

A Appendix

Contents of the appendix

Contents of the appendix	1
Appendix 1: Quorum sensing domains	2
Appendix 2: Transcription factors involved in yeast quorum sensing	8
Appendix 3: Full list of descriptors	11
Key to classes and colour-code.....	11
Table of descriptors	11
Appendix 4: List of publications.....	18
Publications relevant to this thesis	18
Other publications.....	18
References for the appendix	19

Deinococcus-Thermus												
<i>Deinococcus geothermalis</i> DSM 11300											1	
<i>Deinococcus radiodurans</i>											1	
<i>Thermus thermophilus</i> HB27											1	
<i>Thermus thermophilus</i> HB8											1	
Acidobacteria												
<i>Acidobacteria bacterium</i> Ellin345												
Firmicutes/Bacilli												
<i>Bacillus anthracis</i> Ames											1	1
<i>Bacillus anthracis</i> Ames 0581											1	1
<i>Bacillus anthracis</i> str Sterne											1	1
<i>Bacillus cereus</i> ATCC 10987											1	1
<i>Bacillus cereus</i> ATCC14579											1	1
<i>Bacillus cereus</i> ZK											1	1
<i>Bacillus clausii</i> KSM-K16											1	1
<i>Bacillus halodurans</i>				1							1	1
<i>Bacillus licheniformis</i> ATCC 14580							1		1			
<i>Bacillus licheniformis</i> DSM 13								1	1			
<i>Bacillus subtilis</i>								1	1			
<i>Bacillus thuringiensis</i> konkukian									1			1
<i>Enterococcus faecalis</i> V583				1					1			1
<i>Geobacillus kaustophilus</i> HTA426									1			
<i>Lactobacillus acidophilus</i> NCFM									2			2
<i>Lactobacillus delbrueckii</i> bulgaricus									1			
<i>Lactobacillus johnsonii</i> NCC 533									1			2
<i>Lactobacillus plantarum</i>				1	1				1			5
<i>Lactobacillus sakei</i> 23K												1
<i>Lactobacillus salivarius</i> UCC118									1			1
<i>Lactococcus lactis</i>									1			
<i>Listeria innocua</i>				1	1				1			1
<i>Listeria monocytogenes</i>				1	1				1			1
<i>Listeria monocytogenes</i> 4b F2365				1	1				1			1
<i>Oceanobacillus iheyensis</i>									1			
<i>Staphylococcus aureus</i> MRSA252				1	1				1			1
<i>Staphylococcus aureus</i> MSSA476				1	1				1			1
<i>Staphylococcus aureus</i> COL				1	1				1			1
<i>Staphylococcus aureus</i> Mu50				1	1				1			1
<i>Staphylococcus aureus</i> MW2				1	1				1			1
<i>Staphylococcus aureus</i> N315				1	1				1			1
<i>Staphylococcus aureus</i> NCTC 8325				1	1				1			1
<i>Staphylococcus aureus</i> RF122				1	1				1			1
<i>Staphylococcus aureus</i> USA300				1	1				1			1
<i>Staphylococcus epidermidis</i> ATCC 12228				1	1				1			1
<i>Staphylococcus epidermidis</i> RP62A				1	1				1			1
<i>Staphylococcus haemolyticus</i>				1	1				1			1
<i>Staphylococcus saprophyticus</i>				1	1				1			1
<i>Streptococcus agalactiae</i> 2603												
<i>Streptococcus agalactiae</i> A909									1			1
<i>Streptococcus agalactiae</i> NEM316									1			
<i>Streptococcus mutans</i>									1			1
<i>Streptococcus pneumoniae</i> R6									1			2
<i>Streptococcus pneumoniae</i> TIGR4									1			2
<i>Streptococcus pyogenes</i> M1 GAS									1			3
<i>Streptococcus pyogenes</i> MGAS10270									1			3
<i>Streptococcus pyogenes</i> MGAS10394									1			3

Appendix 2: Transcription factors involved in yeast quorum sensing

Transcription factors that are likely to regulate the genes differentially expressed after various treatments (rows) according to each of the three networks (columns). Only transcription factors where $df > 1$ and where $p < 0.05$ are included (see main text for details). Transcription factors that regulate the expression of the differentially expressed genes in a treatment according to more than one network are highlighted in bold. PheOH: phenylethanol treatment, TrpOH: tryptophol treatment, TyrOH: tyrosol treatment, 3OH: treatment with a combination of all three aromatic alcohols. The three transcriptional networks CC, GRD and GROE were obtained using three different methods (see main text for details).

	CC network				GRD network				GROE network			
PheOH	name	p	d	abs	name	p	d	abs	name	p	d	abs
	CAT8	0	3.6696	8	RDS1	0	2.7641	8	MSN4	0	2.5023	23
	SIP4	0.0006	1.9915	8	YER130C	0	2.7116	9	MSN2	0	2.0511	26
	SKO1	0.0244	1.5586	5	ECM22	0.0079	2.7116	3	MIG1	0	1.8324	23
	PIP2	0.0224	1.1842	8	YER184C	0	2.6803	13	SPS18	0	1.643	17
	UME6	0.0001	1.1264	28	RGM1	0	2.6292	17	ADR1	0	1.6308	26
	MIG1	0.0386	1.1264	7	YPR196W	0.0003	2.519	6	RGM1	0.0004	1.4126	14
	MSN2	0.0008	1.0152	23	CAT8	0.0121	2.519	3	GIS1	0	1.3251	46
					WTM2	0	2.4746	12	MBP1	0	1.2682	26
					HAP3	0.0172	2.3491	3	SIP4	0.0447	1.1651	6
					YPR022C	0.002	2.3491	5	HSF1	0.0006	1.1019	20
					FLO8	0	2.2769	13	CAT8	0.0105	1.0458	12
					SRB8	0.0001	2.246	9	RSC30	0.0086	1.0301	13
					HIR1	0.0235	2.1971	3	CST6	0.0001	1.0141	30
					MBF1	0.0013	2.1971	6				
					YHP1	0.0037	2.1685	5				
					YLR278C	0.0006	2.1564	7				
					ARG80	0.0001	2.1498	9				
					CUP2	0	2.1267	10				
					GAT3	0.0017	2.1267	6				
				CSE2	0.0008	2.0975	7					
				HAL9	0.0307	2.0596	3					
				MGA2	0	2.0565	43					
				PHD1	0.0391	1.934	3					
				YFL052W	0.0391	1.934	3					
				YDR266C	0.0078	1.934	5					
				SET2	0.0078	1.934	5					
				SPT2	0.0017	1.934	7					
				CBF1	0.0002	1.8984	10					
				RPI1	0.0215	1.8466	4					
				ESC2	0.0081	1.7116	6					
				SWI4	0.0001	1.6539	14					
				SMK1	0.07	1.6121	3					
				SSN2	0.0282	1.5011	5					
				MIG1	0.0373	1.398	5					
				RIC1	0.0065	1.376	9					
				RPN4	0.0144	1.2902	8					
				RSC2	0.0345	1.2711	6					
				PDR1	0.0478	1.1614	6					

TrpOH	name p d abs				name p d abs				name p d abs			
	RDR1	0.0418	4.5817	1	MIG2	0.0439	4.511	1	MSN4	0	2.6226	15
SIP4	0.0018	2.1667	6	CAF17	0.0056	3.926	2	MSN2	0	2.0876	16	
PDR3	0.0063	2.0456	5	YER130C	0.0002	2.7036	6	SPS18	0.0001	1.8774	12	
PDR1	0.0014	1.3633	13	ECM22	0.0347	2.7036	2	MIG1	0.0001	1.7463	13	
NRG1	0.0032	1.0101	18	NGG1	0.0105	2.6365	3	RAP1	0.0001	1.4552	17	
				RGM1	0	2.5781	11	ADR1	0.0035	1.2523	12	
				YER184C	0	2.5568	8	GAT3	0.0078	1.1751	11	
				POP2	0.0449	2.511	2	HSF1	0.0082	1.1019	12	
				HIR2	0.0449	2.511	2	CRZ1	0.0023	1.0983	16	
				DAL82	0.0043	2.511	4	RFX1	0.006	1.0517	14	
				YBR239C	0.0449	2.511	2	MET31	0.0364	1.0458	8	
				CBF1	0	2.4753	10	TOS8	0.0276	1.0413	9	
				RDS1	0.0067	2.341	4	MGA1	0.0037	1.0356	16	
				CAD1	0.021	2.2886	3	CST6	0.0024	1.0141	18	
				SSN3	0.021	2.2886	3					
				WTM2	0.0004	2.2739	7					
				SNF11	0.0039	2.189	5					
				FLO8	0.0003	2.1534	8					
				RPN4	0.0002	2.037	9					
				SWI4	0	2.0085	12					
				SRB8	0.0076	1.9749	5					
				RFX1	0.0189	1.926	4					
				ARG80	0.0101	1.8787	5					
				CSE2	0.0218	1.8671	4					
				SSN2	0.0283	1.7561	4					
				CUP2	0.0169	1.7036	5					
				GCR2	0.0092	1.537	7					
				RSC1	0.0055	1.5337	8					
				MGA2	0	1.4551	19					
				RIC1	0.0267	1.368	6					
				SUM1	0.0032	1.3557	11					
TyrOH	name p d abs				name p d abs				name p d abs			
	CAT8	0	3.7215	5	MIG2	0.034	4.8787	1	MSN4	0	2.4871	11
SIP4	0.0065	2.0435	5	YER184C	0.0034	2.2464	5	MSN2	0	2.3065	15	
IXR1	0.0497	1.5121	4	YER130C	0.0325	2.0713	3	MIG1	0	2.0582	13	
GAL4	0.0176	1.2621	8	RGM1	0.0026	2.0713	6	SIP4	0.0373	1.6291	4	
PDR1	0.0157	1.1246	10	SRB8	0.0155	2.0207	4	TOS8	0.0021	1.5052	10	
TEC1	0.041	1.1195	7	CBF1	0.0116	1.8431	5	GIS1	0	1.4943	25	
YAP6	0.0032	1.0849	16	RPN4	0.0062	1.8198	6	RAP1	0.0003	1.4871	14	
				RSC2	0.0382	1.6308	4	ADR1	0.0018	1.4387	11	
				SWI4	0.0075	1.5986	7	RSC30	0.0101	1.3785	8	
				RSC1	0.0187	1.4864	6	MOT3	0.0015	1.2041	15	
				RIC1	0.0323	1.4727	5	XBP1	0.0003	1.2036	19	
				SUM1	0.0053	1.4339	9	SPS18	0.0446	1.1894	6	
				MGA2	0.0063	1.1599	12	ABF1	0.0117	1.1651	10	
								MET31	0.0335	1.1651	7	
								GAT4	0.0007	1.1571	18	
								CST6	0.0015	1.1561	16	
								MGA1	0.003	1.1549	14	
								GZF3	0.0177	1.1492	9	
								YOX1	0.0057	1.056	14	
3OH	name p d abs				name p d abs				name p d abs			
	RDR1	0.0404	4.6295	1	WAR1	0.0358	4.8031	1	MSN4	0	2.8856	18
CAT8	0	4.1149	7	HIR1	0.0045	3.0662	3	MSN2	0	2.48	21	
SIP4	0	2.7994	9	RIF2	0.0237	2.9958	2	SPS18	0	2.1994	15	
STB4	0.0423	1.5851	4	HCM1	0.0237	2.9958	2	MIG1	0	2.0458	16	
UME6	0.0005	1.205	19	RDS1	0.0003	2.9551	5	ADR1	0.0001	1.5742	15	
SKN7	0.0004	1.1453	21	YER184C	0	2.8489	8	TOS8	0.0003	1.5718	13	
PDR1	0.0235	1.0326	10	STP2	0.0309	2.8031	2	RGM1	0.0033	1.5121	9	
				RTF1	0.0309	2.8031	2	GIS1	0	1.4927	31	
				YLR278C	0.0001	2.8031	6	XBP1	0	1.3986	27	
				YER130C	0.0007	2.7328	5	GAT4	0	1.3191	25	
				PHO23	0.0388	2.6332	2	RAP1	0.0009	1.2746	15	
				RGM1	0	2.5808	9	CRZ1	0.0003	1.2682	18	

	CAD1	0.0122	2.5808	3	ABF1	0.0027	1.2317	13
	SRB8	0.0004	2.5301	6	CAT8	0.0206	1.1978	8
	YOX1	0.0474	2.4812	2	GAT3	0.0078	1.1751	11
	RTT107	0.0148	2.4812	3	RFX1	0.0022	1.1513	15
	YDR266C	0.0048	2.4812	4	HAP4	0.0047	1.0876	14
	GAT3	0.0058	2.4108	4	YHP1	0.0013	1.048	19
	YPR196W	0.0178	2.3881	3	MGA1	0.0037	1.0356	16
	WTM2	0.0009	2.3437	6	YOX1	0.003	1.0242	17
	ESC2	0.0027	2.3177	5	CST6	0.0024	1.0141	18
	FLO8	0.0005	2.2529	7	MBP1	0.0105	1.0052	13
	ARG80	0.0043	2.1709	5				
	RPN4	0.0003	2.1593	8				
	CSE2	0.011	2.1593	4				
	CHA4	0.0326	2.0662	3				
	MBF1	0.0326	2.0662	3				
	CBF1	0.0029	2.0306	6				
	CUP2	0.0074	1.9958	5				
	SPT2	0.0164	1.9958	4				
	MGA2	0	1.9588	22				
	MIG1	0.0185	1.9452	4				
	ISW1	0.0419	1.9287	3				
	HOG1	0.0419	1.9287	3				
	ARO80	0.047	1.8645	3				
	SRB2	0.0073	1.7734	6				
	RIC1	0.0106	1.6602	6				
	RSC2	0.045	1.5552	4				
	SW4	0.0099	1.523	7				
	RSC1	0.0237	1.4108	6				

Appendix 3: Full list of descriptors

List of all 236 descriptors used in the analysis, together with a short description, an indication (type) whether the descriptor is discrete (D) or continuous (C), a source or reference, and a number that indicates for how many genes data is available (size). The descriptors are sorted and colour-coded by class. References for those descriptors that were obtained from previously published datasets are at the end of the appendix (*References for the appendix*).

Key to classes and colour-code

Class (colour)	Comments	Number of descriptors
interaction network	derived from interaction networks, including protein interaction networks and transcriptional networks	27
evolutionary metrics	mainly descriptors derived from comparison to orthologues	12
dosage effect	the phenotypic effects of deleting or overexpressing the gene products	7
membrane associated	specific to membrane-associated proteins	3
gene location	information about the genomic location where the gene is encoded	6
information flow	associated with transcription, translation and degradation	14
sequence derived	computed from the nucleic or amino acid sequence of the gene	22
chemogenomic	viability of deletion mutants in the presence of various small molecules	23
protein location	cellular location of the protein	23
epigenetic	histone occupancy	13
gene expression	changes in gene expression in response to various stimuli and stresses	23

Table of descriptors

Descriptor name	Comments	Type	Source or reference	Size
interaction network				
2hybrid_cc	clustering coefficient of node in protein network created from two hybrid screening	C	[1]	1'964
2hybrid_centrality	how many shortest paths pass through node in protein network created from two hybrid screening	C	[1]	1'840
2hybrid_closeness	how many steps are required to access every other vertex in a protein network created from two hybrid screening	C	[1]	3'112
2hybrid_degree	number of neighbours in a protein network created from two hybrid screening (degree k)	C	[1]	3'112
affinity_cc	clustering coefficient of node in protein network created from affinity capture data	C	[1]	3'297
affinity_centrality	how many shortest paths pass through node in protein network created from affinity capture data	C	[1]	3'131
affinity_closeness	how many steps are required to access every other vertex in a protein network created from affinity capture data	C	[1]	4'030
affinity_degree	number of neighbours in a protein network created from affinity capture data (degree k)	C	[1]	4'030
avgpcc	if hub, correlation coefficient of mRNA expression of hub and its interacting proteins	C	[2]	199
cc_incoming_tfs	the number of transcription factors that regulate the gene according to ChIP-chip experiments (in-degree)	C	[3]	4'405
cc_target_num	the number of transcriptional targets regulated by the gene according to ChIP-chip experiments	C	[3]	157

complexsize	number of proteins in the protein's core complex	C	[4]	1'146
epi_cc	clustering coefficient of node in protein network created from epistatic interaction data	C	[1]	2'179
epi_centrality	how many shortest paths pass through node in protein network created from epistatic interaction data	C	[1]	2'008
epi_closeness	how many steps are required to access every other vertex in a protein network created from epistatic interaction data	C	[1]	3'010
epi_degree	number of neighbours in a protein network created from epistatic interaction data (degree k)	C	[1]	3'010
grd_incoming_tfs	the number of transcription factors that regulate the gene according to TF deletion experiments (in-degree)	C	[5]	2'567
grd_target_num	the number of transcriptional targets regulated by the gene according to TF deletion experiments	C	[5]	267
groe_incoming_tfs	the number of transcription factors that regulate the gene according to TF overexpression experiments (in-degree)	C	[6]	1'951
groe_target_num	the number of transcriptional targets regulated by the gene according to TF overexpression experiments	C	[6]	55
incoming_kinases	number of kinases that phosphorylate this protein according to kinase chips	C	[7]	1'182
metabolic_cc	clustering coefficient of gene in a directed enzyme network	C	[8]	486
metabolic_centrality	centrality of gene in a directed enzyme network	C	[8]	476
metabolic_connectivity	connectivity of gene in a directed enzyme network	C	[8]	579
metabolic_incoming	indegree of gene in a directed enzyme network	C	[8]	540
metabolic_outgoing	outdegree of gene in a directed enzyme network	C	[8]	
party_date_hub	whether the protein is a party hub (P) or a date hub (D)	D	[2]	199
evolutionary metrics				
expr_div	interspecies expression divergence between four different <i>Saccharomyces</i> species	C	[9]	4'035
identity_s_mikatae	FASTA sequence identity score with the best bidirectional hit in <i>Saccharomyces mikatae</i>	C	derived from the genome sequences	5'490
identity_sc_pombe	FASTA sequence identity score with the best bidirectional hit in <i>Schizosaccharomyces pombe</i>	C	derived from the genome sequences	3'084
kaks_average	ratio of synonymous to nonsynonymous mutations, average of comparison of orthologues of multiple species	C	[10]	3'384
kaks_s_mikatae	ratio of synonymous to nonsynonymous mutations in comparison to <i>Saccharomyces mikatae</i>	C	[10]	2'126
ks_s_mikatae	proportion of synonymous mutations in comparison to the <i>Saccharomyces mikatae</i> orthologue	C	[10]	2'154
ortho_e_coli	whether an orthologue is present in <i>Escherichia coli</i>	D	[11]	6'823
ortho_h_sapiens	whether an orthologue is present in humans	D	[11]	6'823
ortho_s_mikatae	whether an orthologue is present in <i>Saccharomyces mikatae</i>	D	[9]	6'823
ortho_sc_pombe	whether an orthologue is present in <i>Schizosaccharomyces pombe</i>	D	[11]	6'823
paralogues	number of paralogues with 0.9 length coverage threshold and 1.75 score coverage threshold	D	derived from the genome sequence using SEALS blastclust	6'823
positive_evol	whether the gene has undergone positive evolution in the lineage to strain S288	D	[12]	6'823
dosage effect				
essential	is the gene required for viability	D	[13]	5'855
gly_accumulation	glycogen accumulation in null mutants as measured by iodine staining	D	[14]	558
het_av	average fitness of the heterozygous deletion in YPD	C	[13]	5'773
het_mm	fitness of the heterozygous deletion in minimal media supplemented with histidine, uracil, and leucine	C	[13]	5'668
hom_av	average fitness of the homozygous deletion in YPD	C	[13]	4'731
hom_mm	fitness of the homozygous deletion in minimal media supplemented with histidine, uracil, and leucine	C	[13]	4'670
oe_toxicity	toxicity if the gene is overexpressed	D	[15]	768
membrane associated				
c_membranepos	if membrane protein, where is the C terminus predicted to be according to TMHMM	D	[16]	509
n_membranepos	if membrane protein, where is the N terminus predicted to be according to TMHMM	D	[16]	509
tm_helices	if membrane protein, how many transmembrane helices are predicted to be present according to TMHMM	C	[16]	598
gene location				
chromosome	chromosome the gene is encoded on	D	derived from locus name	6'702
chr_region	whether the gene is located in the telomeric region, the centromeric region, or in between	D	derived from genome	6'677

gnumber	order of the open reading frame, counting from the centromere	C	sequence derived from locus name	6'674
neigh_arrangement	arrangement of upstream ORFs, T - tail, H- head	D	[17]	5'557
start_nt	start nucleotide on the chromosome	C	derived from genome sequence	6'686
strand	strand - Watson (W) or Crick (C)	D	derived from locus name	6'674
information flow				
mrna_copy_num	mRNA copy number per cell	C	[18]	5'700
pest_motif	the number of potential PEST motifs that have been identified in the protein sequence	D	derived from the genome sequence using EMBOSS epestfind	6'823
protein_halflife	protein half-life in minutes after inhibition of translation	C	[19]	3'751
protein_level	protein abundance as measured using GFP tags	C	[20]	3'868
protein_per_s	protein per second, calculated by multiplying translation rate with mRNA copy number	C	[18]	5'643
ribosome_num	how many ribosomes were in the fraction in which abundance of mRNA peaked	D	[18]	5'700
ribo_density	the number of ribosomes per unit ORF length	C	[18]	5'700
ribo_occupancy	fraction of transcripts engaged in translation	C	[18]	5'700
tata_box	whether a TATA box is present around the regulatory region of the gene, is a proxy for transcriptional noise	D	[17]	6'823
tbp_occupancy	TATA binding protein occupancy as measured by ChIP-chip	C	[17]	2'370
transc_degrad	poly-A shortening of transcript	C	[21]	5'654
transc_halflife	transcript half-life in minutes after inhibition of transcription	C	[22]	4'686
transl_noise	the noise in protein levels, defined by the ratio of standard deviation to mean abundance	C	[23]	2'204
transl_rate	relative translation rate, calculated by multiplying ribosome occupancy with density	C	[18]	5'700
sequence derived				
aa_length	length of the protein in amino acids	C	derived from genome sequence	6'702
acidic_prop	proportion of acidic residues (D, E)	C	derived from genome sequence using EMBOSS pepstats	6'702
aggreg_prop	proportion of aggregation-prone residues in the protein	C	derived from genome sequence using TANGO	5'660
aliphatic_prop	proportion of aliphatic residues (I, L, V)	C	derived from genome sequence using EMBOSS pepstats	6'702
aromatic_prop	proportion of aromatic residues (F, H, W, Y)	C	derived from genome sequence using EMBOSS pepstats	6'702
basic_prop	proportion of basic residues (H, K, R)	C	derived from genome sequence using EMBOSS pepstats	6'702
cai	codon adaptation index	C	[24]	6'621
charge	charge of the protein	C	derived from genome sequence using EMBOSS pepstats	6'702
charged_prop	proportion of charged residues (D, E, H, K, R)	C	derived from	6'702

disorder	proportion of the protein predicted to be intrinsically unfolded as computed using DISOPRED	C	genome sequence using EMBOSS pepstats derived from the genome sequence using DISOPRED	6'640
gc_content	GC content of the coding sequence of the gene	C	derived from genome sequence	5'879
gravy	score of the hydropathicity of the protein	C	[24]	6'621
minfree_100nt_5p	minimum free energy of RNA, 100 nucleotides of 5' UTR	C	[25]	5'887
minfree_200nt_5p	minimum free energy of RNA, 200 nucleotides of 5' UTR	C	[25]	5'887
minfree_4_53nt	minimum free energy of RNA, nucleotides 4-53 downstream	C	[25]	5'821
minfree_50nt_3p	minimum free energy of RNA, 50 nucleotides of 3' UTR	C	[25]	5'586
minfree_50nt_5p	minimum free energy of RNA, 50 nucleotides of 5' UTR	C	[25]	5'535
pi	isoelectric point of the protein	C	derived from genome sequence using EMBOSS pepstats	6'702
polar_prop	proportion of polar residues (D, E, H, K, N, Q, R, S, T)	C	derived from genome sequence using EMBOSS pepstats	6'702
stackingen	stacking energy of the coding B-DNA sequence of the gene	C	derived from the genome sequence using EMBOSS btwisted	5'879
stackingen_nt	average stacking energy of the coding B-DNA sequence of the gene per dinucleotide	C	derived from the genome sequence using EMBOSS btwisted	5'879
tiny_prop	proportion of tiny residues (A, C, G, S, T)	C	derived from genome sequence using EMBOSS pepstats	6'702
chemogenomic				
cg_benomyl	chemogenomic profile in response to benomyl - microtubule function	D	[26]	4'616
cg_cacl2	chemogenomic profile in response to calcium chloride	D	[26]	4'616
cg_cad	chemogenomic profile in response to cadmium chloride	D	[26]	4'616
cg_caff	chemogenomic profile in response to caffeine	D	[26]	4'616
cg_cyclohex	chemogenomic profile in response to cycloheximide - protein synthesis	D	[26]	4'616
cg_etoh	chemogenomic profile in response to ethanol	D	[26]	4'616
cg_felim	chemogenomic profile in response to iron-limited YPD	D	[26]	4'616
cg_hu	chemogenomic profile in response to hydroxyurea - DNA replication and repair	D	[26]	4'616
cg_hygrob	chemogenomic profile in response to hygromycin B	D	[26]	4'616
cg_lowpo4	chemogenomic profile in response to low-phosphate YPD	D	[26]	4'616
cg_mpa	chemogenomic profile in response to mycophenolic acid - transcriptional elongation	D	[26]	4'616
cg_nacl	chemogenomic profile in response to sodium chloride	D	[26]	4'616
cg_paraq	chemogenomic profile in response to paraquat	D	[26]	4'616
cg_ph3	chemogenomic profile in response to pH 3	D	[26]	4'616
cg_rap	chemogenomic profile in response to rapamycin - protein synthesis	D	[26]	4'616
cg_sorb	chemogenomic profile in response to sorbitol	D	[26]	4'616
cg_uv	chemogenomic profile in response to ultraviolet radiation	D	[26]	4'616
cg_ypgal	chemogenomic profile in response to galactose	D	[26]	4'616
cg_ypgly	chemogenomic profile in response to glycerol	D	[26]	4'616
cg_yplac	chemogenomic profile in response to lactate	D	[26]	4'616

cg_ypraff	chemogenomic profile in response to raffinose	D	[26]	4'616
het_conditions	number of conditions the heterozygous deletion mutant is sensitive to	C	[27]	5'744
hom_conditions	number of conditions the homozygous deletion mutant is sensitive to	C	[27]	4'652
protein location				
actin	whether the protein is associated with actin	D	[20]	6'823
ambiguous_loc	whether the protein is present in an ambiguous location	D	[20]	6'823
bud	whether the protein is present in the bud	D	[20]	6'823
bud_neck	whether the protein is present in the bud neck	D	[20]	6'823
cell_periphery	whether the protein is present in the cell periphery	D	[20]	6'823
cytoplasm	whether the protein is present in the cytoplasm	D	[20]	6'823
early_golgi	whether the protein is present in the early Golgi	D	[20]	6'823
endosome	whether the protein is present in the endosome	D	[20]	6'823
er	whether the protein is present in the endoplasmic reticulum	D	[20]	6'823
er_to_golgi	whether the protein is present in the endoplasmic reticulum to Golgi	D	[20]	6'823
golgi	whether the protein is present in the Golgi	D	[20]	6'823
late_golgi	whether the protein is present in the late Golgi	D	[20]	6'823
lipid_particle	whether the protein is present in the lipid particle	D	[20]	6'823
microtubule	whether the protein is present in the microtubules	D	[20]	6'823
mitochondrion	whether the protein is present in the mitochondrion	D	[20]	6'823
nuclear_periphery	whether the protein is present in the nuclear periphery	D	[20]	6'823
nucleolus	whether the protein is present in the nucleolus	D	[20]	6'823
nucleus	whether the protein is present in the nucleus	D	[20]	6'823
peroxisome	whether the protein is present in the peroxisome	D	[20]	6'823
punctate_composite	whether the protein is present in the punctate composite	D	[20]	6'823
spindle_pole	whether the protein is present in the spindle pole	D	[20]	6'823
vacuolar_membrane	whether the protein is present in the vacuolar membrane	D	[20]	6'823
vacuole	whether the protein is present in the vacuole	D	[20]	6'823
epigenetic				
h3k14acvsh3_5p	nucleosome occupancy of gene's 5'-most tile as measured by ChIP-chip - H3K14ac vs H3	C	[28]	5'524
h3k14acvsh3_h2o2_5p	nucleosome occupancy of gene's 5'-most tile as measured by ChIP-chip - H3K14ac vs H3 in the presence of hydrogen peroxide	C	[28]	5'525
h3k36me3vsh3_5p	nucleosome occupancy of gene's 5'-most tile as measured by ChIP-chip - H3K36me3 vs H3	C	[28]	5'525
h3k4me1vsh3_5p	nucleosome occupancy of gene's 5'-most tile as measured by ChIP-chip - H3K4me1 vs H3	C	[28]	5'517
h3k4me2vsh3_5p	nucleosome occupancy of gene's 5'-most tile as measured by ChIP-chip - H3K4me2 vs H3	C	[28]	5'519
h3k4me3vsh3_5p	nucleosome occupancy of gene's 5'-most tile as measured by ChIP-chip - H3K4me3 vs H3	C	[28]	5'525
h3k79me3vsh3_5p	nucleosome occupancy of gene's 5'-most tile as measured by ChIP-chip - H3K79me3 vs H3	C	[28]	5'525
h3k9acvsh3_5p	nucleosome occupancy of gene's 5'-most tile as measured by ChIP-chip - H3K9ac vs H3	C	[28]	5'515
h3_5p	nucleosome occupancy of gene's 5'-most tile as measured by ChIP-chip - H3	C	[28]	5'522
h3_h2o2_5p	nucleosome occupancy of gene's 5'-most tile as measured by ChIP-chip - H3 in the presence of hydrogen peroxide	C	[28]	5'522
h4acvsh3_5p	nucleosome occupancy of gene's 5'-most tile as measured by ChIP-chip - H4ac vs H3	C	[28]	5'527
h4acvsh3_h2o2_5p	nucleosome occupancy of gene's 5'-most tile as measured by ChIP-chip - H4ac vs H3 in the presence of hydrogen peroxide	C	[28]	5'526
h4_5p	nucleosome occupancy of gene's 5'-most tile as measured by ChIP-chip - H4	C	[28]	5'527
gene expression				
ge_25c_growth	changes in gene expression - 25C growth	C	[29]	6'010
ge_aa_starvation	changes in gene expression - amino acid starvation	C	[29]	5'979
ge_coldshock	changes in gene expression - 37C to 25C temperature shock	C	[29]	6'061
ge_ddt	changes in gene expression - DDT treatment	C	[29]	4'777
ge_diamide	changes in gene expression - diamide treatment	C	[29]	6'006
ge_diauxic_10h	changes in gene expression - 9.5 hours into diauxic shift	C	[29]	5'857
ge_diauxic_12h	changes in gene expression - 11.5 hours into diauxic shift	C	[29]	5'825
ge_diauxic_19h	changes in gene expression - 18.5 hours into diauxic shift	C	[29]	5'969
ge_ethanol	changes in gene expression - ethanol treatment	C	[29]	5'965
ge_galactose	changes in gene expression - galactose treatment	C	[29]	5'929
ge_glucose	changes in gene expression - glucose treatment	C	[29]	6'040
ge_h2o2	changes in gene expression - hydrogen peroxide treatment	C	[29]	5'872
ge_heatshock	changes in gene expression - heat shock	C	[29]	5'331
ge_hypoosmotic	changes in gene expression - hypo-osmotic shock	C	[29]	6'026
ge_irradiation	changes in gene expression - Gamma irradiation causing DNA damage	C	[29]	5'925

ge_menadione	changes in gene expression - menadione treatment	C	[29]	5'947
ge_n_starvation	changes in gene expression - 12 hours into nitrogen depletion	C	[29]	6'015
ge_pheoh	changes in gene expression - addition of phenylethanol	C	[30]	5'840
ge_raffinose	changes in gene expression - raffinose treatment	C	[29]	5'942
ge_sorbitol	changes in gene expression - sorbitol treatment	C	[29]	5'978
ge_statphaseentry	changes in gene expression - 12 hours after entry into stationary phase	C	[29]	5'997
ge_steady_sorbitol	changes in gene expression - steady-state 1M sorbitol	C	[29]	5'837
ge_trpoh	changes in gene expression - addition of tryptophol	C	[30]	5'818
Gene Ontology				
go0000746	GO slim ID GO:0000746 (biological_process): conjugation	D	[31]	6'823
go0000910	GO slim ID GO:0000910 (biological_process): cytokinesis	D	[31]	6'823
go0003677	GO slim ID GO:0003677 (molecular_function): DNA binding	D	[31]	6'823
go0003723	GO slim ID GO:0003723 (molecular_function): RNA binding	D	[31]	6'823
go0003774	GO slim ID GO:0003774 (molecular_function): motor activity	D	[31]	6'823
go0004386	GO slim ID GO:0004386 (molecular_function): helicase activity	D	[31]	6'823
go0004672	GO slim ID GO:0004672 (molecular_function): protein kinase activity	D	[31]	6'823
go0004721	GO slim ID GO:0004721 (molecular_function): phosphoprotein phosphatase activity	D	[31]	6'823
go0004871	GO slim ID GO:0004871 (molecular_function): signal transducer activity	D	[31]	6'823
go0005198	GO slim ID GO:0005198 (molecular_function): structural molecule activity	D	[31]	6'823
go0005215	GO slim ID GO:0005215 (molecular_function): transporter activity	D	[31]	6'823
go0005515	GO slim ID GO:0005515 (molecular_function): protein binding	D	[31]	6'823
go0005618	GO slim ID GO:0005618 (cellular_component): cell wall	D	[31]	6'823
go0005624	GO slim ID GO:0005624 (cellular_component): membrane fraction	D	[31]	6'823
go0005634	GO slim ID GO:0005634 (cellular_component): nucleus	D	[31]	6'823
go0005694	GO slim ID GO:0005694 (cellular_component): chromosome	D	[31]	6'823
go0005730	GO slim ID GO:0005730 (cellular_component): nucleolus	D	[31]	6'823
go0005737	GO slim ID GO:0005737 (cellular_component): cytoplasm	D	[31]	6'823
go0005739	GO slim ID GO:0005739 (cellular_component): mitochondrion	D	[31]	6'823
go0005740	GO slim ID GO:0005740 (cellular_component): mitochondrial envelope	D	[31]	6'823
go0005773	GO slim ID GO:0005773 (cellular_component): vacuole	D	[31]	6'823
go0005777	GO slim ID GO:0005777 (cellular_component): peroxisome	D	[31]	6'823
go0005783	GO slim ID GO:0005783 (cellular_component): endoplasmic reticulum	D	[31]	6'823
go0005794	GO slim ID GO:0005794 (cellular_component): Golgi apparatus	D	[31]	6'823
go0005815	GO slim ID GO:0005815 (cellular_component): microtubule organizing center	D	[31]	6'823
go0005840	GO slim ID GO:0005840 (cellular_component): ribosome	D	[31]	6'823
go0005856	GO slim ID GO:0005856 (cellular_component): cytoskeleton	D	[31]	6'823
go0005886	GO slim ID GO:0005886 (cellular_component): plasma membrane	D	[31]	6'823
go0005938	GO slim ID GO:0005938 (cellular_component): cell cortex	D	[31]	6'823
go0005975	GO slim ID GO:0005975 (biological_process): carbohydrate metabolic process	D	[31]	6'823
go0006091	GO slim ID GO:0006091 (biological_process): generation of precursor metabolites and energy	D	[31]	6'823
go0006259	GO slim ID GO:0006259 (biological_process): DNA metabolic process	D	[31]	6'823
go0006350	GO slim ID GO:0006350 (biological_process): transcription	D	[31]	6'823
go0006412	GO slim ID GO:0006412 (biological_process): translation	D	[31]	6'823
go0006457	GO slim ID GO:0006457 (biological_process): protein folding	D	[31]	6'823
go0006464	GO slim ID GO:0006464 (biological_process): protein modification process	D	[31]	6'823
go0006519	GO slim ID GO:0006519 (biological_process): amino acid and derivative metabolic process	D	[31]	6'823
go0006629	GO slim ID GO:0006629 (biological_process): lipid metabolic process	D	[31]	6'823
go0006725	GO slim ID GO:0006725 (biological_process): aromatic compound metabolic process	D	[31]	6'823
go0006766	GO slim ID GO:0006766 (biological_process): vitamin metabolic process	D	[31]	6'823
go0006810	GO slim ID GO:0006810 (biological_process): transport	D	[31]	6'823
go0006950	GO slim ID GO:0006950 (biological_process): response to stress	D	[31]	6'823
go0006996	GO slim ID GO:0006996 (biological_process): organelle organization and biogenesis	D	[31]	6'823
go0006997	GO slim ID GO:0006997 (biological_process): nuclear organization and biogenesis	D	[31]	6'823
go0007010	GO slim ID GO:0007010 (biological_process): cytoskeleton organization and biogenesis	D	[31]	6'823

go0007047	GO slim ID GO:0007047 (biological_process): cell wall organization and biogenesis	D	[31]	6'823
go0007049	GO slim ID GO:0007049 (biological_process): cell cycle	D	[31]	6'823
go0007114	GO slim ID GO:0007114 (biological_process): cell budding	D	[31]	6'823
go0007124	GO slim ID GO:0007124 (biological_process): pseudohyphal growth	D	[31]	6'823
go0007126	GO slim ID GO:0007126 (biological_process): meiosis	D	[31]	6'823
go0007165	GO slim ID GO:0007165 (biological_process): signal transduction	D	[31]	6'823
go0008233	GO slim ID GO:0008233 (molecular_function): peptidase activity	D	[31]	6'823
go0008289	GO slim ID GO:0008289 (molecular_function): lipid binding	D	[31]	6'823
go0009653	GO slim ID GO:0009653 (biological_process): anatomical structure morphogenesis	D	[31]	6'823
go0012505	GO slim ID GO:0012505 (cellular_component): endomembrane system	D	[31]	6'823
go0016020	GO slim ID GO:0016020 (cellular_component): membrane	D	[31]	6'823
go0016023	GO slim ID GO:0016023 (cellular_component): cytoplasmic membrane-bound vesicle	D	[31]	6'823
go0016044	GO slim ID GO:0016044 (biological_process): membrane organization and biogenesis	D	[31]	6'823
go0016070	GO slim ID GO:0016070 (biological_process): RNA metabolic process	D	[31]	6'823
go0016192	GO slim ID GO:0016192 (biological_process): vesicle-mediated transport	D	[31]	6'823
go0016491	GO slim ID GO:0016491 (molecular_function): oxidoreductase activity	D	[31]	6'823
go0016740	GO slim ID GO:0016740 (molecular_function): transferase activity	D	[31]	6'823
go0016779	GO slim ID GO:0016779 (molecular_function): nucleotidyltransferase activity	D	[31]	6'823
go0016787	GO slim ID GO:0016787 (molecular_function): hydrolase activity	D	[31]	6'823
go0016829	GO slim ID GO:0016829 (molecular_function): lyase activity	D	[31]	6'823
go0016853	GO slim ID GO:0016853 (molecular_function): isomerase activity	D	[31]	6'823
go0016874	GO slim ID GO:0016874 (molecular_function): ligase activity	D	[31]	6'823
go0019725	GO slim ID GO:0019725 (biological_process): cellular homeostasis	D	[31]	6'823
go0030163	GO slim ID GO:0030163 (biological_process): protein catabolic process	D	[31]	6'823
go0030234	GO slim ID GO:0030234 (molecular_function): enzyme regulator activity	D	[31]	6'823
go0030435	GO slim ID GO:0030435 (biological_process): sporulation	D	[31]	6'823
go0030528	GO slim ID GO:0030528 (molecular_function): transcription regulator activity	D	[31]	6'823
go0032196	GO slim ID GO:0032196 (biological_process): transposition	D	[31]	6'823
go0042221	GO slim ID GO:0042221 (biological_process): response to chemical stimulus	D	[31]	6'823
go0042254	GO slim ID GO:0042254 (biological_process): ribosome biogenesis and assembly	D	[31]	6'823
go0045182	GO slim ID GO:0045182 (molecular_function): translation regulator activity	D	[31]	6'823
go0045333	GO slim ID GO:0045333 (biological_process): cellular respiration	D	[31]	6'823
go0046483	GO slim ID GO:0046483 (biological_process): heterocycle metabolic process	D	[31]	6'823
go0051186	GO slim ID GO:0051186 (biological_process): cofactor metabolic process	D	[31]	6'823

Appendix 4: List of publications

Publications relevant to this thesis

- Wuster, A. and M. Madan Babu (2009) "Scorrel: A database for finding associations between cellular descriptors." (*in preparation*)
- Wuster, A. and M. Madan Babu (2009) "Analysis of the global set of associations between cellular descriptors." (*in preparation*)
- Wuster, A. and M. Madan Babu (2009) "Computational approaches to identifying quorum sensing systems in bacteria." (*in preparation*)
- Wuster, A., G. F. Schertler and M. Madan Babu (2009). "Spial: Analysis of subtype-specific features in multiple sequence alignments of proteins." Bioinformatics (*in review*)
- Wuster, A. and M. Madan Babu (2009). "Transcriptional control of the quorum sensing response in yeast." Mol BioSyst, DOI: 10.1039/b913579k
- Wuster, A. and M. Madan Babu (2008). "Chemogenomics and biotechnology." Trends Biotechnol **26**(5): 252-8
- Wuster, A. (2008) "Networking with drugs" Mol BioSyst **4**: 14
- Wuster, A. and M. Madan Babu (2008). "Conservation and evolutionary dynamics of the *agr* cell-to-cell communication system across firmicutes." J Bacteriol **190**(2): 743-6
- Wuster, A. and M. Madan Babu (2008). "Chemical Molecules that Regulate Transcription and Facilitate Cell-to-cell Communication." book chapter in Wiley Encyclopedia of Chemical Biology, Hoboken, New Jersey: John Wiley and Sons, Inc.

Other publications

- Jothi, R., S. Balaji, A. Wuster, J. A. Grochow, J. Gsponer J, T. M. Przytycka, L. Aravind and M. Madan Babu (2009). "Genomic Analysis reveals a tight link between transcription factor dynamics and regulatory network architecture." Mol Syst Biol **5**: 294

References for the appendix

1. Stark, C., et al., *BioGRID: a general repository for interaction datasets*. Nucleic Acids Res, 2006. **34**(Database issue): p. D535-9.
2. Han, J.D., et al., *Evidence for dynamically organized modularity in the yeast protein-protein interaction network*. Nature, 2004. **430**(6995): p. 88-93.
3. Balaji, S., et al., *Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast*. J Mol Biol, 2006. **360**(1): p. 213-27.
4. Gavin, A.C., et al., *Proteome survey reveals modularity of the yeast cell machinery*. Nature, 2006. **440**(7084): p. 631-6.
5. Hu, Z., P.J. Killion, and V.R. Iyer, *Genetic reconstruction of a functional transcriptional regulatory network*. Nat Genet, 2007. **39**(5): p. 683-7.
6. Chua, G., et al., *Identifying transcription factor functions and targets by phenotypic activation*. Proc Natl Acad Sci U S A, 2006. **103**(32): p. 12045-50.
7. Ptacek, J., et al., *Global analysis of protein phosphorylation in yeast*. Nature, 2005. **438**(7068): p. 679-84.
8. Lu, C., et al., *Impacts of yeast metabolic network structure on enzyme evolution*. Genome Biol, 2007. **8**(8): p. 407.
9. Tirosh, I., et al., *A genetic signature of interspecies variations in gene expression*. Nat Genet, 2006. **38**(7): p. 830-4.
10. Kellis, M., et al., *Sequencing and comparison of yeast species to identify genes and regulatory elements*. Nature, 2003. **423**(6937): p. 241-54.
11. Berglund, A.C., et al., *InParanoid 6: eukaryotic ortholog clusters with inparalogs*. Nucleic Acids Res, 2008. **36**(Database issue): p. D263-6.
12. Gu, Z., et al., *Elevated evolutionary rates in the laboratory strain of *Saccharomyces cerevisiae**. Proc Natl Acad Sci U S A, 2005. **102**(4): p. 1092-7.
13. Deutschbauer, A.M., et al., *Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast*. Genetics, 2005. **169**(4): p. 1915-25.
14. Wilson, W.A., Z. Wang, and P.J. Roach, *Systematic identification of the genes affecting glycogen storage in the yeast *Saccharomyces cerevisiae*: implication of the vacuole as a determinant of glycogen level*. Mol Cell Proteomics, 2002. **1**(3): p. 232-42.
15. Sopko, R., et al., *Mapping pathways and phenotypes by systematic gene overexpression*. Mol Cell, 2006. **21**(3): p. 319-30.
16. Osterberg, M., et al., *Phenotypic effects of membrane protein overexpression in *Saccharomyces cerevisiae**. Proc Natl Acad Sci U S A, 2006. **103**(30): p. 11148-53.
17. Basehoar, A.D., S.J. Zanton, and B.F. Pugh, *Identification and distinct regulation of yeast TATA box-containing genes*. Cell, 2004. **116**(5): p. 699-709.
18. Arava, Y., et al., *Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae**. Proc Natl Acad Sci U S A, 2003. **100**(7): p. 3889-94.
19. Belle, A., et al., *Quantification of protein half-lives in the budding yeast proteome*. Proc Natl Acad Sci U S A, 2006. **103**(35): p. 13004-9.
20. Huh, W.K., et al., *Global analysis of protein localization in budding yeast*. Nature, 2003. **425**(6959): p. 686-91.
21. Beilharz, T.H. and T. Preiss, *Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome*. Rna, 2007. **13**(7): p. 982-97.
22. Lackner, D.H., et al., *A network of multiple regulatory layers shapes gene expression in fission yeast*. Mol Cell, 2007. **26**(1): p. 145-55.

23. Newman, J.R., et al., *Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise*. Nature, 2006. **441**(7095): p. 840-6.
24. Lu, P., et al., *Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation*. Nat Biotechnol, 2007. **25**(1): p. 117-24.
25. Ringner, M. and M. Krogh, *Folding free energies of 5'-UTRs impact post-transcriptional regulation on a genomic scale in yeast*. PLoS Comput Biol, 2005. **1**(7): p. e72.
26. Dudley, A.M., et al., *A global view of pleiotropy and phenotypically derived gene function in yeast*. Mol Syst Biol, 2005. **1**: p. 2005 0001.
27. Hillenmeyer, M.E., et al., *The chemical genomic portrait of yeast: uncovering a phenotype for all genes*. Science, 2008. **320**(5874): p. 362-5.
28. Pokholok, D.K., et al., *Genome-wide map of nucleosome acetylation and methylation in yeast*. Cell, 2005. **122**(4): p. 517-27.
29. Gasch, A.P., et al., *Genomic expression programs in the response of yeast cells to environmental changes*. Mol Biol Cell, 2000. **11**(12): p. 4241-57.
30. Chen, H. and G.R. Fink, *Feedback control of morphogenesis in fungi by aromatic alcohols*. Genes Dev, 2006. **20**(9): p. 1150-61.
31. Ashburner, M., et al., *Gene ontology: tool for the unification of biology*. The Gene Ontology Consortium. Nat Genet, 2000. **25**(1): p. 25-9.