Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity

Highlights

- Glutamine-rich repeats are highly enriched in eukaryotic transcriptional regulators

- Repeat variation in a fungal transcription factor changes expression of its targets

- Variable repeats enable tuning of transcription factor solubility and interactions

- Excessive repeat expansion results in TF malfunctioning and aggregation

In Brief

Expanded glutamine-rich repeat domains within regulatory proteins are associated with devastating neurodegenerative diseases. However, Gemayel et al. show that these highly variable glutamine-rich repeats also have a functional role by modulating the solubility, interactions, and function of the transcriptional regulators in which they typically reside.

Accession Numbers

GSE70209
Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity

Rita Gemayel,1,2 Sreenivas Chavali,3 Ksenia Pougach,1,2 Matthieu Legendre,4 Bo Zhu,1,2 Steven Boeynaems,1,2 Elisa van der Zande,1,2 Kris Gevaert,1,6 Frederic Rousseau,7 Joost Schymkowitz,7 M. Madan Babu,3 and Kevin J. Verstrepen1,2,*

1Laboratory of Systems Biology, VIB, Gaston Geenslaan 1, 3001 Heverlee, Belgium
2Laboratory of Genetics and Genomics, Centre of Microbial and Plant Genetics (CMPG), Department M2S, KU Leuven, Gaston Geenslaan 1, 3001 Heverlee, Belgium
3MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK
4Structural and Genomic Information Laboratory, IGS UMR7256, Centre National de la Recherche Scientifique, Aix-Marseille Université, Institut de Microbiologie de la Méditerranée (IMM), 13288 Marseille Cedex 9, France
5Department of Medical Protein Research, VIB, 9000 Ghent, Belgium
6Department of Biochemistry, Ghent University, 9000 Ghent, Belgium
7Switch Laboratory, VIB, Campus Gasthuisberg, KU Leuven, Herestraat 49, 3000 Leuven, Belgium
*Correspondence: kevin.verstrepen@biw.vib-kuleuven.be
http://dx.doi.org/10.1016/j.molcel.2015.07.003
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

SUMMARY

Excessive expansions of glutamine (Q)-rich repeats in various human proteins are known to result in severe neurodegenerative disorders such as Huntington’s disease and several ataxias. However, the physiological role of these repeats and the consequences of more moderate repeat variation remain unknown. Here, we demonstrate that Q-rich domains are highly enriched in eukaryotic transcription factors where they act as functional modulators. Incremental changes in the number of repeats in the yeast transcriptional regulator Ssn6 (Cyc8) result in systematic, repeat-length-dependent variation in expression of target genes that result in direct phenotypic changes. The function of Ssn6 increases with its repeat number until a certain threshold where further expansion leads to aggregation. Quantitative proteomic analysis reveals that the Ssn6 repeats affect its solubility and interactions with Tup1 and other regulators. Thus, Q-rich repeats are dynamic functional domains that modulate a regulator’s innate function, with the inherent risk of pathogenic repeat expansions.

INTRODUCTION

Polyglutamine (polyQ) repeats have been associated with several human neurodegenerative diseases, including Huntington’s disease, Kennedy disease, and multiple spinocerebellar ataxias (reviewed in Orr and Zoghbi, 2007). Like other tandem repeat (TR) sequences, polyQ repeats show a high mutation rate, exceeding that of single-nucleotide polymorphisms by orders of magnitude (Legendre et al., 2007; Lynch et al., 2008). Specifically, repeats often shrink and expand at rates between 10⁻² and 10⁻⁶ per generation. For repeats associated with neurodegenerative diseases, variation within certain limits is not pathogenic, and different individuals in the population often have different repeat lengths (Duitama et al., 2014). However, repeat expansion beyond a certain threshold causes disease, with longer expansions leading to earlier disease onset and faster progression. Pathogenic polyQ expansions have been shown to alter several cellular processes that can lead to neuronal dysfunction. They can modify the conformation of a protein and affect its interaction with partners (Schaffar et al., 2004), lead to depletion of tRNA and subsequent translational frameshifting (Girstmair et al., 2013), cause non-ATG translation and production of abnormal peptides (Pearson, 2011), and even alter normal proteasome function (Park et al., 2013). However, despite the ubiquitous presence of TRs in functional regions of genomes (Duitama et al., 2014; Legendre et al., 2007; Li et al., 2002) and the fact that many repeats are conserved over evolutionary timescales (Schaper et al., 2014), their functional significance outside of the pathological context remains unknown. Natural variation in the length of Q-rich repeats was often dismissed as unimportant neutral drift with no tangible phenotype or physiological role. Isolated studies in several organisms, however, reported instances where TR variation correlated with phenotypic changes (Fondon and Garner, 2004; Sawyer et al., 1997).

Although Q-rich repeats are enriched in eukaryotic transcriptional regulators (Gemayel et al., 2010; Legendre et al., 2007), a comprehensive understanding of their role, specifically whether their variation causes any functional changes, remains unanswered. First, using comparative genomics, we show that targets of Q-rich regulators have elevated gene-expression variation across multiple timescales, suggesting a role of Q-rich repeats in gene-expression regulation. To elucidate how variable Q-rich repeats might influence transcription, we generated multiple repeat variants of the yeast transcriptional regulator Ssn6 (Cyc8). We present direct experimental evidence showing that Ssn6 repeat-length variation affects the expression of target genes. Remember that this is an open access article under the CC BY-NC-ND license.
genes, which in turn results in a broad array of phenotypic changes. Using quantitative proteome analysis, we further demonstrate that Ssn6 solubility and its interaction with partners depend on the length of the repeat region. The Hsp70 chaperone Ssa2 helps to maintain Ssn6 function by reducing its intrinsic, repeat-length-dependent propensity to misfold and aggregate. Together, these results demonstrate that, while excessive repeat expansion is pathogenic, Q-rich repeats with normal lengths are functional domains that can help maintain and tune proper transcriptional regulation.

RESULTS

Glutamine-Rich Transcription Factors Promote Target Gene-Expression Divergence

We scanned the open reading frames of all protein coding genes in genomes that span the eukaryotic diversity (yeast, fruit fly, zebrafish, mouse, human) using Tandem Repeat Finder (Benson, 1999). We find that 14%–20% of eukaryotic genes are enriched in TRs (Table S1). We defined repeats as Q rich if at least 85% of their translated sequence comprised glutamine residues (Table S1). Gene ontology analysis of these Q-rich genes versus all genes with repeats revealed a significant enrichment for regulatory functions such as transcriptional regulation and chromatin modification (Table S2). This is consistent with previous studies investigating the functional enrichment of repeat-containing proteins in various eukaryotic genomes (Faux et al., 2005; Gemayel et al., 2010; Legendre et al., 2007; Young et al., 2000). TRs are often unstable, with even closely related individuals or species showing differences in the number of repeated units in a “homologous” TR. This prompted us to ask whether repeats in transcription factors (TFs) can influence the variability of target gene expression. To address this, we first reconstructed a comprehensive yeast transcriptional regulatory network (TRN) by combining a previously published TRN (Balaji et al., 2006) with the recent genome-wide in vivo binding map of yeast regulatory proteins (Venters et al., 2011) (Figure 1A). We classified the target genes as those regulated by Q-rich TFs and those by non-repeat containing TFs (NR-TFs). We next analyzed gene-expression variation over long, intermediate, or short timescales by examining published datasets on yeast gene-expression variation across species (Tirosh et al., 2006), among strains (Choi and Kim, 2008), across generations (Landry et al., 2007), and between genetically identical cells at an instant (Newman et al., 2006). We find that target genes that are regulated by Q-rich TFs show significantly higher levels of expression divergence, expression variability, mutational variance, and expression noise than targets of NR-TFs (Figure 1B). Importantly, these differences in expression patterns of the targets are not explained by changes in expression patterns of the respective TFs (Figure S1A) or by differences in average transcript levels of targets (Figure S2A). Even if we only consider TFs with low expression variation across multiple time scales, the targets of Q-rich TFs still show higher expression diversity and variability and higher mutational variance than targets of NR-TFs (Figure S1B).

While the targets of NR-TFs are enriched in processes related to gene-expression, those of Q-rich TFs were enriched for metabolic processes (Figure S2B). This prompted us to investigate whether Q-rich TF targets are on average more amenable to dynamic gene-expression regulation. To test this, we compared expression plasticity, which captures the magnitude of gene-expression variation to various conditions (Tirosh and Barkai, 2008). The results show that on average, Q-rich TF target genes show a larger range of expression modulation compared with NR-TF targets (Figure 1C) and those with high expression plasticity showed higher variation across different timescales (Figure 1D). Moreover, irrespective of the nature of the stimuli they respond to—exogenous (stress) or endogenous (developmental transitions, etc.)—Q-rich TF targets have higher expression variation across multiple timescales than targets of NR-TFs (Figures S2C and S2D). Taken together, our results suggest that among genes with highly dynamic gene-expression modulation, Q-rich TF targets tend to have enhanced expression variability at multiple timescales (Figures 1C and 1D).

Whole-Genome Transcriptomics Reveal Gradual and Quantitative Changes in Target Expression as a Result of SSN6 Repeat Variation

To assess whether the variability in Q-rich repeats within transcriptional regulators directly influences the expression of downstream targets, we chose the S. cerevisiae transcriptional regulator Ssn6 (Cyc8) (Chen et al., 2013; Tzamarias and Struhl, 1994; Wong and Struhl, 2011). Ssn6 is a TF that controls a large number of targets that show high expression variability (Figure 1E). The SSN6 coding sequence comprises two Q-rich repeat regions: a first polyQ stretch (residues 15–30), henceforth referred to as Tandem Repeat 1 (TR1), and a second Q-rich region, comprised of an array of QA repeats directly followed by a polyQ stretch (residues 499–587), referred to as TR2 (Figure 2A).

To check for variability in the SSN6 repeats, we sequenced both repeat regions in a number of genetically and ecologically diverse S. cerevisiae strains (Liti et al., 2009) (Figures 2B and 2C; Table S3). Both regions showed variability in repeat number between yeast strains, with TR2 having a wider range of variation and also showing variation between closely related strains (Figure 2C). Similarly, we find Q-rich repeats in other transcriptional regulators (SWI1 and SNF5) to be variable between natural S. cerevisiae strains (Figure S3), indicating that the variability of the Ssn6 Q-rich repeat is typical for repeats in regulatory proteins.

To investigate the direct consequences of SSN6 repeat variation, we created a series of variants of the SSN6 TR1 and TR2 regions in the reference yeast strain Sigma1278b (Figure 2D). These variants included natural repeat lengths, as well as extremely short and long forms. This set of mutants allowed us to explore the outcomes of natural variability, as well as complete loss or long expansions of Q-rich repeats in Ssn6 controlled by its native promoter, in an otherwise isogenic background. As TR2 shows higher variability within the natural strains, we chose nine TR2 variants (natural variations: TR2-33, TR2-55, TR2-63; short variants: TR2-0, TR2-14, TR2-20, TR2-27; long variants: TR2-90, TR2-105) to investigate the influence of Q-rich repeats on gene-expression regulation. Using RNA sequencing, we profiled the transcriptome of the SSN6 TR2 variants in glucose-rich medium and during carbon starvation. As many as 153 genes...
Figure 1. Q-Rich TFs Influence Expression Variation of Targets across Different Timescales

(A) Yeast TRN reconstructed based on data from Balaji et al. (2006) and Venters et al. (2011). For the latter, promoter occupancy cutoff of at least 3-fold higher than background was considered. Based on the presence of Q-rich repeats, the TFs and their targets were categorized.

(B) Distribution of variation of expression among species and strains, across generations, and among genetically identical cells, of targets regulated by non-repeat containing TFs (NR-TFs) and TFs with Q-rich repeats (Q-rich TFs). The boxes represent the first and third quartile with the median at the black line. The notches correspond to 95% confidence interval for the median. The whiskers show data points up to 1.5 times the interquartile range. Statistical significance was assessed using Wilcoxon rank sum test. The effect sizes are represented by the common language effect size (CLES) statistic, describing the probability that a randomly selected target of Q-rich TFs will have higher expression variation than a randomly selected target of NR-TFs.

(C) Influence of expression plasticity on gene-expression variability. Distribution of expression plasticity of NR-TF and Q-rich TF targets is shown. The panels represent the median of expression variation across different timescales of targets of NR-TFs and Q-rich TFs in low (bottom 33.3%), medium (middle 33.3%), and high (top 33.3%) expression plasticity bins defined using tertile cuts of the distribution of all genes. p values were estimated using Wilcoxon rank sum test.

(D) Proposed model of target gene-expression variability over different time-scales facilitated by Q-rich TFs among genes with high dynamic expression modulation.

(E) Enrichment of targets with expression variation across different timescales for each Q-rich TF. Enrichment of targets with expression variation values higher than that of median of all Q-rich TF targets was tested using a permutation test. In each permutation, every target of a Q-rich TF was replaced with a random target from the TRN. The number of random targets with expression variation values equal or higher than the median of Q-rich TF targets was noted for 10,000 iterations. The color intensity in the heatmap represents Z scores, which indicate the distance of the number of real targets to the mean of random expectation in SD units. Statistically significant enrichment is highlighted with a red border. p values were estimated as the ratio of the average number of random targets with expression variation more than or equal to that of Q-rich TF targets over the total number of random samples (10,000).

See also Figures S1 and S2 and Tables S1 and S2.
showed significant changes in expression (log2 fold-change cut-off of 0.8 and false discovery rate (FDR) p < 0.01) in either the SSN6 TR2 deletion (TR2-0) or expansion (TR2-105) variant compared with the WT strain (TR2-63) in either culture conditions (Table S4). To further discriminate between noise and targets whose regulation is directly influenced by SSN6 TR2 number variation, we computed the autocorrelation function, which detects non-randomness in data, considering the SSN6 TR2 number as a series. By selecting SSN6 target genes that showed an autocorrelation coefficient \( R \geq 0.2 \) and the same expression trend in at least two consecutive repeat variants, we identified 89 targets whose regulation was TR2 length dependent (Figure 3A).

In the carbon-starved condition, we observe a positive correlation between the length of TR2 and the expression levels of a subset of genes (Group 1) during carbon starvation are mostly involved in alternative carbon transport and metabolism, including key genes of gluconeogenesis (PCK1, FBP1, ICL1-2, MLS1, IDP2). Genes that show the opposite expression trend (Group 2) are mostly involved in glycolysis (MAE1, CDC19, ENO2, GPM1), flocculation, and filamentous growth (FLO11, TIR3, NRG1, PHD1) or, notably, are TFs that (physically) interact with the Ssn6-Tup1 complex (YAP6, CIN5, NRG1, PHD1, SUT1) (Hanlon et al., 2011) (Figure 3A, left). The genes whose expression in glucose-rich medium is anti-correlated with TR2 length are mostly involved in hexose transport and metabolism (HXT5-6-7, HXT1), stress response (GPH1, HSP42, HSP12, GAD1, TSL1, DDR2), or flocculation and filamentous growth (FLO11, FLO10, BSC1, HMX1, MGA1) (Figure 3A, right). Importantly, we also find several highly induced cytosolic chaperones, including the “antiaaegrase” HSP26, the inducible Hsp70 SSA4, and HSP30 and HSP33 only in the TR2-0 or TR2-105 variants (Table S4), suggesting a response of the protein folding machinery to deletion or expansion of the Ssn6 repeat.

**Figure 2.** The Q-Rich Repeats in *Saccharomyces cerevisiae* Transcriptional Regulator Ssn6 Show Variability between Natural Yeast Strains

(A) Schematic representation of the Ssn6 protein showing the two repeat regions. TR1 denotes the N-terminal polyQ (residues 15 to 30), and TR2 denotes the central Q-rich repeat (residues 493 to 587). The natural range of repeat number variation is indicated underneath each repeat region. See also Figure S3.

(B) The TR1 region of SSN6 from various *S. cerevisiae* strains was amplified by PCR.

(C) Amplification of the SSN6 TR2 region from various *S. cerevisiae* strains. The asterisks denote genetically close strains YPS606 (*) and YPS128 (**) with different TR2 sizes. See also Table S3.

(D) TR sizes in representative SSN6 variants constructed for this study. Total repeat numbers are given. The asterisk indicates the repeat number in the WT strain. Repeat numbers falling within the range observed in natural strains (C) are indicated.
Using published datasets on protein-protein, protein-DNA, and metabolic interactions, we constructed a network between SSN6 (CYC8) and its targets identified by RNA-seq (Figure 3B). This network shows that most targets are connected to SSN6 through direct protein-DNA interactions with TFs (CIN5, PHD1, YAP6, NRG1, and SUT1) that act as hub proteins connected to SSN6 by direct protein-DNA as well as protein-protein interactions. Furthermore, we find significant overlap between the genes whose expression is identified as SSN6 TR2-dependent in this study and (1) de-repressed genes upon Tup1 nuclear depletion (Wong and Struhl, 2011) and (2) SSN6 target promoters identified by chromatin immunoprecipitation (ChIP) (Venter et al., 2011) (Figure 3C). This observation, along with the fact that most genes identified by our RNA-seq could be included in an interaction network, suggests that these genes are bona fide SSN6 targets and that the observed changes in their expression are most likely a direct consequence of changes in the number of SSN6 TR2.

To investigate whether this variation in gene expression under changing nutrient availability has phenotypic consequences, we measured growth of the TR2 variants in alternative carbon or nitrogen sources using the Phenotype MicroArrays (Biolog). While we find no significant differences in growth when glucose is the sole carbon source, the fitness of the variants increases linearly with TR2 number when fructose or galactose is used (Figure 3D). Interestingly, this fitness trend closely resembles the expression pattern of SUC2 (the major sucrose hydrolyzing enzyme) (highlighted in Figure 3A, left). Moreover, growth in sucrose mirrors the more complex expression pattern of SUC2 (the major sucrose hydrolyzing enzyme) (highlighted in Figure 3A, right) and growth in the presence of α-Amino-N-Valeric Acid (Norvaline) as the sole nitrogen source is anti-correlated with TR2 number (Figure 3D).

We next analyzed the SSN6 TR2 variants for phenotypes associated with FLO11 and IMA1, two target genes (highlighted in Figure 3B) that underlie very specific and quantifiable phenotypes: FLO11 mediates complex colony morphology (induced by growth on sucrose) (Voordekers et al., 2012b), invasive growth (induced by glucose starvation) (Brücker and Mösch, 2012), adhesion to plastic (Reynolds and Fink, 2001), and flocculation (Verstrepen and Klis, 2006), while IMA1 is a glucosidase required for growth on palatinose (a naturally occurring disaccharide) (Brown et al., 2010; Voordekers et al., 2012a).

All the Flo11-related phenotypes presented a common response to variation in SSN6 TR2 number: colonies showed more “wrinkled” surfaces, stronger agar invasion and plastic adhesion, and higher flocculation levels in the short SSN6 TR2 variants (Figure 4D). These phenotypes gradually decrease with increasing TR2 number. On the other hand, the Ima1-dependent growth rates on palatinose closely follow Ima1 levels in these TR2 variants (Figure 4D). Together, this analysis demonstrates that Ssn6 TR2 variation results in changes in the expression levels of its target genes and also results in corresponding changes in fitness in different nutrient environments. Moreover, these transcriptomics and phenotypic data are in keeping with the results of the genome-scale analysis showing that the targets of Q-rich TFs are enriched in metabolic processes (Figure S2B).

To confirm that Q-rich repeats can be functional domains in transcriptional regulators and that our main observations are not limited to Ssn6, we investigated the effect of deletion of the polyQ repeat in Snf5, a subunit of the SWI/SNF chromatin remodeling complex in S. cerevisiae. Deletion of the polyQ resulted in lower expression of known Snf5 targets (Figure S5A). Additionally, using a fluorescent reporter for PHO84 (a high-affinity phosphate transporter), we show that Snf5 polyQ deletion leads to lower steady-state expression of PHO84 and to differences in PHO84 induction in response to variable concentrations of extracellular phosphate (Figure S5B). Together, these results show that the Q-rich repeat influences the function of Snf5, similar to what we observed with Ssn6.

**Variable SSN6 Repeats Contribute to Target Expression Noise**

We next asked whether SSN6 repeat variability affects the expression noise of its targets (i.e., stochastic cell-to-cell differences in gene expression), as suggested by our genome-scale observations (Figure 1B). We grew fluorescent reporter strains of three targets (FLO11, IMA1, and CIN5) identified by RNA-seq (highlighted in Figure 3B) in appropriate media for more than ten doublings, maintaining them at low cell density to ensure steady-state levels of expression. First, protein levels measured using reporter constructs closely follow the changes in mRNA levels, with the exception of Cin5 levels in the TR2-105 variant (Figure 4A). Second, the fluorescence distributions indicate that changes in SSN6 TR2 variants select target-specific effects on expression noise (Figure 4B).

Specifically, flo11::YFP distributions show remarkable variability between the TR2 variants (Figures 4B and 4C), whereas ima1::YFP distributions only show greater dispersion in the TR2-105 variant and Cin5-RFP fluorescence is homogeneous between all the variants (Figure 4B). This indicates that Q-rich repeat variation in the regulator could be one of the factors that contributes to expression noise and this effect is target dependent.

**Ssn6 Q-Rich Repeat Variation Results in Changes in Its Stability and Interactome**

How does repeat variation lead to changes in Ssn6 function? Quantitative real-time PCR analysis showed that the expression of SSN6 is similar in all the TR2 variants and does not differ significantly from the WT allele (TR2-63) (Figure S4B). Moreover, expanded SSN6 repeats (TR2-105) are stable over multiple (>30) generations (Figure S6A), ruling out that TR2 repeat heterogeneity within a population would be responsible for the increased expression noise in some Ssn6 targets.

To check for changes in the Ssn6 stability and interactome, we fused the Ssn6 repeat variants (TR2-0, TR2-55, TR2-63, TR2-90, and TR2-105) to six copies of an HA epitope tag at their C terminus. Using gel-bound anti-HA antibodies, we isolated the tagged Ssn6 from soluble protein extracts. Bound proteins were separated by SDS-PAGE, followed by in-gel digestion and identification by mass spectrometry (Figure 5A, left). This qualitative approach shows that the interactomes of the repeat deletion and expansion variants are larger than that of the WT and also contain common elements (Figure 5A, left). Notably, these common interactors are involved in active chromatin remodeling.
Figure 3. Whole-Genome Transcriptomics Reveal that Variation in the SSN6 TR2 Region Influences the Expression of Target Genes

(A) Expression profiles for genes (rows) in different SSN6 TR2 variants (columns) in carbon-starved (left) or glucose-rich medium (right) measured by RNA-seq. The data are represented as relative to the expression levels in the WT (TR2-63), and similar colors indicate similar changes in expression relative to the WT strain. Enriched biological processes (Gene Ontology [GO] categories) of the target genes are shown (p < 0.05). Genes highlighted with arrows are key genes involved in alternative carbon transport and catabolism. The SUC2 gene that encodes the major sucrose-hydrolyzing enzyme is also highlighted. See also Figure S4 and Table S4.
(e.g., the histone variant H2AZ, Dot6, Yku80), transcription (Ppr1, Tfg1), and signaling (Pkc1, Bmh1, Ypt1). To confirm these results and to accurately quantify TR2-dependent changes in the Ssn6 interactome, we repeated the pull-down experiment and eluted the antibody-bound protein complexes. The eluted fractions were digested, and the peptide mixtures from the WT were labeled with a light isotope ($^{12}$C$_3$-proprionate), while those from the TR2 variants were labeled with a heavy isotope ($^{13}$C$_3$-proprionate). The labeled peptides from each TR2 variant and the WT were mixed 1:1 and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Figure 5A, right). Among the 85 proteins detected, we considered the ones identified by at least two different peptides in all five TR2 variants tested (Table S5). We find >2-fold depletion of Ssn6 and its known interaction partner Tup1 in the TR2-105 variant relative to the WT (Figure 5A, right). Concurrently, we find a significant enrichment of the Hsp70 chaperone Ssa2 in the TR2 expansion (TR2-90 and TR2-105) and deletion (TR2-0) variants. Ssa2 is a cytosolic chaperone required for protein folding, translocation, and degradation of unfolded proteins. It also partners with Hsp40 co-chaperones to promote the refolding of aggregated or unfolded proteins (Verghese et al., 2012). Interestingly, we find that Ssa2 co-localizes with expanded Ssn6 in the nucleus, but not with the WT or TR2-0 variant (Figure 5B).

To assess whether Ssa2 is required for proper Ssn6 function, we deleted this chaperone in the flo11::YFP reporter strains. In general, we observed higher FLO11 levels in the Jsssa2 background relative to the parental strains, indicative of a loss of (repressive) Ssn6 function (Figure 5C). The most dramatic functional change is observed in the TR2-105 variant, where median flo11::YFP levels increase ~20-fold in the absence of Ssa2 and become homogeneous within the population (less expression noise) (Figure 5G).

The involvement of Ssa2 points to a TR2-dependent effect on the conformational dynamics of Ssn6. Aggregation is an intrinsic property of glutamine-rich repeats, and its propensity increases with repeat number (Krobitsch and Lindquist, 2000). Using fluorescence microscopy on Ssn6-YFP fusion proteins, we checked whether Ssn6 formed inclusions. We did not detect any inclusions in actively dividing cells (not shown), but when cells reached the stationary phase, the TR2-105 variant formed nuclear inclusions visible as bright foci (Figure 5D). Notably, deletion of Ssa2 leads to an increase in fluorescent foci in the TR2-105 variant and even the appearance of foci in the WT as well as in the TR2-0 variant (Figure 5D). This suggests that Ssa2 maintains proper Ssn6 folding and counteracts its intrinsic property to aggregate. We did not detect any change in the subcellular localization of Ssn6-YFP in any TR2 variant (not shown).

Under specific experimental conditions, Ssn6 can misfold into self-templating prion-like amyloid fibers (Patel et al., 2009). To determine whether the TR2-105 inclusions were prion aggregates, we devised a strategy to detect protein-only (i.e., non-mendelian) inheritance typical of prions. We crossed the TR2-105 variant with the TR2-0 variant or with the WT strain and scored the progeny for its colony morphology on YP-sucrose. TR2-105 colonies are smooth, TR2-0 very wrinkled, and TR2-63 moderately wrinkled on YP-sucrose. In both crosses, the morphology of the resulting progeny followed 2:2 segregation (Figure S6B), indicating mendelian inheritance of phenotype and ruling out Ssn6 prionization under these conditions.

To provide direct biochemical evidence that repeat expansion results in misfolding and aggregation of Ssn6, we performed an aggregation assay to assess the soluble and insoluble fractions of Ssn6 in the different TR2 variants (Figure 5E). While more than 80% of the WT Ssn6 can be found in the soluble fraction, almost 60% of TR2-105 and about 40% of the short TR2 variants (TR2-0 and TR2-55) of Ssn6 were recovered from the insoluble fraction (Figure 5E). Ssn6 TR2 is predicted to adopt an $\alpha$-helical coiled-coil structure (Fiumara et al., 2010). To investigate whether this conformation is essential for Ssn6 function, we substituted two or three QA residues in the WT Ssn6 TR2 with single prolines, a strategy commonly employed for helix disruption (Figure 5F). In these QAP mutants, FLO11 levels are identical to those in the TR2-0 variant, indicating that partial or complete impairment of the TR2 structure results in loss of Ssn6 function (Figure 5G).

To further confirm that repeat expansion causes effects directly related to Ssn6 function, rather than being a general consequence of aggregation of a Q-rich repeat, we overexpressed, in the WT, a construct bearing a long, aggregation-prone Q-rich repeat (Krobitsch and Lindquist, 2000). Our results show that expression of this artificial repeat does not have any effect on FLO11 expression (Figure S6C), suggesting that the observed effects of TR2 variation are indeed directly related to Ssn6 function rather than being non-specific effects of polyQ aggregates.

Together, these results indicate that the Q-rich repeat length modulates the interaction of Ssn6 and its partner Tup1 and maintains a balance between correctly folded versus an aggregated Ssn6 and that Ssa2 is critical for maintaining a functional Ssn6.

**DISCUSSION**

TRs are present in ~20% of genes in the human genome, yet their functional significance is still not fully understood. In this study, we provide a detailed and direct investigation of the outcome of Q-rich repeat variation in a physiological context. We engineered *S. cerevisiae* strains that only differ in the number of repeats in the transcriptional regulator *SSN6* (*CYC8*) to mimic the natural repeat variation as well variations outside the natural range. We provide evidence that even moderate variations in Ssn6 repeat number result in detectable functional changes.
Figure 4. Correlation between Ssn6 TR2-Dependent Variation in Target mRNA, Protein Levels, Expression Noise, and Associated Phenotypes

(A) Mean mRNA levels, measured by RNA-seq are expressed as fold change relative to the WT (TR2-63 variant). Mean fluorescence of the corresponding reporter was measured by flow cytometry. Data points represent mean ± SEM, n = 5. See also Figure S5.

(legend continued on next page)
These are likely caused by changes in Ssn6 solubility and interaction with other proteins, including its prime partner Tup1 (Figure 6). Our study thus shows that Q-rich repeat variation results in changes in TF stability and interaction, which in turn leads to changes in the expression of its target genes and the phenotypes they control, supporting the notion that repeats could function as evolvable genetic tuning knobs of TF function (King et al., 1997).

**Role of Variable Q-Rich Repeats in Modulating the Function of Ssn6**

Q-rich repeats are enriched in transcriptional regulators in eukaryotic genomes from yeasts to humans. We first show, using genome-scale analyses in yeast, that targets of Q-rich regulators have higher expression variability across multiple timescales compared with targets of regulators devoid of repeats. We then experimentally demonstrate, using Q-rich repeat (TR2) variants of the yeast Ssn6, that repeat variation in a regulator results in expression changes of its direct targets (Figure 3A).

Interestingly, expression of some Ssn6 targets directly correlates with its TR2 length, whereas the expression of other targets follows an inverse trend. This is likely due to the fact that the Ssn6-Tup1 complex acts as a transcriptional repressor for some targets (Smith and Johnson, 2000) and a co-activator for others (Papamichos-Chronakis et al., 2002; Wong and Struhl, 2011). In line with these observations, we find that 22 of the 37 (~60%) genes whose expression shows a positive correlation with Ssn6 TR2 number (Group 1) have Cat8 and/or Adr1 binding sites in their promoters (Tachibana et al., 2005). These are carbon-responsive transcriptional activators required for transcription in low glucose. On the other hand, 17 of the 29 (~60%) genes whose expression follows an opposite trend to increasing Ssn6 TR2 (Group 2) contain Phd1 (or its homolog Cin5) and/or Nrg1 sites in their promoters. The network analysis further demonstrates how Ssn6 can act as a master regulator controlling the expression of these key TFs that are both direct Ssn6 targets themselves (protein-DNA interactions) and Ssn6 partners (protein-protein interactions) (Figure 3B). Moreover, deletion of the Q-rich repeat in another transcription regulator, Snf5, resulted in similar changes in its target gene expression, suggesting that our observations for Ssn6 may represent a general phenomenon.

The almost identical response of diverse targets to Ssn6 repeat variation points to a common underlying mechanism. We found that Ssn6 solubility and its interaction with Tup1 depend on its TR2 length and that the helical conformation of TR2 is essential for Ssn6 function (Figures 5F and 5G). This repeat region could behave as a flexible hydrophilic tract that increases the solubility of Ssn6 and also allows its flanking domains to come to spatial proximity, similar to the polyQ domain of huntingtin (Caron et al., 2013). Repeat shortening or expansion could impair the flexibility of this tract and consequently the conformational dynamics of Ssn6. In addition, Ssn6 repeat deletion or expansion variants acquire common interactors (Figure 5A), which could further contribute to the observed (and in some cases identical) functional changes. The Hsp70 chaperone Ssa2 likely counteracts the intrinsic, repeat-length dependent propensity of Ssn6 to aggregate.

**Contribution of Unstable Q-Rich Repeats to Phenotypic Diversity and Evolvability**

Expression divergence is key to the emergence of variable phenotypes. Here we provide evidence that links Ssn6 repeat-dependent gene expression variation to phenotypic diversity. The Ssn6 TR2 variants show clear differences in fitness when grown on alternative carbon or nitrogen sources (Figure 3D). Importantly, these changes in fitness mirror the expression patterns of key metabolic genes induced during nutritional shifts (Figure 3A). We also observe changes in multicellular growth forms in the TR2 variants that are directly linked to changes in FLO11 expression (Figure 4D). These observations suggest that no single SSN6 repeat length offers optimal fitness in different environments. It is therefore tempting to speculate that repeat-dependent variation in gene expression might help to increase the phenotypic evolvability as it facilitates the emergence of variants fit in distinct environments. This would imply that the inherent instability of Q-rich repeats in TFs might in some cases be advantageous instead of purely neutral or even detrimental. Indeed, we observe that Ssn6 repeat expansion can, in some instances, have beneficial outcomes (e.g., higher fitness in alternative carbon sources). Chaperones like Ssa2 might help buffer the potentially negative effects of repeat variation, except in times of stress where the chaperone system can be overstretched, which can lead to the uncovering of the previously buffered effects of Q-repeat variation (Jarosz et al., 2010).

Our results agree with previous studies that show correlations between variation in TRs and phenotypes (Gemayel et al., 2010). A key study by Fondon and Garner (2004) implicated variable repeats in the regulatory gene Runx-2 as mediators of morphological evolution in domestic dogs. In Arabidopsis thaliana, variation in polyQ repeats of ELF3, a component of the circadian clock, also leads to morphological changes (Undurraga et al., 2012). Interestingly, in the human, mouse, and fly genomes, polyQ tracts are preferentially encoded by pure CAG repeats (Schafer et al., 2015).
Figure 5. TR2 Variation Modulates Ssn6 Solubility and Protein Interactions

(A) Identification of Ssn6 interactors. Shared and unique interacting proteins between WT Ssn6 (TR2-63) and the repeat deletion (TR2-0) or expansion (TR2-105) forms are shown. Quantification is shown of enriched or depleted proteins in pull-downs of representative TR2 variants (labeled with 13C isotope, H) relative to the WT (TR2-63) (labeled with 12C isotope, L). Data points represent mean ± SD, n = 2. p values were evaluated using the unpaired two-tailed t test. Asterisks indicate p < 0.05. See also Table S5.

(B) Co-localization of the Hsp70 chaperone Ssa2 and the expanded Ssn6 (TR2-105). Live cells of TR2 variants containing Ssn6-YFP and Ssa2-RFP fusion proteins were visualized by fluorescence and DIC microscopy.

(legend continued on next page)
et al., 2012). This might indicate purifying (negative) selection, as without such selection, repeats are expected to gather non-synonymous mutations and consequently become less variable (Legendre et al., 2007).

Understanding the Pathogenesis of Polyglutamine Repeat Expansions

While our study primarily describes the physiological role of polyQ repeats, there are also links to polyQ expansion pathogenesis. First, our study shows how repeat variation in a regulator leads to changes in the expression of downstream targets and how repeat expansion beyond a certain threshold could lead to a global transcriptional dysregulation, a common observation in multiple polyQ expansion diseases (Gatchel and Zoghbi, 2005). In addition, we show that expanded Ssn6 acquires private interactors (Figure 5A). In line with what has been previously documented, these interactors are mainly involved in transcription/translation, chromatin modification, and protein folding (Friedman et al., 2007; Lam et al., 2006; Park et al., 2013; Schaffar et al., 2004). The loss of key cellular regulators through aberrant interactions (Schaffar et al., 2004) could also result in transcriptional dysregulation.

Proteostasis is another factor contributing to polyQ pathogenesis. PolyQ aggregates may not be directly responsible for pathogenesis, but are rather remnants of an overwhelmed proteasome system (Douglas and Dillin, 2010). This “proteasome hijacking” mechanism is directly linked to the pathogenesis of polyQ aggregates (Park et al., 2013; Tsvetkov et al., 2013). Our study identifies the Hsp70 chaperone Ssa2 as a key functional interactor of the Ssn6-Tup1 complex. In the absence of Ssa2, we observe increased aggregation of expanded Ssn6 (TR2-105). These findings highlight the importance of molecular chaperones in counteracting the pathogenesis of polyQ expansions and consolidate the emerging notion that bolstering the proteostasis system might be a therapeutic target for neurodegenerative diseases (Tsvetkov et al., 2013).

In summary, our study establishes a direct link between Q-rich repeat variation and changes in transcription regulator function and demonstrates that repeats may be more than just potentially harmful elements in genomes. Instead, by modulating the proteostasis and the protein interactions of the regulator in which they reside, repeat domains may help tune the expression levels of target genes and diversify underlying phenotypes.
EXPERIMENTAL PROCEDURES

Molecular Techniques and Yeast Strains
Yeast genetic manipulations were performed using standard methods (Amberd et al., 2005). A detailed explanation of the strategy used to create the SSN6 repeat variants in the S21278a strain is provided in the Supplemental Experimental Procedures.

Transcriptome Analysis
Strains were grown overnight in YM 4% glucose then inoculated in fresh medium at an OD600 = 0.0006 and grown for 20 hr after which all samples had comparable (~2%) glucose levels left. Each culture was subsequently divided in two equal volumes. One pellet was frozen at ~80°C, and the other was used to inoculate fresh YM with no glucose and incubated for 2 hr at 30°C. Total RNA was isolated using the standard phenol-chloroform method (Amberd et al., 2005). For details on mRNA sequencing and analysis, see Supplemental Experimental Procedures.

Immunoprecipitation of Ssn6
Yeast cells were grown in YM 4% glucose until exponential phase, transferred to YM medium lacking any carbon source for 2 hr, and then harvested by centrifugation. Cell pellets were lysed by vortexing with glass beads in cold lysis buffer (50 mM Tris-HCl [pH 7.5], 100 mM NaCl, 1 mM EDTA, 10% glycerol, complete protease inhibitors; Roche), and extracts were cleared twice by centrifugation (10 min, 14,000 rpm at 4°C). Protein concentration in the soluble extracts was measured by the Protein Quantification kit (Sigma-Aldrich) and labeled and quantified by LC-MS/MS (see Supplemental Experimental Procedures).

Protein Aggregation Assay
Yeast cells were grown in YM 4% glucose until stationary phase, harvested by centrifugation, and lysed as described above. Extracts were cleared twice by centrifugation, and the supernatant contains the insoluble fraction. The resulting extracts was measured by the Protein Quantification kit (Sigma-Aldrich) and labeled and quantified by LC-MS/MS (see Supplemental Experimental Procedures).

REFERENCES


