity; GO:0003714; ; GO:0003714; ; GO:0003714
CG7804	CG7804	1	1	RNA binding; ; GO:0003723; ; GO:0003723; transcription regulator activity; ; GO:0030528; ; GO:0030528

CG8171	dup	1	3	DNA binding; Mutants affect: S-phase, embryo, imaginal disc
CG8394	CG8394	1	2	gamma-aminobutyric acid transporter activity; ; GO:0005331; ; GO:0005331
CG8545	CG8545	1	2	nucleic acid binding; ; GO:0003676; ; GO:0003676
CG8805	wun2	1	3	phosphatidate phosphatase activity; ; GO:0008195; EC:3.1.3.4; ; GO:0008195; EC:3.1.3.4; ; GO:0008195; EC:3.1.3.4; ; GO:0008195; EC:3.1.3.4
CG9023	Drip	1	2	water transporter activity; ; GO:0005372; ; GO:0005372; carrier activity; ; GO:0005386; ; GO:0005386; water channel activity; ; GO:0015250; ; GO:0005372; ; GO:0015250
CG11450	net	2	1,3	RNA polymerase II transcription factor activity; interacts with rho and EGFR, wing vein patterning
CG15279	CG15279	2	3,2	cation:amino acid symporter activity; ; GO:0005416; neurotransmitter:sodium symporter activity; ; GO:0005328; ; GO:0005416; ; GO:0005328
CG17337	CG17337	2	1,2	/
CG17843	CG17843	2	3,1	flavin-linked sulfhydryl oxidase activity; GO:0016971; glucosidase activity; GO:0015926; EC:3.2.1.-; ; GO:0015926; EC:3.2.1.-; ; GO:0016971
CG2863	Nle	2	1,3	Notch signaling pathway; ; GO:0007219
CG3036	CG3036	2	1,1	sodium:phosphate symporter activity; ; GO:0005436; ; GO:0005436
CG4591	Tsp86D	2	3,2	receptor binding; GO:0005102; ; GO:0005102
CG5114	CG5114	3	3,3,1	/
CG6751	CG6751	2	2,1	general RNA polymerase II transcription factor activity; ; GO:0016251; ; GO:0016251

B. Unsupervised search for putative regulatory elements

GENE	Count of TATCGATA	Ratio in TD	TATCGATA positions upstream		Function
CG10223	1	-1.334	-37		Top2; DNA binding; ; GO:0003677; RNA binding; ; GO:0003723; ; GO:0003677; ; GO:0003723; ATPase activity; ; GO:0003723; topoisomerase (ATP-hydrolyzing) activity; ; GO:0003918; EC:5.99.1.3; ; GO:0
CG1512	1	-1.103	-20		cul-2; proteolysis and peptidolysis; ; GO:0006508; regulation of cell cycle; ; GO:0000074
CG15218	1	-1.164	-486		CycK; transcription regulator activity; ; GO:0030528; kinase activator activity; ; GO:0019209; cyclin-dependent protein kinase activity; ; GO:0016538; EC:2.7.1.-; ; GO:0030528; ; GO:0019
CG15484	1	-1.5095	-125		CG15484
CG10642	1	-1.122	-171		Klp64D; structural constituent of cytoskeleton; ; GO:0005200; microtubule motor activity; ; GO:0003777; ; GO:0003777; ; GO:0005200; ; GO:0003777
CG11255	1	-1.057	-574		/
CG5553	1	-1.04345	-45		DNAprim; nucleic acid binding; GO:0003676; DNA primase activity; GO:0003896; EC:2.7.7.-; ; GO:0003896; EC:2.7.7.-; ; GO:0003896; EC:2.7.7.-; ; GO:0003896; EC:2.7.7.-; ; GO:0003
CG5837	2	-1.558	-438	-39	Hem; receptor binding; GO:0005102; ; GO:0005102
CG6143	2	-1.064	-140	-45	/
CG14685	1	-1.318	-96		CG14685
CG1483	1	-1.245	-83		Map205; microtubule binding; ; GO:0008017; ; GO:0008017; ; GO:0008017
CG1646	2	-1.121	-38	-15	CG1646; pre-mRNA splicing factor activity; GO:0008248; ; GO:0008248
CG18572	1	-1.124	-299		/
CG4303	1	-1.062	-217		Bap60; chromatin binding; ; GO:0003682; general RNA polymerase II transcription factor activity; ; GO:0016251; ; GO:0016251
CG4954	1	-1.745	-220		eIF3-S8; translation initiation factor activity; ; GO:0003743; ; GO:0003743; ; GO:0003743

CG9253	1	-1.388	-47			CG9253; nucleic acid binding; ; GO:0003676; ; ; GO:0003676; ATP-dependent RNA helicase activity; ; GO:0004004; ; GO:0004004
CG6945	1	-1.4435	-13			CG6945
CG9128	1	-1.138	-269			Sac1; polyphosphoinositide phosphatase activity; ; GO:0017120; ; ; GO:0017120
CG10353	1	-1.183	-67			CG10353
CG11190	1	-1.279	-44			CG11190; attachment of GPI anchor to protein; GO:0016255
CG5902	2	-1.093	-80	-24		CG5902; molecular_function unknown; GO:0005554; ; GO:0005554
CG1242	1	-1.614	-240			Hsp83; ATPase activity, coupled; ; GO:0042623; ; ; GO:0042623
CG7425	1	-1.325	-140			/
CG7769	1	-1.289	-209			DDB1; damaged DNA binding; GO:0003684; ; GO:0003684
CG6801	1	-1.149	-104			/
CG8188	2	-1.139	-100	-92		CG8188; ubiquitin conjugating enzyme activity; ; GO:0004840; ligase activity; ; GO:0016874 ; EC:6.-.-.-; ; ; GO:0016874 ; GO:0004840
CG6120	1	-1.03835	-131			Tsp96F; receptor binding; GO:0005102; ; GO:0005102
CG11259	2	-1.89	-2570	-56		/
CG12301	2	-1.118	-4265	-42		CG12301
CG2135	2	-1.011	-4767	-363		CG2135; beta-glucuronidase activity; ; GO:0004566 ; EC:3.2.1.31; ; ; GO:0004566 ; EC:3.2.1.31
CG5874	2	-1.176	-4605	-116		CG5874; RNA binding; GO:0003723; ; GO:0003723
CG1810	3	-1.1	-4789	-560	-13	mRNA-capping-enzyme; protein tyrosine/serine/threonine phosphatase activity; ; GO:0008138 ; EC:3.1.3.-; mRNA guanylyltransferase activity; ; GO:0004484 ; EC:2.7.7.50; ; ; GO:0008138 ; EC:3.1.3.-; ; ; GO:0004484 ; E
CG13900	3	-1.039	-623	-61	-32	CG13900; poly(A) binding; ; GO:0008143; damaged DNA binding; ; GO:0003684; ; ; GO:0003684; pre-mRNA splicing factor GO:0008248; ; ; GO:0008143; ; ; GO:0008248
CG2165	2	-1.14	-1262	-18		CG2165; calcium-transporting ATPase activity; ; GO:0005388 ; EC:3.6.3.8; ; ; GO:0005388 ; EC:3.6.3.8
CG10687	2	-1.286	-1848	-390		BcDNA:GH06451/ RNA binding; ; GO:0003723; asparagine-tRNA ligase activity; ; GO:0004816 ; EC:6.1.1.22; ; ; GO:0004816 ; EC:6.1.1.22; ; ; GO:0003723; ; ; GO:0004816 ; EC:6.1.1.22; RNA binding; ; GO:0003723; asparagine-tRNA ligase activity; ; ; EC:6.1.1.22; ; ; GO:0004816 ; EC:6.1.1.22; ; ; GO:0003723; ; ; GO:0004816 ; EC:6.1.1.22
CG13691	1	-1.2	-1996			CG13691
CG9915	1	-1.123	-1771			CG9915
CG6204	1	-1.2847	-956			CG6204; nucleic acid binding; GO:0003676; ; GO:0003676; DNA helicase activity; GO:0003678; RNA helicase activity; GO:0003678; EC:2.7.7.-; helicase activity; GO:0004386; ; GO:0003724 ; EC:2.7.7
CG2666	2	-1.045	-962	-950		kkv; chitin synthase activity; GO:0004100 ; EC:2.4.1.16; ; GO:0004100 ; EC:2.4.1.16; ; GO:0004100 ; EC:2.4.1.16; ; GO:0004100 ; EC:2.4.1.16; ; GO:0004100 ; EC:2.4.1.16
CG9191	1	-1.87	-1818			Klp61F; motor activity; ; GO:0003774; structural constituent of cytoskeleton; ; GO:0005200; microtubule motor activity; ; GO:0003777; plus-end-directed microtubule motor activity; ; GO:0003777
CG6767	1	-1.2725	-1580			/
CG12325	1	-1.425	-1987			CG12325; cell cycle; ; GO:0007049; cytokinesis; ; GO:0000910; mitosis; ; GO:0007067
CG1341	1	-1.243	-1380			Rpt1; endopeptidase activity; ; GO:0004175; ATPase activity; ; GO:0016887; ; ; GO:0016887; ; ; GO:0004175
CG1874	1	-1.0517	-1899			CG1874
CG8502	1	-1.7881	-1887			CG8502; structural constituent of larval cuticle (sensu Insecta); ; GO:0008010; ; ; GO:0008010
CG7033	2	-1.056	-2794	-1196		CG7033; ATPase activity, coupled; GO:0042623; ; GO:0042623
CG10363	1	-1.456	-4476			TepIV; protease inhibitor activity; ; GO:0030414; ; ; GO:0030414
CG10697	1	-1.3805	-3555			/
CG10954	1	-1.5	-3186			Arc-p34; actin binding; ; GO:0003779; ; ; GO:0003779; ; ; GO:0003779; structural constituent of cytoskeleton; ; GO:0005200

CG11822	1	-1.026	-2561		nAcRbeta-21C; nicotinic acetylcholine-activated cation-selective channel activity; ; GO:0004889; ; ; GO:0004889; ; ; GO:000
CG5001	1	-1.0995	-2235		CG5001; defense response; ; GO:0006952; protein folding; ; GO:0006457; response to heat; ; GO:0009408; response to str GO:0006950
CG5776	1	-1.241	-3231		CG5776; ATPase activity; ; GO:0016887; ; ; GO:0016887
CG7294	1	-1.116	-2600		CG7294
CG7296	1	-1.307	-4881		CG7296
CG9188	2	-1.26	-4220	-3581	sip2, CG9188
CG9246	1	-1.112	-4643		CG9246; molecular_function unknown; ; GO:0005554; ; ; GO:0005554
CG11073	2	-1.451	-3243	-2348	CG11073
CG3830	1	-3.682	-4023		vg; wing margin morphogenesis; ; GO:0008587; wing morphogenesis; ; GO:0007476
CG6530	1	-1.477	-3818		mthl3; G-protein coupled receptor activity; ; GO:0004930; ; ; GO:0004930; ; ; GO:0004930
CG8414	1	-1.569	-3875		CG8414
CG8811	1	-1.146	-4240		muskelin; cell adhesion; ; GO:0007155; cytoskeleton organization and biogenesis; ; GO:0007010
CG14059	1	-2.0365	-4434		/
CG17947	1	-2.06	-2135		alpha-Cat; actin binding; GO:0003779; ; GO:0003779; cytoskeletal protein binding; GO:0008092; ; GO:0008092; cadherin b GO:0045296; structural constituent of cytoskeleton; GO:0005200; ; GO:0
CG1898	1	-1.228	-3431		HBS1; translation release factor activity; ; GO:0003747; translation elongation factor activity; ; GO:0003746; ; ; GO:000374 GO:0003747
CG3424	1	-1.11605	-4385		/
CG6650	1	-1.081	-2348		CG6650
CG6897	1	-1.132	-4729		CG6897
CG11990	1	-1.058	-4250		CG11990
CG2512	1	-1.289	-3269		alphaTub84D; GTP binding; GO:0005525; ; GO:0005525; structural constituent of cytoskeleton; GO:0005200; tubulin binding GO:0015631; ; GO:0005200; ; GO:0015631
CG5913	1	-1.15885	-4977		CG5913
CG6378	2	-1.511	-4572	-3750	BM-40-SPARC; structural molecule activity; GO:0005198; growth factor activity; GO:0008083; ; GO:0005198; ; GO:0008083
CG6962	1	-1.352	-3502		CG6962
CG8064	1	-1.2695	-4227		CG8064
CG8165	1	-1.508	-2928		CG8165; transcription regulator activity; GO:0030528; ; GO:0030528
CG9373	1	-1.002	-2753		CG9373; RNA binding; GO:0003723; ; GO:0003723
CG31998	1	-1.01353	-4331		CG11572, CG11578
CG11750	1	-1.26	-3436		CG11750