Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity
Rita Gemayel, Sreenivas Chavali, Ksenia Pougach, Matthieu Legendre, Bo Zhu, Steven Boeynaems, Elisa van der Zande, Kris Gevaert, Frederic Rousseau, Joost Schymkowitz, M. Madan Babu, and Kevin J. Verstrepen
Figure S1

A

| Few million years | Few hundred years | Few weeks | Isogenic cells at an instant |
|-------------------|-------------------|-----------|-----------------------------|
| **Expression divergence** | **Expression variability** | **Mutational variance** | **Expression noise** |

- S. cerevisiae (BY4743)
- S. paradoxus (CBS 432)
- S. paradoxus (NRLY-17217)
- S. mikatae (IFO1815)
- S. kudriavzevii (IFO1802)

- BY4716
- RM11-1a

B

| Expression divergence | Expression variability | Mutational variance | Expression noise |
|-----------------------|------------------------|--------------------|-----------------|
| n=1903 [n=31]        | n=1972 [n=21]          | n=890 [n=22]       | n=163 [n=4]    |
| CLES 45.9%            | 37.5%                  | 49.3%              | NA              |

Target of NR-TFs with low expression phenotype
Target of Q-rich TFs with low expression phenotype
Figure S2

A

Expression divergence

Expression variability

Mutational variance

Expression noise

CLES

42.4%

Expression divergence

Expression variability

Mutational variance

Expression noise

CLES

44.1%

B
Regulation of cellular protein metabolic process (GO:0032268)
Posttranscriptional regulation of gene expression (GO:0010608)
Intracellular transport (GO:0046907)
Regulation of translation (GO:0006417)
Oxidation reduction (GO:0055114)
Translational elongation (GO:0006414)
Steroid metabolic process (GO:0008202)
Sterol metabolic process (GO:0016125)
Cofactor metabolic process (GO:0051186)

-log P-value

0 0.5 1 1.5 2 2.5 3 3.5

Targets of NR-TFs
Targets of Q-rich TFs

C

Expression divergence

Expression variability

Mutational variance

Expression noise

CLES

42.4%

Expression divergence

Expression variability

Mutational variance

Expression noise

CLES

43.0%

D

Expression divergence

Expression variability

Mutational variance

Expression noise

CLES

44.1%

Expression divergence

Expression variability

Mutational variance

Expression noise

CLES

44.9%
Figure S3

A

\[SW1 ORF\]

\[
\begin{array}{c}
1 & 348 & 372 & 1314 \\
N & Q & C
\end{array}
\]

B

\[SNF5 ORF\]

\[
\begin{array}{c}
1 & 218 & 268 & 905 \\
N & Q & C
\end{array}
\]
Figure S4

A  Carbon starvation

| RNA-seq | RT-qPCR | MLS1 |
|---------|---------|------|
| ![Graph](image1) | ![Graph](image2) | ![Graph](image3) |
| $R^2 = 0.892$ | $R^2 = 0.778$ |

| RNA-seq | RT-qPCR | HXX1 |
|---------|---------|------|
| ![Graph](image4) | ![Graph](image5) | ![Graph](image6) |
| $R^2 = 0.43$ | $R^2 = 0.805$ |

| FOX2 | FLO11 | NRG1 |
|------|------|------|
| ![Graph](image7) | ![Graph](image8) | ![Graph](image9) |
| $R^2 = 0.772$ | $R^2 = 0.744$ | $R^2 = 0.869$ |

| CIN5 | FLO11 |
|------|------|
| ![Graph](image10) | ![Graph](image11) |
| $R^2 = 0.835$ | $R^2 = 0.757$ |

B  Glucose rich

| RNA-seq | RT-qPCR |
|---------|---------|
| ![Graph](image12) | ![Graph](image13) |
| $R^2 = 0.743$ | $R^2 = 0.805$ |

| CIN5 | FLO11 |
|------|------|
| ![Graph](image14) | ![Graph](image15) |
| $R^2 = 0.835$ | $R^2 = 0.757$ |

B  Glucose-rich medium

| Carbon-starved medium | Glucose-rich medium |
|-----------------------|---------------------|
| ![Graph](image16) | ![Graph](image17) |
| $R^2 = 0.743$ | $R^2 = 0.805$ |

SSN6 ORF:

- **SSN6 TR2 Number**
  - primers 1
  - primers 2

- **SSN6 ORF**
  - primers 1
  - primers 2

| **SSN6 TR2 Number** | **SSN6 TR2 Number** |
|---------------------|---------------------|
| 0 20 40 60 80 100 120 | 0 20 40 60 80 100 120 |
| **SSN6 ORF** N | **SSN6 ORF** C |
| **TR1** | **TR2** |
Figure S5

A

**PHO84**

|        | Glucose | Glycerol |
|--------|---------|----------|
| 0Q     |         |          |
| 51Q (WT)| **+**   | **+**    |

**ZEO1**

|        | Glucose | Glycerol |
|--------|---------|----------|
| 0Q     | 10      | 20       |
| 51Q (WT)| 15     | 25       |

B

pho84::YFP fluorescence (a.u.)

10µM Pi  50µM Pi  100µM Pi  250µM Pi  10mM Pi

0Q 51Q
Figure S6

A

MW markers

TR2 number 0 63 105 Controls

TR2 number 0 63 105 Controls

B

TR2-105

Parent 1

TR2-0

Parent 2

Dissected tetrads

Tetrads showing 2:2 segregation of phenotype

31

31

31

31

C

103Q-GFP fluorescence (a.u.)

flo11::CFP fluorescence (a.u.)

replicate 1

replicate 2

replicate 3

replicate 4

pgPD 103Q-GFP

no plasmid

pgPD 103Q-GFP

no plasmid

pgPD 103Q-GFP

no plasmid

pgPD 103Q-GFP

no plasmid
SUPPLEMENTAL INFORMATION

SUPPLEMENTAL DATA

Figure S1: Expression variation of targets of Q-rich TFs across different time-scales is independent of that of the TFs, Related to Figure 1

(A) Expression variation of Q-rich TFs and non-repeat containing TFs are comparable across different time-scales. As the number of TFs were low, we could not compare the expression variation distributions of Q-rich TFs and nonRCP TFs using Wilcoxon rank-sum test. Therefore, we categorized each expression variation measure into low (bottom 33.3%), medium (middle 33.3%) and high (top 33.3%) using tertile-cuts of the distribution of all genes. Subsequently, we counted the number of nonRCP TFs and Q-rich TFs in each category and assessed for differences in their distribution using Fisher’s exact test. Though, predominantly Q-rich TFs tend to show low expression variation, these differences are not statistically significant owing to few datapoints. Therefore, we interpret that Q-rich TFs and non-repeat containing TFs have comparable expression variation across different time-scales.

(B) Distribution of expression divergence, expression variability, mutational variance and expression noise of targets regulated by NR-TFs and Q-rich TFs, which had low expression variation across different time scales (as defined in Figure S1A). The number of targets in each class is provided below each box, and the number of TFs that qualified our criterion are between brackets. Statistical significance was assessed using Wilcoxon rank sum test. The effect sizes are represented by the common language effect size (CLES) statistic.

Figure S2: Expression variation of targets of Q-rich TFs across different time-scales is independent of endogenous and exogenous conditions, Related to Figure 1
Distribution of transcript abundance of targets of NR-TFs and Q-rich TFs in yeast grown in YPD (Holstege et al., 1998). Both classes of targets have comparable expression levels ruling out the possibility of measurement errors influencing our observations. P-values were estimated using Wilcoxon rank sum test.

Gene Ontology biological process enrichment among targets of NR-TFs and Q-rich TFs obtained using DAVID server (Huang da et al., 2009). FDR values are presented as –log P values.

Distribution of stress-responsive targets (Luscombe et al., 2004; Gasch et al., 2000) of NR-TFs and Q-rich TFs for expression variation across different time-scales. Statistical significance was assessed using Wilcoxon rank sum test.

Distribution of nonstress-responsive targets of NR-TFs and Q-rich TFs for expression variation across different time-scales. All targets that were not found to be stress-response genes in Figure S2C were classified as nonstress-response genes. Statistical significance was assessed using Wilcoxon rank sum test. The effect sizes are represented by the common language effect size (CLES) statistic.

Figure S3. The Q-rich repeats in other S. cerevisiae transcriptional regulators show variability between natural strains, Related to Figure 2

Schematic representation of Swi1 (a subunit of the SWI/SNF chromatin remodeling complex) showing the Q-rich repeat region (residues 348 to 372). Amplification of the Q-rich region of SWI1 from various S. cerevisiae isolates.

Schematic representation of Snf5 (another subunit of the SWI/SNF chromatin remodeling complex) showing the Q-rich repeat region (residues 218 to 268). Amplification of the Q-rich region of SNF5 from various S. cerevisiae isolates.
Figure S4. Confirmation of changes in expression of the SSN6 targets by real-time quantitative PCR, Related to Figure 3

(A) Expression profiles for representative genes, identified by RNAseq, were confirmed by real-time quantitative PCR. Log$_2$ expression fold changes relative to the WT (TR2-63) are given. Data points represent mean ± SD; n=2.

(B) The expression of SSN6 in all the TR2 variants was measured by real-time quantitative PCR using primer pairs annealing before the TR2 region (primers 1) and after the TR2 region (primers 2). SSN6 expression was normalized to the ACT1 levels and the Log$_2$ expression fold change relative to the WT (TR2-63) is given. Data points represent mean ± SD; n=2.

Figure S5. Loss of polyQ in the chromatin modifier Snf5 results in loss of transcriptional activity, Related to Figure 4

(A) Expression of PHO84 and ZEO1 was measured in Snf5 polyQ deletion variant (0Q) and in the WT (51Q) by real-time quantitative PCR. Cultures were grown until exponential phase in rich medium containing 2% glucose or 2% glycerol as a carbon source. Expression values were normalized to the expression of the RPS16A gene. Data points represent mean ± SD; n=3. A two-tailed, two sample unequal variance t-test was used to determine P values.

(B) Flow cytometry profiles of pho84::YFP expression in the Snf5 polyQ deletion variant (0Q) (green traces) and in the WT strain (51Q) (blue traces) in SC-glucose medium containing various phosphate concentrations.

Figure S6. Expanded SSN6 repeats are stable over multiple generations, the Ssn6 TR2-105 aggregates are not prions and overexpression of an aggregation-prone polyQ does not recapitulate the observations made with expanded Ssn6, Related to Figure 5
(A) Two biological replicates (starting clones 1-2) and two technical replicates (a–b) of the SSN6 TR2-105 variant were cultured over > 30 generations using 4 intermediate dilutions. Genomic DNA was extracted from all the cultures and the SSN6 TR2 region was amplified by PCR.

(B) The SSN6 TR2-105 variant was crossed with either the TR2-0 variant or the WT (TR2-63) strain. The resulting diploids were allowed to sporulate and 31 tetrads resulting from each cross were dissected to obtain separate spores. This progeny was then spotted on YP-sucrose medium and colonies were scored for their morphology. A Mendelian, 2:2 segregation of the colony morphology phenotype was obtained for all tetrads.

(C) Expression of the N-terminal fragment of human huntingtin containing 103 glutamines controlled by the GPD promoter in WT cells (SSN6 TR2-63) with a flo11::CFP reporter. GFP and CFP fluorescence from 4 independent replicates was measured by flow cytometry. Cells expressing the 103Q-GFP fragment (y-axis) have similar flo11::CFP (x-axis) expression compared to control (no plasmid) cells.
Table S1. Genome-scale analysis of the occurrence of Q-rich repeats in various eukaryotic organisms, Related to Figure 1

Tandem repeats of $\geq 9$ units were first extracted from the genomes of *S. cerevisiae*, *D. melanogaster*, *D. rario*, *M. musculus*, and *H. sapiens*. From this list we subsequently extracted Q-rich repeats, defined as repeats with at least 85 % of glutamines in the translated sequence. Systematic gene names are given.

This table provided as a separate Excel file.

Table S2. Functional enrichment of genes containing Q-rich repeats in various eukaryotic organisms, Related to Figure 1

Genes containing Q-rich repeats are enriched for particular biological functions (column B) when compared to all repeat-containing genes in various eukaryotic genomes. The statistical significance value (adjusted $P$-value) for each enriched category is given in column C.

This table is provided as a separate Excel file.
Table S3. Sequencing of SSN6 tandem repeats in various *Saccharomyces cerevisiae* strains, Related to Figure 2

Q, glutamine; A, alanine

| Strain                     | TR1 | TR2 |
|----------------------------|-----|-----|
|                            | Unit number (Q) | Repeat length (bp) | Unit number (QA) | Unit number (Q) | Unit number (Q+A) | Repeat length (bp) |
| SK1                        | 20  | 60  | 22  | 9  | 31  | 159               |
| NCYC110                    | 20  | 60  | 22  | 9  | 31  | 159               |
| DBVPG6044                  | 20  | 60  | 23  | 9  | 32  | 165               |
| Y12                        | 11  | 33  | 25  | 26 | 51  | 228               |
| Y9                         | 11  | 33  | 25  | 26 | 51  | 228               |
| L-1374                     | 20  | 60  | 36  | 23 | 59  | 285               |
| DBVPG1373                  | 16  | 48  | 36  | 24 | 60  | 288               |
| YS2                        | 16  | 48  | 36  | 24 | 60  | 288               |
| DBVPG1853                  | 16  | 48  | 37  | 24 | 61  | 294               |
| Σ1278b                     | 16  | 48  | 32  | 31 | 63  | 285               |
| S288c                      | 16  | 48  | 32  | 31 | 63  | 285               |
| UWOPS83-787.3             | 16  | 48  | 34  | 31 | 65  | 297               |
| 378804X                    | 16  | 48  | 34  | 32 | 66  | 300               |
| UWOPS03-461.4             | 16  | 48  | 34  | 32 | 66  | 300               |
| UWOPS05-227.2             | 16  | 48  | 34  | 32 | 66  | 300               |
| YJM 789                    | 16  | 48  | 40  | 28 | 68  | 324               |
Table S4. Target genes showing significant induction or repression in the SSN6 TR2-0 or TR2-105 variants in carbon starved or glucose-rich medium, Related to Figure 3

The data is represented as Log$_2$ fold change in expression relative to the expression in the WT strain (TR2-63). A cutoff of Log$_2$ fold change ≥ 0.8, FDR-adjusted $P< 0.01$, in at least one growth condition, was used to select the significantly up- or down-regulated genes in these variants.

This table is provided as a separate Excel file.
Table S5. Interactors of Ssn6 identified by LC-MS/MS, Related to Figure 5
These proteins were detected in the pull-downs of every Ssn6 TR2 variant tested (i.e. TR2-0, TR2-55, TR2-63, TR2-90 and TR2-105) by at least two different peptides. In bold are the proteins that show enrichment or depletion relative to the WT (TR2-63) in one or more TR2 variant. Subcellular localization predicted by pSORT.

| Gene name | Description | Subcellular localization |
|-----------|-------------|--------------------------|
| 1 SSN6    | General transcriptional corepressor CYC8 | nuclear |
| 2 TUP1    | General transcriptional corepressor TUP1 | nuclear |
| 3 SSA2    | Heat shock protein SSA2 | cytosolic |
| 4 SAR1    | Component of COPII coat of vesicles | cytosolic |
| 5 PAB1    | Polyadenylate-binding protein, cytoplasmic and nuclear | cytosolic/nuclear |
| 6 TEF1    | Elongation factor 1-alpha | cytosolic |
| 7 DED1    | ATP-dependent RNA helicase DED1 | nuclear |
| 8 RPS19A,RPS19B | 40S ribosomal protein S19A, S19B | cytosolic |
| 9 RPL2B,RPL2A | 40S ribosomal protein S1-A, S1-B | cytosolic |
| 10 RPL3   | 60S ribosomal protein L3 | cytosolic |
| 11 HRP1   | Nuclear polyadenylated RNA-binding protein 4 | nuclear |
| 12 NPL3   | RNA-binding protein, Nucleolar protein 3 | cytosolic |
| 13 NOP1   | Histone glutamine methyltransferase | cytosolic |
| 14 YDJ1   | Yeast dnaJ protein 1 | cytosolic |
| 15 ACH1   | Acetyl-CoA hydrolase | cytosolic |
| 16 PSP2   | Asparagine-rich cytoplasmic protein | cytosolic/nuclear |
| 17 ADE12  | Adenylosuccinate synthetase | cytosolic |
SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Genome-Scale Analyses

Genomic sequences and gene annotations for *Danio rerio* (assembly Zv9, r73), *Drosophila melanogaster* (r5.53), *Homo sapiens* (GRCh37, r73), *Mus musculus* (GRCm38, r73) and *Saccharomyces cerevisiae* (S288c, r64) were downloaded from the Ensembl database (Flicek et al., 2013). We scanned these genomes with Tandem Repeat Finder (TRF) v4.07b (Benson, 1999) with standard parameters. We then selected repeats overlapping protein-coding regions. Repeats coding for glutamines were flagged as “Q repeats” if glutamines composed at least 50% of the translated sequence. Repeats with at least 85% of glutamines in the translated sequence were annotated as “Q-rich repeats”. Gene ontology enrichment analyses were done using the GoStat tool (Beissbarth and Speed, 2004). For each species, genes were split into 4 categories: all genes, genes with repeats, genes with Q repeats and genes with Q-rich repeats. We searched for enriched Biological Process GO terms with significant corrected P-value (FDR P-value ≤ 0.05) by comparing genes with Q repeats to genes with repeats. The trends were then confirmed by computing the proportion of genes in the other categories for significantly enriched GO terms. Functional annotations for the targets of SSN6 repeat variants were determined using the clustering algorithm of DAVID v6.7 (Huang da et al., 2009) with the default stringency settings.

Construction of the Yeast SSN6 Repeat Variants

To create the SSN6 repeat variants, we first amplified the *Hph* gene, conferring resistance to hygromycin B, using primers 2361-2362 which have 60 bp overhangs that allow homologous recombination of the PCR product downstream of the SSN6 open reading frame in the Sigma1278b strain. We then used the genomic DNA from the resulting strain (RG518, WT
strain), the reverse primer 2362 and a series of forward primers (2363, 2364, 2857 till 2860, 4126, 4127, 4256, 4257, 4258) each designed to produce a PCR product with a different repeat number. These PCR products were used to transform a fresh Sigma1278b strain and transformants were selected for hygromycin B resistance. To generate a Snf5 polyQ deletion variant in the lab S288c strain, we used the same strategy described above and the following primers 2355, 2356, 2357. The repeat regions of the newly created variants were verified by sequencing with primers flanking the repeats. To create YFP tagged versions of the Ssn6 repeat variants we amplified a cassette containing yEVenus-kanR using primers 4037-4038 and plasmid pKT103 (Euroscarf). To create 6xHA-tagged versions of the Ssn6 repeat variants, we amplified a 6xHA-kanR cassette with primers 4491-4492 and plasmid pYM14 (Euroscarf). To make flo11::YFP or CFP promoter fusions we used primers 4356-4357 and plasmid pKT103 (YFP) or pKT102 (CFP), for IMA1-YFP protein fusions we used primers KP6 – KP7 and plasmid pKT103, and for CIN5-mCherry protein fusions we used primers Cin5-yEVL-F and Cin5-yEVL-R and plasmid pSR101. We then transformed the SSN6 repeat variants and selected transformants for the adequate resistance. The pho84::YFP (promoter fusion) strains were generated by inserting a cassette containing the native PHO84 promoter followed by a yEVenus coding sequence at a neutral genomic locus. These cells contain a functional PHO84 copy (in its native locus) to maintain physiological responses to varying Pi levels and prevent constitutive expression of the PHO genes.

**Media and Growth Conditions**

Unless otherwise stated, strains were grown in liquid YP medium (1% yeast extract, 2% peptone) supplemented with 4% glucose or on YPD (2% glucose) plates. Flo11::YFP reporter strains were grown in YP 4% glucose, IMA1-YFP reporter strains in YP 2% palatinose and CIN5-RFP reporter strains in YP 0.5% glucose. Strains harboring plasmid p416 103Q GPD
(Addgene) were grown in SC-URA medium (for plasmid selection) supplemented with 4% glucose. Snf5 polyQ deletion variant containing *pho84::YFP* reporter was grown in SC 2% glucose medium containing various phosphate concentrations prepared as described (Thomas and O'Shea, 2005). To avoid depletion of extracellular phosphate, the cells were periodically diluted to obtain a density of $5 \times 10^5$ cells/ml after 18 hrs of growth. The remaining phosphate in the medium was measured at the end of the experiment and was > 80 % of the initial concentration.

**Flow Cytometry**

Strains were grown until exponential phase in appropriate medium according to the reporter genes (see above). Analytical flow cytometry was performed using a BD Influx flow cytometer equipped with the appropriate lasers and filters. Before each analysis, we ran stable multicolored fluorescent beads in order to standardize the instrument and allow for correct comparisons between experiments.

**Yeast Phenotypic Analyses**

We used the Phenotype MicroArray technology (BIOLOG, Inc.) for high throughput screening of growth in various carbon and nitrogen sources following the manufacturer’s instructions. Growth kinetics were extracted from the growth plots using the OmniLog PM software. Growth rate measurements in palatinose were done using the Bioscreen C (Growthcurves USA) as in (New et al., 2014). To evaluate colony morphology, strains were grown overnight in liquid YPD, normalized to the same density and spotted on YP 2% Sucrose plates. After 4 days of growth at 30°C, the plates were photographed at 0.5 X magnification using a Nikon AZ 100M macroscope with a DS-Ri1 camera. To evaluate invasive growth, strains were pre-grown and normalized as previously and spotted on YPD
plates. After 11 days at 30°C, the plates were photographed before and after removal non-invasive cells by washing under running water.

Adhesion to plastic was quantified by staining with crystal violet as described elsewhere (Reynolds and Fink, 2001) and flocculation measured as follows: dense overnight cultures were normalized to OD$_{600} = 10$. Cultures were then left to sediment and at multiple time points 20 µl of culture was taken from a fixed sampling point and OD$_{600}$ measured. Quantification was based on counts of free cells in a flocculating culture over total cell density using the formula: [1- (free cells/total cells)] x 100.

**Transcriptome Analysis by RNA-seq and Real-Time Quantitative PCR**

RNA sample preparation and sequencing on an Illumina Hiseq2000, which generated single-end 50 bp reads, was carried out at the Genomics Core Facility (EMBL – Heidelberg). To ensure data accuracy for the subsequent analyses, a minimum of 50X coverage was required. Raw reads were filtered using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) with Q30 cutoff. After removing the low quality reads and adaptors, the RNA-Seq reads were aligned to the *S. cerevisiae* S288C reference genome (version genebank 64) using Tophat 2.0.7 (Trapnell et al., 2009) allowing a maximum of 2 mismatches. If reads mapped to more than one location, only the one with the highest score was kept. After obtaining the reads number for every sample, the edgeR Bioconductor package with the Trimmed Mean of M-values (TMM) normalization method (Robinson and Oshlack, 2010) was used to determine the differentially expressed genes (DEGs). Significantly differentially expressed genes ($P$ value < 0.01 and Log$_2$ fold change > 0.8) in the TR2-0 or the TR2-105 variants were selected (Table S4). To detect meaningful expression trends between all the TR2 variants, we treated the *SSN6* TR2 numbers as a time series and analyzed the autocorrelation of the expression pattern of the above selected genes.
The expression fold change (compared to WT, TR2-63) trajectories of different SSN6 TR2 variants are expected to be autocorrelated, whereas those of false positives should be uncorrelated between two consecutive repeat variants. We rejected DEGs with an autocorrelation coefficient < 0.2. The GENE-E software (http://www.broadinstitute.org/cancer/software/GENE-E/) was used to generate a heat map for the expression fold change of genes fulfilling the above criteria.

For quantitative real-time PCR measurements, complementary DNA was synthesized from 1 µg total RNA using the QuantiTect Reverse Transcription (Qiagen). Real-time quantitative PCR was performed using the StepOnePlus system (Applied Biosystems) and Power SYBR Green PCR master mix (Applied Biosystems). The PCR reaction conditions were as follows: 10 min at 95°C, followed by 40 cycles of 95°C for 15 sec, 60°C for 1 min, and 95°C-60°C (melt curve). Expression values were normalized to the expression of the ACT1 gene (SSN6 targets) or the RPS16A gene (SNF5 targets).

Network Construction and Visualization

Protein-protein interaction networks were extracted from the STRING interaction database (Franceschini et al., 2013), protein-DNA interactions from YEASTRACT (Teixeira et al., 2014) and metabolic networks from the KEGG database (Kanehisa et al., 2014). Network analysis and visualization were performed using Cytoscape (Smoot et al., 2011).

Identification and Quantification of the Ssn6 Interactome

Peptides originating from the ‘in solution’ endoproteinase-LysC digestion of the pull-downs were labelled using N-hydroxysuccinimide esters of different versions of propionic acid: 12C₃-propionate (Light isotope, L) for the WT pull-downs and 13C₃-propionate (Heavy isotope, H) for the TR2 variants, as described (Ghesquiere et al., 2011). Following acidification to pH 3
by adding TFA and a centrifugation step for 5 min at 16,000g to remove insolubilities, equal amounts of WT and TR2 variants samples were mixed. Of these peptide mixtures, 5 µL was analyzed by LC-MS/MS on a Q Exactive mass spectrometer as described (Stes et al., 2014). Subsequent peptide identification and quantification was performed with the MaxQuant software (Cox and Mann, 2008) (version 1.4.0.3) in the Swiss-Prot database with restriction to *Saccharomyces cerevisiae* proteins. Here, acetylation of the protein N-terminus, pyroglutamic acid formation of N-terminal glutamine and methionine oxidation were set as possible modifications. Mass tolerances of peptide precursor ions and fragments ions were set to 4.5 ppm and 20 ppm, respectively. The false discovery rate on peptide-to-spectrum matches was set at 0.01, and a minimum peptide length of 7, the option “matching between all runs” and a minimum of 2 unique or razor peptides for protein identification were used. Peptides from the trypsin in-gel digestion were analyzed by MS/MS on an Orbitrap XL mass spectrometer and identified against the Swiss-Prot *Saccharomyces cerevisiae* proteins.

**Fluorescence Imaging**

Fluorescence imaging of live cells was carried out using an inverted Nikon Eclipse Ti microscope equipped with a 100x oil-immersion objective, a Lambda XL fluorescent lamp (Sutter Instrument Company), an Andor DL-604M - #VP camera (Andor™ Technology) and MetaMorph software (Molecular Devices) for image acquisition.
## Primers used in this study

| Primer name | Primer sequence |
|-------------|-----------------|
| **Primers for SSN6 repeat sequencing** | |
| 2332-TRsz-CYC8-F2 | TCAATGGTACAACAACAGCATCCTGCCTGCTAA |
| 2333-TRsz-CYC8-R2 | TATGGTTGCCCTTGTAGGATTTAATCAAT |
| 2531-TR1sz-CYC8-F1 | AGACTAGTACTACAATACAACAGCA |
| 2532-TR1sz-CYC8-R1 | GTTTCTGCAAAAAGAAGCAGAAT |
| **Primers for SSN6 repeat variant construction** | |
| 2361-dTR-CYC8-HYG-F | AACGAGAAAATGTTGTACATCACAATTTCATTATGTTTGAGATA |
| 2362-dTR-CYC8-HYG-R | TTTGTGATTACAATCCCATATGTCAATTTCCCATTCAGAGGAG |
| 2363-dTR1-CYC8-F | CAAGCTTCGTTCTCAGAGTCAGGGCAGCTCCCATGAGCAGCAAC |
| 2364-dTR2-CYC8-F | GCAATCCTCGTCAAGAAAAGCTGATACTTCATTAGAGAGGAGAAG |
| 2857-dTR5Q-CYC8-F | AAAATAATGGAGCAACAAGCTAGAGCAGCTTCAGAGGAGGAG |
| 2858-dTR10Q-CYC8-F | GGGCTGACAAATGATAATGGAACAAAGCTAGAGCAGCTTCAGAGGAG |
| 2859-TR32Q-CYC8-F | ACAAGCCAAAGCTAGAGCAGCTTCAGAGGAGGAG |
| 2860-TR51Q-CYC8-F | ACAAGCCAAAGCTAGAGCAGCTTCAGAGGAGGAG |
| 4126-TR2-64Q-CYC8-F | ACAAGCCAAAGCTAGAGCAGCTTCAGAGGAGGAG |
| 4127-TR2-48QA-CYC8-F | ACAAGCCAAAGCTAGAGCAGCTTCAGAGGAGGAG |
| 4256-SB-TR2-6QA-CYC8-F | ACAAGCCAAAGCTAGAGCAGCTTCAGAGGAGGAG |
| 4257-SB-TR2-12QA-CYC8-F | ACAAGCCAAAGCTAGAGCAGCTTCAGAGGAGGAG |
| 4258-SB-TR2-24QA-CYC8-F | ACAAGCCAAAGCTAGAGCAGCTTCAGAGGAGGAG |

**Primers for tagged Ssn6p variant construction**

| Primer name | Primer sequence |
|-------------|-----------------|
| 2361-dTR-CYC8-HYG-F | AACGAGAAAATGTTGTACATCACAATTTCATTATGTTTGAGATA |
| 2362-dTR-CYC8-HYG-R | TTTGTGATTACAATCCCATATGTCAATTTCCCATTCAGAGGAG |
| 2363-dTR1-CYC8-F | CAAGCTTCGTTCTCAGAGTCAGGGCAGCTCCCATGAGCAGCAAC |
| 2364-dTR2-CYC8-F | GCAATCCTCGTCAAGAAAAGCTGATACTTCATTAGAGAGGAGAAG |
| 2857-dTR5Q-CYC8-F | AAAATAATGGAGCAACAAGCTAGAGCAGCTTCAGAGGAGGAG |
| 2858-dTR10Q-CYC8-F | GGGCTGACAAATGATAATGGAACAAAGCTAGAGCAGCTTCAGAGGAG |
| 2859-TR32Q-CYC8-F | ACAAGCCAAAGCTAGAGCAGCTTCAGAGGAGGAG |
| 2860-TR51Q-CYC8-F | ACAAGCCAAAGCTAGAGCAGCTTCAGAGGAGGAG |
| 4126-TR2-64Q-CYC8-F | ACAAGCCAAAGCTAGAGCAGCTTCAGAGGAGGAG |
| 4127-TR2-48QA-CYC8-F | ACAAGCCAAAGCTAGAGCAGCTTCAGAGGAGGAG |
| 4256-SB-TR2-6QA-CYC8-F | ACAAGCCAAAGCTAGAGCAGCTTCAGAGGAGGAG |
| 4257-SB-TR2-12QA-CYC8-F | ACAAGCCAAAGCTAGAGCAGCTTCAGAGGAGGAG |
| 4258-SB-TR2-24QA-CYC8-F | ACAAGCCAAAGCTAGAGCAGCTTCAGAGGAGGAG |
4037-RG-CYC8-yEVL-F

4038-RG-CYC8-yEVL-R

4491-RG-CYC8-6xHA-F

4492-RG-CYC8-6xHA-R

Primer to create QA to P variants

4553-RG-SSN6-QAtoP-F

Primer to create QA to P variants

2322-TRsz-SNF5-F1

2323-TRsz-SNF5-R1

Primers for SNF5 repeat deletion construction

2355-dTR-SNF5-HYG-F

2356-dTR-SNF5-HYG-R

2357-dTR-SNF5-F

Primer to create target promoter or protein–YFP or RFP fusions

4356-FLO11-yEVL-F

4357-FLO11-yEVL-R

RG-Cin5-yEVL-F

RG-Cin5-yEVL-R

KP6

KP7

3440-PHO84-prom3-F

3441-PHO84pré-yEVL3-R

Primers to delete SSA2

4934-RG-del-SSA2-NAT-F

4935-RG-del-SSA2-NAT-R

Real-time quantitative PCR primers

36-ACT1-RT-F1

37-ACT1-RT-R1

2873-ZEO1-F

2874-ZEO1-R

3314-PHO84-qPCR-F

CAAGATCCACACTGCTGAAAAGAGAA
| Primer Set | Gene | Sequence |
|------------|------|----------|
| 3315-PHO84-qPCR-R | CCAACCGTTAATTGCCATGT |
| 3442-RPS16A-qPCR2-F | CGATGAAACATCCAAGAAACGA |
| 3443-RPS16A-qPCR2-R | GAATCAGCAAATAAAGGTTCTG |
| 4397-BS-FLO11-Sigma-qPCR-F | CACTTTTGAGTCTATGCCACACAA |
| 4398-BS-FLO11-Sigma-qPCR-R | CATGCAATTCAGGCAGCAT |
| 4500-RG-qPCR-ssn6-F1 | ATCCCTCCGATGCCACTACA |
| 4501-RG-qPCR-ssn6-R1 | TGCGCAGTATAATCTGTCTAATCA |
| 4502-RG-qPCR-ssn6-F2 | CCCCCACCTTAAATCCAGCAT |
| 4503-RG-qPCR-ssn6-R2 | GATTGCACCTTCATGTATCTTTT |
| 4815-RG-qPCR-FOX2-F | GGAATGCACAAGCAAGAGCTTAA |
| 4816-RG-qPCR-FOX2-R | AACGTGGGCAAACATTGGAA |
| 4817-RG-qPCR-MLS1-F | CGTTGTAAAGGCCTGCTAATCTG |
| 4818-RG-qPCR-MLS1-R | CCCCTCGGTAGTAATCTCACATT |
| 4823-RG-qPCR-CIN5-F | AAGAGAGCTGCCAAAATCG |
| 4824-RG-qPCR-CIN5-R | TTCTCACGACGCTGCTAATCT |
| 4825-RG-qPCR-NRG1-F | CGCAGCTCCCGAAACTCT |
| 4826-RG-qPCR-NRG1-R | CGGGGCTTTTCAACGTGTTT |
| 4827-RG-qPCR-HXK1-F | GACAGTGTCAGCAGCAAG |
| 4828-RG-qPCR-HXK1-R | ACCGCAGCTGAAACGACATC |
### Yeast strains used in this study

| Strain name | Genotype | Source |
|-------------|----------|--------|
| KV447       | S288c (Mat a) | Fink lab |
| KV449       | Sigma1278b (Mat a) | Fink lab |
| RG1         | KV447 + HYG downstream SNF5 (51Q, WT strain) | This study |
| RG518       | KV449 + HYG downstream SSN6 (TR1-16, TR2-63 repeats; WT strain) | This study |

**SNF5 polyQ deletion containing PHO84pr-YFP fusions**

| Strain name | Genotype | Source |
|-------------|----------|--------|
| RG22        | SNF5 0Q | This study |
| RG258       | RG22 (0Q) YRO2::PHO84pr-yEVenus-KAN | This study |
| RG256       | RG1 (51Q) YRO2::PHO84pr-yEVenus-KAN | This study |

**SSN6 TR1 variants**

| Strain name | Genotype | Source |
|-------------|----------|--------|
| RG480       | SSN6 TR1-0 | This study |
| RG476       | SSN6 TR1-5 | This study |
| RG478       | SSN6 TR1-10 | This study |
| RG516       | SSN6 TR1-32 | This study |
| RG479       | SSN6 TR1-51 | This study |

**SSN6 TR2 variants**

| Strain name | Genotype | Source |
|-------------|----------|--------|
| RG475       | SSN6 TR2-0 | This study |
| SB13        | SSN6 TR2-14 | This study |
| SB37        | SSN6 TR2-20 | This study |
| SB35        | SSN6 TR2-27 | This study |
| SB42        | SSN6 TR2-31 | This study |
| SB34        | SSN6 TR2-33 | This study |
| SB46        | SSN6 TR2-53 | This study |
| SB44        | SSN6 TR2-55 | This study |
| SB21        | SSN6 TR2-90 | This study |
| SB51        | SSN6 TR2-105 | This study |
| SB57        | SSN6 TR2-105 | This study |

**Tagged SSN6 TR variants (Ssn6-YFP or Ssn6-HA)**

| Strain name | Genotype | Source |
|-------------|----------|--------|
| RG697       | RG475 (TR2-0) SSN6-yEVenus-KAN | This study |
| RG666       | RG518 (TR2-63) SSN6-yEVenus-KAN | This study |
| RG668       | SB57 (TR2-105) SSN6-yEVenus-KAN | This study |
| SB198       | SB51 (TR2-105) SSN6-yEVenus-KAN | This study |
| RG870       | RG697 (TR2-0) SSA2-mCherry-URA3 | This study |
| RG872       | RG666 (TR2-63) SSA2-mCherry-URA3 | This study |
| RG874       | RG668 (TR2-105) SSA2-mCherry-URA3 | This study |
| RG888       | RG697 (TR2-0) ssa2::NAT | This study |
| RG890       | RG666 (TR2-63) ssa2::NAT | This study |
| RG892       | RG668 (TR2-105) ssa2::NAT | This study |
| SSN6 TR variants containing FLO11 promoter - YFP fusions | SSN6 TR variants containing IMA1 protein - YFP fusions | SSN6 TR variants containing FLO11 promoter - YFP and CIN5 protein - RFP fusions |
|---------------------------------------------------------|-------------------------------------------------------|------------------------------------------------------------------|
| SB187 RG480 (TR1-0) flo11::yEVenus-KAN | RG894 SB309 (TR2-0) ssa2::NAT | RG934 SB309 (TR2-0) CIN5-mCherry-URA3 |
| SB650 RG476 (TR1-5) flo11::yEVenus-KAN | RG895 SB309 (TR2-0) ssa2::NAT | RG931 SB205 (TR2-14) CIN5-mCherry-URA3 |
| SB652 RG478 (TR1-10) flo11::yEVenus-KAN | RG897 SB342 (TR2-63) ssa2::NAT | |
| SB653 RG516 (TR1-32) flo11::yEVenus-KAN | RG898 SB342 (TR2-63) ssa2::NAT | |
| SB184 SB309 (TR2-0) CIN5-mCherry-URA3 | RG899 SB211 (TR1-105) ssa2::NAT | |
| SB205 SB13 (TR2-14) flo11::yEVenus-KAN | RG900 SB211 (TR1-105) ssa2::NAT | |
| SB173 SB37 (TR2-20) flo11::yEVenus-KAN | GB832 SB21 (TR2-90) IMA1-yEVenus-KAN | |
| SB170 SB35 (TR2-27) flo11::yEVenus-KAN | GB836 SB35 (TR2-90) IMA1-yEVenus-KAN | |
| SB176 SB42 (TR2-31) flo11::yEVenus-KAN | GB856 SB34 (TR2-33) IMA1-yEVenus-KAN | |
| SB167 SB34 (TR2-33) flo11::yEVenus-KAN | GB848 SB44 (TR2-55) IMA1-yEVenus-KAN | |
| SB182 SB46 (TR2-53) flo11::yEVenus-KAN | GB888 SB44 (TR2-55) IMA1-yEVenus-KAN | |
| SB179 SB44 (TR2-55) flo11::yEVenus-KAN | GB858 GB518 (TR2-63) IMA1-yEVenus-KAN | |
| SB342 GB518 (TR2-63) IMA1-yEVenus-KAN | GB900 SB211 (TR2-90) IMA1-yEVenus-KAN | |
| GB303 SB21 (TR2-90) flo11::yEVenus-KAN | RG902 GB57 (TR2-105) IMA1-yEVenus-KAN | |
| GB221 SB51 (TR2-105) flo11::yEVenus-KAN | GB225 SB57 (TR2-105) IMA1-yEVenus-KAN | |
| RG926 RG475 (TR2-0) IMA1-yEVenus-KAN | GB228 SB35 (TR2-90) IMA1-yEVenus-KAN | |
| GB834 SB35 (TR2-27) IMA1-yEVenus-KAN | GB228 SB35 (TR2-90) IMA1-yEVenus-KAN | |
| GB856 SB34 (TR2-33) IMA1-yEVenus-KAN | GB228 SB35 (TR2-90) IMA1-yEVenus-KAN | |
| GB948 SB44 (TR2-55) IMA1-yEVenus-KAN | GB228 SB35 (TR2-90) IMA1-yEVenus-KAN | |
| GB828 GB518 (TR2-63) IMA1-yEVenus-KAN | GB228 SB35 (TR2-90) IMA1-yEVenus-KAN | |
| GB832 SB21 (TR2-90) IMA1-yEVenus-KAN | GB228 SB35 (TR2-90) IMA1-yEVenus-KAN | |
| GB952 SB57 (TR2-105) IMA1-yEVenus-KAN | GB228 SB35 (TR2-90) IMA1-yEVenus-KAN | |
| Strain     | Description                                      | Source          |
|------------|--------------------------------------------------|-----------------|
| RG929      | SB170 (TR2-27) CIN5-mCherry-URA3                 | This study      |
| RG930      | SB179 (TR2-55) CIN5-mCherry-URA3                 | This study      |
| RG935      | SB342 (TR2-63) CIN5-mCherry-URA3                 | This study      |
| RG933      | SB303 (TR2-90) CIN5-mCherry-URA3                 | This study      |
| RG932      | SB211 (TR2-105) CIN5-mCherry-URA3                | This study      |
|            | **SSN6 WT overexpressing expanded human Huntington (Ht-103Q)** |                |
| RG944      | RG518 (TR2-63) flo11::CFP-KAN                    | This study      |
| RG959      | RG944 (TR2-63) pGPD-103Q-GFP-URA3                | This study      |
| RG960      | RG944 (TR2-63) pGPD-103Q-GFP-URA3                | This study      |
|            | **SSN6 QA to P mutants**                         |                |
| RG735      | SSN6 29QA/3P- 30Q/1P                             | This study      |
| RG736      | SSN6 29QA/3P- 30Q                               | This study      |
| RG738      | SSN6 26QA/2P- 31Q                               | This study      |
| RG740      | RG735 flo11::yEVenus-KAN                         | This study      |
| RG743      | RG736 flo11::yEVenus-KAN                         | This study      |
| RG746      | RG738 flo11::yEVenus-KAN                         | This study      |
|            | **Diploids resulting from the cross to check SSN6 prionization** |                |
| RG690      | SSN6 TR2-90/TR2-105                             | This study      |
| RG691      | SSN6 TR2-0/TR2-105                              | This study      |
| RG692      | SSN6 TR2-63/TR2-105                             | This study      |
SUPPLEMENTAL REFERENCES

Beissbarth, T., and Speed, T.P. (2004). GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics 20, 1464-1465.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27, 573-580.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26, 1367-1372.

Flicek, P., Ahmed, I., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., et al. (2013). Ensembl 2013. Nucleic Acids Res 41, D48-55.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41, D808-815.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell 11, 4241-4257.

Ghesquiere, B., Jonckheere, V., Colaert, N., Van Durme, J., Timmerman, E., Goethals, M., Schymkowitz, J., Rousseau, F., Vandekerckhove, J., and Gevaert, K. (2011). Redox proteomics of protein-bound methionine oxidation. Mol Cell Proteomics 10, M110 006866.

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell 95, 717-728.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4, 44-57.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 42, D199-205.

Luscombe, N.M., Babu, M.M., Yu, H., Snyder, M., Teichmann, S.A., and Gerstein, M. (2004). Genomic analysis of regulatory network dynamics reveals large topological changes. Nature 431, 308-312.

New, A.M., Cerulus, B., Govers, S.K., Perez-Samper, G., Zhu, B., Boogmans, S., Xavier, J.B., and Verstrepen, K.J. (2014). Different levels of catabolite repression optimize growth in stable and variable environments. PLoS Biol 12, e1001764.
Reynolds, T.B., and Fink, G.R. (2001). Bakers' yeast, a model for fungal biofilm formation. Science 291, 878-881.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11, R25.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27, 431-432.

Stes, E., Laga, M., Walton, A., Samyn, N., Timmerman, E., De Smet, I., Goormachtig, S., and Gevaert, K. (2014). A COFRADIC protocol to study protein ubiquitination. J Proteome Res 13, 3107-3113.

Teixeira, M.C., Monteiro, P.T., Guerreiro, J.F., Goncalves, J.P., Mira, N.P., dos Santos, S.C., Cabrito, T.R., Palma, M., Costa, C., Francisco, A.P., et al. (2014). The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in Saccharomyces cerevisiae. Nucleic Acids Res 42, D161-166.

Thomas, M.R., and O'Shea, E.K. (2005). An intracellular phosphate buffer filters transient fluctuations in extracellular phosphate levels. Proc Natl Acad Sci U S A 102, 9565-9570.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105-1111.