## Applied Microbiology and Biotechnology

Insights into the prevalence and underlying causes of clonal variation through transcriptomic analysis in *Pichia pastoris* 

Rochelle Aw<sup>1,2</sup>, Geraint R Barton<sup>3</sup> and David J. Leak<sup>1,4\*</sup>

<sup>1</sup>Department of Life Sciences, Imperial College London, London, SW7 2AZ

<sup>2</sup>Centre for Synthetic Biology and Innovation, Imperial College London, London, SW7 2AZ

<sup>3</sup>Centre for Integrative Systems Biology and Bioinformatics, Imperial College London, London, SW7 2AZ

<sup>4</sup> Current address: Department of Biology & Biochemistry, University of Bath, Bath BA2

7AY, UK

\*Corresponding author

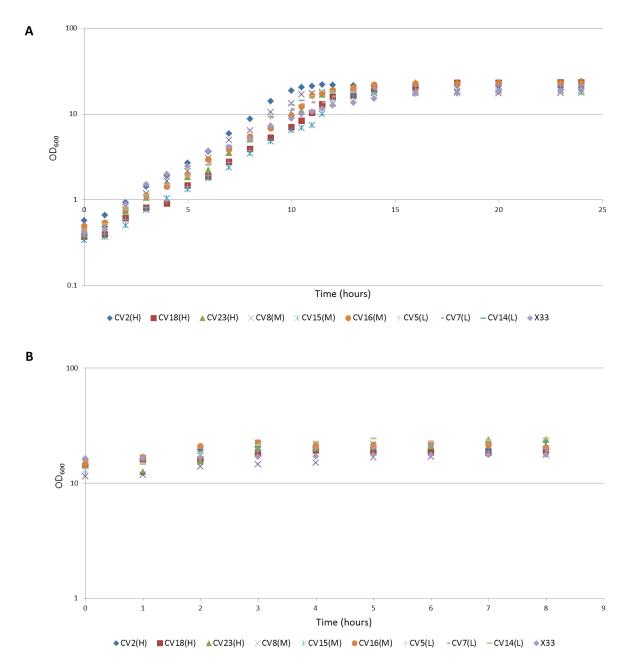
Email: <u>d.j.leak@bath.ac.uk</u>

Tel: +44 (0) 1225 384309

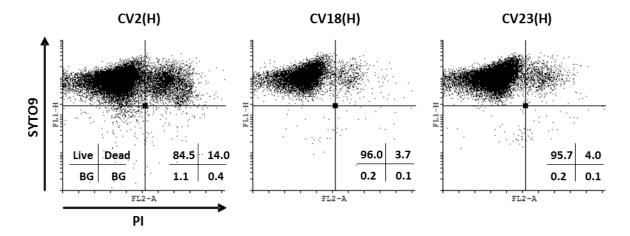
## **Supplementary Materials**

	460	470	480	490	500	510	520	530	540
		. <mark>  </mark>			.				
CV2	TGG <mark>CCC</mark> AAAAC <mark>T</mark> GA	CAGTTTAAAC	G <mark>C</mark> TGTCTTGGA	ACC TAATA	T G <mark>acaaaa Gc</mark> g	TGATCTCATC	CAAGA <mark>T</mark> GAAC	T <mark>AA</mark> G TTTGGT	г <mark>С</mark> GTTG
CV5	TGG <mark>CCC</mark> AAAAC <mark>T</mark> GA	CAGTTTAAAC	GCTGTCTTGGA	ACCTAATA	T G <mark>acaaaa gc</mark> g	TGATCTCATC	CAAGA <mark>T</mark> GAAC	TAAGTTTGGT'	г <mark>с</mark> дттс
CV7	TGG <mark>CCC</mark> AAAAC <mark>T</mark> GA	CAGTTTAAAC	GCTGTCTTGGA	ACC TAATA	T G <mark>acaaaa gc</mark> g	TGATCTCATC	CAAGATGAAC	TAAGTTTGGT	г <mark>с</mark> дттс
CV8	TGG <mark>CCC</mark> AAAAC <mark>T</mark> GA	CAGTTTAAAC	GCTGTCTTGGA	ACC TAATA	T G <mark>acaaaa gc</mark> g	TGATCTCATC	CAAGA <mark>T</mark> GAAC	TAAGTTTGGT	г <mark>с</mark> дттс
CV14	TGG <mark>CCC</mark> AAAAC <mark>T</mark> GA	CAGTTTAAAC	GCTGTCTTGGA	ACC TAATA	T G <mark>acaaaa gc</mark> g	TGATCTCATC	CAAGA <mark>T</mark> GAAC	T <mark>AA</mark> G TTTGGT	г <mark>с</mark> дттс
CV15	TGG <mark>CCC</mark> AAAAC <mark>T</mark> GA	CAGTTTAAAC	GCTGTCTTGGA	ACC TAATA	T GACAAAA GCG	TGATCTCATC	CAAGATGAAC	TAAG TTTGGT	г <mark>с</mark> сттс
CV16	TGG <mark>CCC</mark> AAAAC <mark>T</mark> GA	CAGTTTAAAC	GCTGTCTTGGA	ACC TAATA	T GACAAAA GCG	TGATCTCATC	CAAGATGAAC	TAAG TTTGGT	г <mark>с</mark> сттс
CV18	TGG <mark>CCC</mark> AAAAC <mark>T</mark> GA	CAGTTTAAAC	GCTGTCTTGGA	ACC TAATA	T GACAAAA GCG	TGATCTCATC	CAAGATGAAC	TAAG TTTGGT	г <mark>с</mark> дттс
CV23	TGG <mark>CCC</mark> AAAACTGA	CAGTTTAAAC	GCTGTCTTGGA	ACC TAATA	T GACAAAA GCG	TGAT CTCATC	CAAGA <mark>T</mark> GAAC	TAAG TTTGGT	г <mark>с</mark> дттс
	550	560	570	580	590	600	610	620	630
	550 	560 	570	580 	590 .	600 	610 	620 	630 
CV2	550 ••••• ••••• ••••• <mark>AAAT</mark> GCTAACGGCC		1			1	1	1	
CV2 CV5		 Ag <mark>tt</mark> gg <mark>tcaa</mark>	 AAA <mark>GAAACTT</mark> O		.     GG <mark>CATACC</mark> GTT	 TGT <mark>C</mark> TTGTTT	 GG <mark>TATTGA</mark> TT	 Gacgaa <mark>t</mark> gct	 СААААА
0.12	AAA <mark>T GCT</mark> AACGGCC	 AGTT GG <mark>TCAA</mark> AGTT GG <mark>TCAA</mark>	 AAAGAAACTTC AAAGAAACTTC	CAAAAG <mark>TC</mark> CAAAAG <mark>TC</mark>	.     GG <mark>CATACC</mark> GTT GG <mark>CATACC</mark> GTT	 TGTCTTGTTT TGTCTTGTTT	 GG <mark>TATTGATT</mark> GG <mark>TATTGATT</mark>	GACGAATGCT	 Сааааа Сааааа
CV5	AAATGCTAACGGCC AAAT <mark>GCT</mark> AACGGCC	 AGTTGG <mark>TC</mark> AA AGTTGGTCAA AG <mark>TT</mark> GG <mark>TC</mark> AA	 AAAGAAA CTTC AAAGAAA CTTC AAAGAAA CTTC	CAAAAGTC CAAAAGTC CAAAAGTC	GG <mark>CATACC</mark> GTT GG <mark>CATACC</mark> GTT GG <mark>CATACC</mark> GTT	TGTCTTGTTT TGTCTTGTTT TGTCTTGTTT	. GGTATTGATT GGTATTGATT GGTATTGATT	GACGAATGCT GACGAATGCT GACGAATGCT	 CAAAAA CAAAAA CAAAAA
CV5 CV7	AAATGCTAACGGCC AAATGCTAACGGCC AAATGCTAACGGCC AAA <mark>TGCT</mark> AACGG <mark>CC</mark>	 AG <mark>TT</mark> GGTCAA AGTTGGTCAA AGTTGGTCAA AGTTGGTCAA	AAAGAAACTTC AAAGAAACTTC AAAGAAACTTC AAAGAAACTTC AAAGAAACTTC	CAAAAGTC CAAAAGTC CAAAAGTC CAAAAGTC	GGCATACCGTT GGCATACCGTT GGCATACCGTT GG <mark>CATACC</mark> GT	TGTCTTGTTT TGTCTTGTTT TGTCTTGTTT TGTCTTGTTT	GGTATTGATT GGTATTGATT GGTATTGATT GG <mark>TATTGA</mark> TT	GACGAATGCT GACGAATGCT GACGAATGCT GACGAATGCT	 Сааааа Сааааа Сааааа Сааааа
CV5 CV7 CV8	AAATGCTAACGGCC AAATGCTAACGGCC AAATGCTAACGGCC AAATGCTAACGGCC AAATGCTAACGGCC	AGTTGGTCAA AGTTGGTCAA AGTTGGTCAA AGTTGGTCAA AGTTGGTCAA	AAAGAAACTTO AAAGAAACTTO AAAGAAACTTO AAAGAAACTTO AAAGAAACTTO AAAGAAACTTO	CAAAAGTC CAAAAGTC CAAAAGTC CAAAAGTC CAAAAGTC	GGCATACCGTT GGCATACCGTT GGCATACCGTT GGCATACCGTT GGCATACCGTT	TGTCTTGTTT TGTCTTGTTT TGTCTTGTTT TGTCTTGTTT TGTCTTGTTT	GGTATTGATT GGTATTGATT GGTATTGATT GGTATTGATT GGT <mark>ATT</mark> GATT	GACGAATGCT GACGAATGCT GACGAATGCT GACGAATGCT GACGAATGCT	CAAAAA CAAAAA CAAAAA CAAAAA CAAAAA
cv5 cv7 cv8 cv14	AAAT GCTAACGGCC AAAT GCTAACGGCC AAAT GCTAACGGCC AAAT GCTAACGGCC AAAT GCTAACGGCC	AGTT GGTCAA AGTT GGTCAA AGTT GGTCAA AGTT GGTCAA AGTT GGTCAA AGTT GGTCAA	AAAGAAACTTC AAAGAAACTTC AAAGAAACTTC AAAGAAACTTC AAAGAAACTTC AAAGAAACTTC AAAGAAACTTC	CAAAAGTC CAAAAGTC CAAAAGTC CAAAAGTC CAAAAGTC CAAAAGTC	GGCATACCGTT GGCATACCGTT GGCATACCGTT GGCATACCGTT GGCATACCGTT GGCATACCGTT	TGTCTTGTTT TGTCTTGTTT TGTCTTGTTT TGTCTTGTTT TGTCTTGTTT TGTCTTGTTT	GGTATTGATT GGTATTGATT GGTATTGATT GGTATTGATT GGTATTGATT GGTATTGATT	GACGAATGCT GACGAATGCT GACGAATGCT GACGAATGCT GACGAATGCT GACGAATGCT	CAAAAA CAAAAA CAAAAA CAAAAA CAAAAA CAAAAA
CV5 CV7 CV8 CV14 CV15	AAATGCTAACGGCC AAATGCTAACGCCC AAATGCTAACGCCC AAATGCTAACGCCC AAATGCTAACGCCC AAATGCTAACGCCC AAATGCTAACGCCC	 AGTTGGTCAA AGTTGGTCAA AGTTGGTCAA AGTTGGTCAA AGTTGGTCAA AGTTGGTCAA	AAAGAAACTTC AAAGAAACTTC AAAGAAACTTC AAAGAAACTTC AAAGAAACTTC AAAGAAACTTC AAAGAAACTTC	CAAAAGTC CAAAAGTC CAAAAGTC CAAAAGTC CAAAAGTC CAAAAGTC CAAAAGTC	GGCATACCGTT GGCATACCGTT GGCATACCGTT GGCATACCGTT GGCATACCGTT GGCATACCGTT GGCATACCGTT	TGTCTTGTTT TGTCTTGTTT TGTCTTGTTT TGTCTTGTTT TGTCTTGTTT TGTCTTGTTT TGTCTTGTTT	GGTATTGATT GGTATTGATT GGTATTGATT GGTATTGATT GGTATTGATT GGTATTGATT	GACGAATGCT GACGAATGCT GACGAATGCT GACGAATGCT GACGAATGCT GACGAATGCT GACGAATGCT	CAAAAA CAAAAA CAAAAA CAAAAA CAAAAA CAAAAA CAAAAA CAAAAA

**Supplementary Fig. S1** Sequence alignment of the *HSA* integration site for clones used in microarray analysis. Sequence alignment of the region around the *Pme*I directed integration site of clones CV2, CV5, CV7, CV8, CV14, CV15, CV16, CV18 and CV23. *Pme*I recognises and cuts at the sequence GTTTAAAC, located at 467 and highlighted in yellow



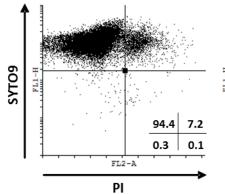
**Supplementary Fig. S2** Growth curves of clonal variation strains. Growth curves were monitored by sampling the cultures every hour and the OD<sub>600</sub> was recorded. Growth in both glycerol and methanol medium was recorded. A) Growth curves in BMGY medium for 25 hours. B) Growth curves in BMMY medium for 8 hours.

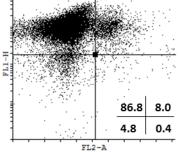


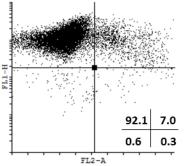


CV15(M)





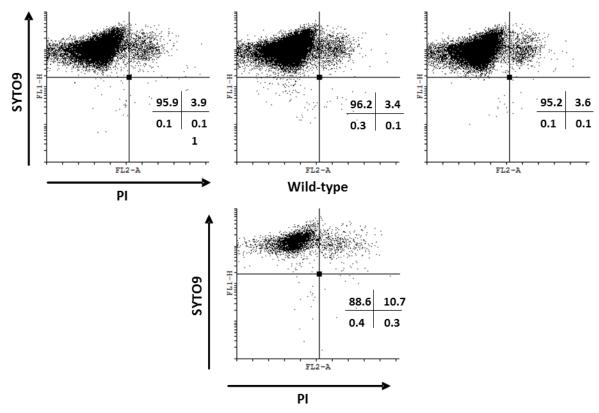






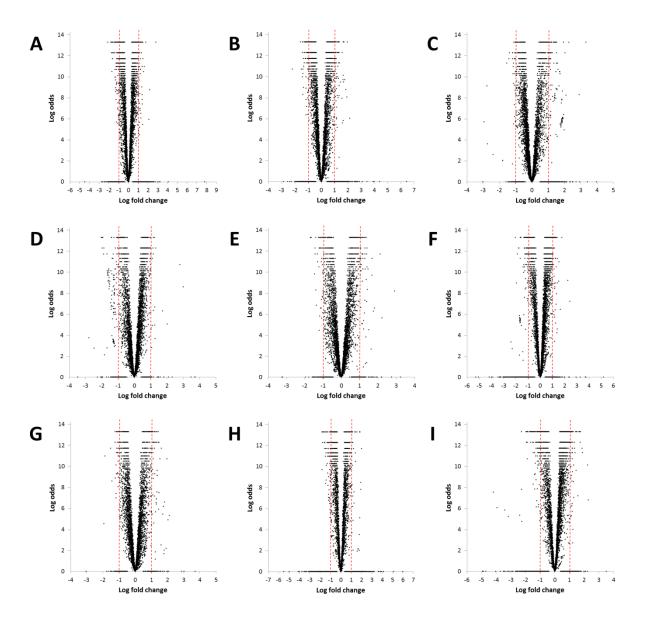




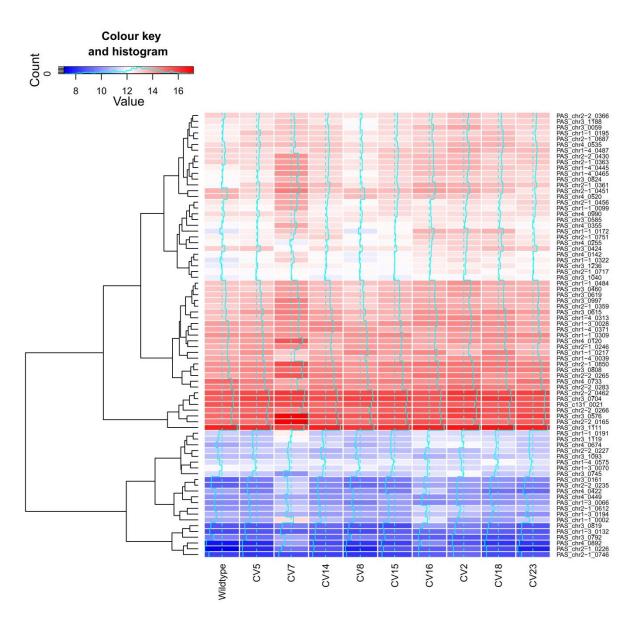


Supplementary Fig. S3 FACS analysis of clonal variants grown for 24 hours in glycerol containing

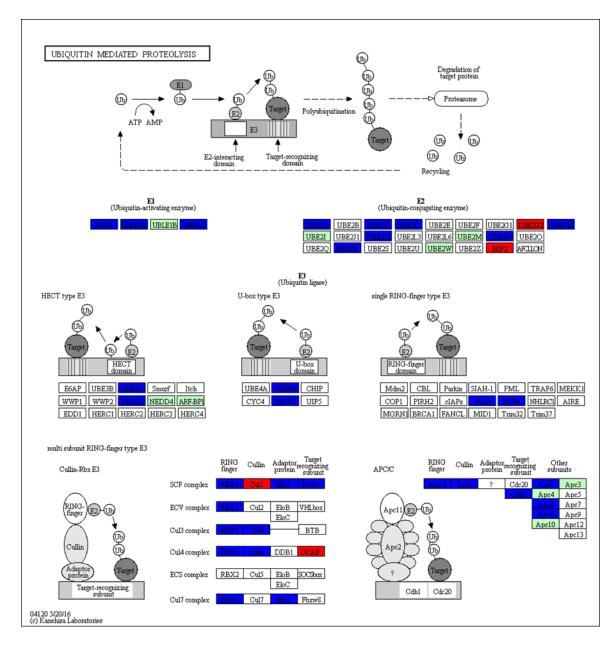
medium. FL1-H: SYTO9<sup>®</sup> a green fluorescent nucleic acid stain indicating live cells, FL2-A: propidium iodide a red fluorescent stain indicating dead cells with a damaged membrane. BG: Background noise. Quadrant displays proportion of live/dead cells and background noise.



**Supplementary Fig. S4** Volcano plots of fold changes vs. adjusted *p* values. A) CV2(H); B) CV18(H); C) CV23(H); D) CV8(M); E) CV15(M); F) CV16(M); G) CV5(L); H) CV7(L); I) CV14(L). Red lines are added for reference to show the boundary of a hypothetical 2 fold change cut-off. No cut-off was used in the analysis presented here, which relied purely on having a false discovery probability of <0.05 (Smyth 2005).



**Supplementary Fig. S5** Expression levels of genes encoding proteins involved in oxidative phosphorylation. Heat maps of log 2 normalised expression levels of genes encoding proteins involved in oxidative phosphorylation, defined by KEGG families, compared to those of wild-type after 24 hour induction with methanol. The associated trees cluster genes with similar expression profiles across all conditions



**Supplementary Fig. S6** Genes up- or downregulated in the ubiquitin mediated proteolysis pathway in CV7(L) compared to wild-type. Microarray data was mapped using KEGG Mapper Search & Colour pathway. Blue genes indicate downregulation compared to wild-type and red genes indicate upregulation. Green boxes indicate organism-specific pathways. The figure was generated using the KEGG Search&Color Pathway programme (Kanehisa and Goto 2000; Kanehisa et al. 2012).

## Supplementary Table S1. Upregulated pathways determined by KOBAS pathway analysis

CV5 (L)	CV7 (L)	CV14 (L)	CV8 (M)	CV15 (M)	CV16 (M)	CV2 (H)	CV18 (H)	CV23 (H)
	Alanine,	Arginine and	Alanine,				Basal	
Fatty acid	aspartate and	proline	aspartate and			Glycerophosph	transcription	Cell cycle -
metabolism	glutamate	metabolism	glutamate	ABC		olipid	factors	yeast
	metabolism		metabolism	transporters	DNA replication	metabolism		
	Aminoacyl-	Biosynthesis of	Arginine and	Arginine and		Nicotinate and	Biosynthesis of	Homologous
Peroxisome	tRNA	secondary	proline	proline	Oxidative	nicotinamide	unsaturated	recombination
	biosynthesis	metabolites	metabolism	metabolism	phosphorylation	metabolism	fatty acids	recombination
Purine	Biosynthesis of	Cysteine and	Biosynthesis of	Biosynthesis of	Porphyrin and		Oxidative	
metabolism	secondary	methionine	secondary	secondary	chlorophyll	Oxidative	phosphorylation	Meiosis - yeast
metabolism	metabolites	metabolism	metabolites	metabolites	metabolism	phosphorylation	priospriorylation	
Ribosome	Cysteine and			Glycolysis /				
biogenesis in	methionine	Peroxisome	Fatty acid	Gluconeogenes	Purine		Peroxisome	Mismatch repair
eukaryotes	metabolism		metabolism	is	metabolism	Proteasome		
RNA	Lysine	Purine	Glycerophosph	THE MACHINE AND A			Ribosome	Oxidative
		metabolism	olipid	Oxidative	Pyrimidine		biogenesis in	phosphorylation
polymerase	biosynthesis	metabolism	metabolism	phosphorylation	metabolism	Ribosome	eukaryotes	phosphorylation
			Pentose and					
Sphingolipid	Metabolic	Pyrimidine	glucuronate			Ribosome	RNA	Peroxisome
metabolism	pathways	metabolism	interconversion			biogenesis in	degradation	Peroxisome
	patriways		s	Peroxisome	Ribosome	eukaryotes	-	
	mRNA		Pentose	Porphyrin and	Ribosome			Phosphatidylino
Spliceosome	surveillance	Ribosome	phosphate	chlorophyll	biogenesis in		Spliceosome	sitol signaling
	pathway		pathway	metabolism	eukaryotes	RNA transport		system
	and another second	Ribosome			·		0 11 1	Ribosome
	Oxidative	biogenesis in			RNA	Steroid	Sulfur relay	biogenesis in
	phosphorylation	eukaryotes	Peroxisome	Ribosome	polymerase	biosynthesis	system	eukaryotes
				Ribosome				
	Peroxisome	RNA	Phenylalanine	biogenesis in		Tyrosine		RNA
		polymerase	metabolism	eukaryotes	RNA transport	metabolism		degradation
		1.	Porphyrin and					
	Phenylalanine	RNA transport	chlorophyll					Sphingolipid
	metabolism		metabolism	RNA transport				metabolism
	B							
	Phenylalanine,	Ubiquinone and						
	tyrosine and	other terpenoid-						Spliceosome
	tryptophan	quinone	Purine	Sulfur				
	biosynthesis	biosynthesis	metabolism	metabolism				
						-		
				Ubiquinone and				Ubiquinone and
	Ribosome		Ribosome	other terpenoid-				other terpenoid-
			biogenesis in	quinone				quinone
			eukaryotes	biosynthesis				biosynthesis
	Sulfur							
	metabolism		Spliceosome					
	Terpenoid							
	backbone							
	biosynthesis							
	Tyrosine							
	metabolism							
	Ubiguinone and							
	other terpenoid-							
	quinone							
	biosynthesis							
	Valine, leucine							
	and isoleucine							
	biosynthesis	1	1	1	1	1	1	1

## Supplementary Table S2 Downregulated pathways determined by KOBAS pathway analysis

CV5 (L)	CV7 (L)	CV14 (L)	CV8 (M)	CV15 (M)	CV16 (M)	CV2 (H)	CV18 (H)	CV23 (H)
Aminoacyl- tRNA biosynthesis	Basal transcription factors	Amino sugar and nucleotide sugar metabolism	Cell cycle - yeast	Basal transcription factors	Aminoacyl- tRNA biosynthesis	Alanine, aspartate and glutamate metabolism	Alanine, aspartate and glutamate metabolism	Basal transcription factors
Fatty acid biosynthesis	Base excision repair	Basal transcription factors	MAPK signaling pathway - yeast		Cyanoamino acid metabolism	Cell cycle - yeast	Amino sugar and nucleotide sugar metabolism	Endocytosis
Glutathione	Cell cycle -	Cell cycle -	pairway youst	One carbon		Fatty acid	Aminoacyl- tRNA	
metabolism	yeast	yeast	Mismatch repair	pool by folate	Endocytosis	biosynthesis	biosynthesis	Proteasome
Glycine, serine and threonine metabolism Phenylalanine,	DNA replication	Endocytosis	Protein processing in endoplasmic reticulum	Proteasome	Glycine, serine and threonine metabolism	Lysine biosynthesis	Biosynthesis of secondary metabolites	Protein processing in endoplasmic reticulum
tyrosine and tryptophan biosynthesis Protein	Endocytosis	Galactose metabolism	Ubiquitin mediated proteolysis	Riboflavin metabolism SNARE	Lysine biosynthesis	MAPK signaling pathway - yeast	Glycine, serine and threonine metabolism	Riboflavin metabolism SNARE
processing in endoplasmic reticulum	Folate biosynthesis	Glutathione metabolism	Various types of N-glycan biosynthesis	interactions in vesicular transport	MAPK signaling pathway - yeast		Lysine biosynthesis	interactions in vesicular transport
Riboflavin metabolism	Homologous recombination	MAPK signaling pathway - yeast		Ubiquitin mediated proteolysis	Mismatch repair	Pentose phosphate pathway Phenylalanine,	MAPK signaling pathway - yeast	Steroid biosynthesis
Ubiquitin mediated proteolysis	MAPK signaling pathway - yeast	Mismatch repair			Natural killer cell mediated cytotoxicity	tyrosine and tryptophan biosynthesis	Natural killer cell mediated cytotoxicity	
		Natural killer cell mediated			N-Glycan	Starch and sucrose	Nitrogen	
	Meiosis - yeast	cytotoxicity N-Glycan			biosynthesis Nucleotide	metabolism Vitamin B6	metabolism One carbon	
	Mismatch repair				excision repair	metabolism	pool by folate	
	Natural killer cell mediated cytotoxicity	Non- homologous end-joining			Phenylalanine, tyrosine and tryptophan biosynthesis		Phenylalanine metabolism Phenylalanine,	
	N-Glycan biosynthesis	Nucleotide excision repair Protein			Regulation of autophagy		tyrosine and tryptophan biosynthesis Protein	
	Nucleotide excision repair Phosphatidylino	processing in endoplasmic reticulum			Starch and sucrose metabolism		processing in endoplasmic reticulum	
	sitol signaling	Regulation of autophagy SNARE			Ubiquitin mediated proteolysis		Riboflavin metabolism	
	Proteasome Protein	interactions in vesicular transport					Starch and sucrose metabolism	
	processing in endoplasmic reticulum	Starch and sucrose metabolism					Steroid biosynthesis	
	Regulation of autophagy	Ubiquitin mediated proteolysis					Tyrosine metabolism Ubiquitin	
	RNA degradation SNARE						mediated proteolysis	
	interactions in vesicular transport							
	Ubiquitin mediated proteolysis Various types							
	of N-glycan biosynthesis							