UNIVERSITY OF CALIFORNIA, SAN DIEGO

Exploring mechanisms of trinucleotide expansion and studying the normal function of the ataxin-7 yeast ortholog Sgf73

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biology

by

Amanda Gayle Mason

Committee in Charge:

Professor Albert La Spada, Chair
Professor Lorraine Pillus, Co-Chair
Professor Eric Bennett
Professor Trey Ideker
Professor Tracy Johnson

2014

The Dissertation of Amanda Gayle Mason is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California, San Diego

2014

EPIGRAPH

"*If you knew what you were doing,*
it wouldn't be called research"
-Albert Einstein

"Getting something to work that has never been done before can be exceedingly
frustrating because you may never know how close you were to success,
and failures quite often teach you nothing."
-Craig Mello

TABLE OF CONTENTS

LIST OF SUPPLEMENTARY FILES

Supplementary Table 3: Significant genes from Cerebellar and Striatal microarray studies

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEGEMENTS

Chapter 1, in full, is a reprint of material that has been published in the Journal of Human Molecular Genetics. Amanda G. Mason, Stephanie Tomé, Jodie Simard, Randell T. Libby, Theodor K. Bammler, Richard P. Beyer, A. Jennifer Morton, Christopher E. Pearson & Albert R. La Spada, 2013. The thesis author was the primary investigator and author of this paper.

Chapter 2, in part, is a reprint of material that is currently submitted for publication. Mark A. McCormick*, Amanda G. Mason* Stephan J. Guyenet*, Weiwei Dang, Renee M. Garza, Marc K. Ting, Rick M. Moller, Shelley L. Berger, Matt Kaeberlein, Lorraine Pillus, Albert R. La Spada, and Brian K. Kennedy, 2014. The thesis author was one of the primary investigators and authors of this paper.

VITA

2008 Bachelor of Science in Human Biology, University of California, San Diego

2009 Master of Science in Biology, University of California, San Diego

2009-2010 Staff Research Associate, University of California, San Diego

2010-2014 Graduate Researcher, University of California, San Diego

2014 Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Exploring plant immune signaling through the Arabidopsis *Pseudomonas syringae* interaction. A thesis submitted in partial satisfaction of the requirements for the degree of Master of Science in Biology, 2009.

Mitochondrial dysfunction in NnaD mutant flies and Purkinje cell degeneration mice reveals a role for Nna proteins in neuronal bioenergetics. Chakrabarti L, Zahra R, Jackson SM, Kazemi-Esfarjani P, Sopher BL, **Mason AG**, Toneff T, Ryu S, Shaffer S, Kansy JW, Eng J, Merrihew G, MacCoss MJ, Murphy A, Goodlett DR, Hook V, Bennett CL, Pallanck LJ, La Spada AR.
Neuron. 2010 Jun 24;66(6):835-47.

Protein interaction analysis of senataxin and the ALS4 L389S mutant yields insights into senataxin post-translational modification and uncovers mutant-specific binding with the brain cytoplasmic RNA-encoded peptide. Bennett CL, Chen Y, Vignali M, Lo RS, **Mason AG,** Unal A, Saifee NPH, Fields S, La Spada AR.. PLoS one 8(11): e78837. Nov 2013 doi:10.1371/journal.pone.0078837

Expression levels of DNA replication and repair genes predict regional somatic repeat instability in the brain but are not altered by polyglutamine disease protein expression or age. **Mason AG,** Tomé S, Simard JP, Libby RT, Bammler TK, Beyer RP, Morton AJ, Pearson CE, La Spada AR. Hum. Mol. Genet. EPub ahead of print Nov 2013 PMID: 24191263

FIELDS OF STUDY

Major field of study: Biology

Studies in Human and Yeast Biology, Ph.D. thesis work
Professors Albert R. La Spada, and Lorraine Pillus

Studies in Plant Biology, Masters thesis work
Professor Steven P. Briggs

ABSTRACT OF THE DISSERTATION

Exploring mechanisms of trinucleotide expansion and studying the normal function of the ataxin-7 yeast ortholog Sgf73

by

Amanda Gayle Mason

Doctor of Philosophy in Biology

University of California, San Diego, 2014

Professor Albert R. La Spada, Chair
Professor Lorraine Pillus, Co- Chair

In this dissertation I explore trinucleotide repeat instability as well as the characterization of the ataxin-7 protein through the interrogation of its yeast ortholog Sgf73. Upon expansion of its CAG tract, ataxin-7 is known to cause the neurodegenerative disease, spinocerebellar ataxia type 7 (SCA7). Using microarray expression analysis, we sought to identify the genetic basis of regional instability differences between two different brain tissues, the striatum and the cerebellum. These

tissues are known to have high and low levels of trinuclotide expansion, respectively. We identified eight candidate genes involved in DNA replication and repair enriched in cerebellum, and validated five – *Pcna, Rpa1, Msh6*, *Fen1, and Lig1*. Expression levels of these candidates are significantly higher in the cerebellum than in the striatum, but were not dependent on disease status in a line of transgenic Huntington disease mice. In Chapters 2 and 3, we studied the function of the ataxin-7 yeast ortholog, Sgf73, a member of the deubiquitinase module (DUBm) in the SAGA transcriptional co-activator complex. It is known that the yeast *sgf73Δ* mutant has an extended replicative life span (RLS). We tested the role of different SAGA complex members in RLS and found that other DUBm mutants have increased RLS, but not to the degree seen in *sgf73Δ*. Deleting other non-DUBm SAGA components did not extend RLS. Additionally we found that the *sgf73Δ* RLS extension is dependent on the presence of the histone deacetylase, Sir2, which physically interacts with Sgf73. Lastly, we performed ChIP-Seq to determine the DNA binding locations of Sgf73. We identified 389 genomic regions bound by Sgf73, with an enrichment in the region 5' to ribosomal protein encoding genes. Additionally, 31 of the 389 identified Sgf73 genomic targets had previously been described to be involved in RLS. Using these data and transcriptome data from SAGA-related mutants, we were able to pinpoint several high-confidence Sgf73 targets for further studies on the mechanism of RLS in yeast as well as expanded ataxin-7 mediated neurodegeneration in humans.

General Introduction

Microsatellite repeat instability is the known cause of more than 40 genetic disorders. Among the microsatellite repeat diseases are a class of genetically inherited tri-nucleotide repeat (TNR) expansion disorders including CAG/CTG tri-nucleotide expansion disorders also known as poly-glutamine or poly Q disorders. Currently there are nine known poly Q repeat disorders caused by coding poly Q expansions at different loci: dentatorubral-pallidoluysian atrophy (DRPLA), Huntington's disease (HD), spinobulbar muscular atrophy (SBMA), and six forms of spinocerebellar ataxias (SCA1, 2, 3, 6, 7, and 17)[1]. The poly Q disorders display the phenomenon known as anticipation, whereby expanded repeat tracts have the propensity to further expand from one generation to the next, resulting in dramatic intergenerational expansions. There is an inverse correlation between repeat size and age at disease onset as well as disease severity, with disease manifesting earlier and being more severe in the individuals with larger repeats. Furthermore, there is a strong tendency for these expanded trinucleotide repeats to further expand rather than contract in somatic tissues resulting in extensive somatic mosaicism.

Somatic mosaicism refers to the variation in repeat length both between and within tissues, and it has been documented for a number of poly-Q disease alleles [2-4]. A shared feature of somatic mosaicism across multiple disorders is that there tends to be greater repeat length variability within the different regions of the central nervous system (CNS), with HD being the first CAG repeat disease found to show CNS somatic mosaicism [5]. Analysis of autopsied HD patient brains indicated that the striatum displayed the largest CAG repeats with occasional enormous expansions,

whereas the cerebellum contained CAG repeats that were on average much smaller [6]. Additional studies of DRPLA, DM1, SBMA, SCA1, and SCA3 have similarly shown that the largest repeat expansions are observed in striatum, and the smallest expansions are seen in cerebellum [4,5,7-11]. Thus the cerebellum is considered a very stable tissue with regard to CAG repeat length stability, whereas the striatum is an unstable tissue, showing extensive instability with larger expansions occurring with advancing age.

Spinocerebellar ataxia type 7 (SCA7) is one of the aforementioned nine genetically inherited neurodegenerative disorders caused by a coding CAG/polyglutamine repeat expansion, with the CAG track being within the ataxin-7 gene. The poly Q tract in ataxin-7 ranges in size from 4 - 35 glutamines in unaffected individuals, and is expanded to 37 to more than 300 glutamines in affected patients [12,13]. It is known that ataxin-7 undergoes extensive anticipation of ~20years/generation [14]. With no current treatments for SCA7 the translated polyQ track causes ataxin-7 to be aggregation prone. Ataxin-7 protein accumulation then causes neuronal dysfunction and death in the retina, cerebellum, and cerebellar-associated structures. In addition findings in SCA7 mice have implicated that expanded ataxin-7 protein has altered functions that contribute to SCA7 disease pathogenesis [15]. Ataxin-7 is a highly conserved member of the Spt3-Taf9-Gcn5 acetyltransferase complex (STAGA), a major transcriptional coactivator complex in mammalian cells [16,17], and is part of the USP22 histone deubiquitinase module. *SGF73* was identified as the *Saccharomyces cerevisiae* ortholog of ataxin-7, and was subsequently identified as a member of the STAGA complex in yeast equivalent, Spt-

Ada-Gcn5 acetyltransferase complex (SAGA) [16,18], and is a member of the Ubp8 histone deubiquitinase module[19-21].

The budding yeast *S. cerevisiae* has long been used as a model organism for studies into a number of human processes due to the high level of conservation between the organisms and its easy genetic manipulation. In particular yeast have proven to be valuable in elucidating molecular mechanisms behind aging, wherein the cell replicative lifespan (RLS) is measured by the number of daughter cells one mother cell can produce [22]. Additionally yeast have proven to be a useful in studying neurodegenerative processes (reviewed[23]) These previous sucessful model studies made *S. cerevisiae* an ideal organism to further study ataxin-7 function through the interrogation of its yeast ortholog *SGF73*. The deletion of *SGF73* results in drastic RLS extension, thereby providing a phenotype for mechanistic studies to unravel the role of *SGF73* in aging. Neurodegenerative disorders have many common features associated with the normal aging processes including: accumulation of oxidative stress, declining mitochondrial function, telomere erosion, protein inclusion formation, and lowered resistance to excitotoxic, genotoxic and metabolic stress. Therefore further deciphering the role of *SGF73* will have implications for both the yeast aging field as well as for the field of neurodegeneration.

There are many unanswered questions for the role of ataxin-7 in both a diseased and non-diseased state and using *S. cerevisiae* as a model allows for examination through a number of genetic and biochemical experiments. In particular SAGA and STAGA are transcriptional co-activators[16,17 22], with Sgf73 having a zinc-finger domain known to bind nucleosomes [24], thus Sgf73 is likely instrumental in

SAGA's transcriptional activities. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is a method to identify the locations in the genome that are bound by a DNA binding protein[25]. Therefore, an understanding of Sgf73 occupancy and associated proteins occupancy can be obtained using ChIP-seq. This approach uncovers genes whose regulation may be important to yeast aging and that will likely correlate to ataxin-7 regulation, thus sheading new light on ataxin-7 function and its role in SCA7.

This thesis describes an exploration into general repeat instability wherein microarray expression analysis was used to identify genes whose differential expression levels in striatum and cerebellum of aged wild-type C57BL/6J mice might explain the regional instability differences seen in poly Q disease models. Studies on the normal function of ataxin-7 were then completed through the interrogation of its yeast ortholog Sgf73, by exploring the molecular basis of replicative lifespan extension in *sgf73Δ* yeast. Lastly, I describe studies into the chromatin occupancy of Sgf73 and Ubp8 to explore regulation of genes that may prove to be essential to yeast aging as well as the study of neurodegeneration processes.

References:

1. Orr, H.T. & Zoghbi, H.Y. Trinucleotide repeat disorders. *Annu Rev Neurosci* **30**, 575-621 (2007).

2. Cleary, J.D. & Pearson, C.E. The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence. *Cytogenet Genome Res* **100**, 25-55 (2003).

3. La Spada, A.R. Trinucleotide repeat instability: genetic features and molecular mechanisms. *Brain Pathol* **7**, 943-63 (1997).

4. Aoki, M., Abe, K., Tobita, M., Kameya, T., Watanabe, M. & Itoyama, Y. Reduction of CAG expansions in cerebellar cortex and spinal cord of DRPLA. *Clin Genet* **50**, 199-201 (1996).

5. Telenius, H., Kremer, B., Goldberg, Y.P., Theilmann, J., Andrew, S.E., Zeisler, J., Adam, S., Greenberg, C., Ives, E.J., Clarke, L.A. & et al. Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat Genet* **6**, 409-14 (1994).

6. Kennedy, L., Evans, E., Chen, C.M., Craven, L., Detloff, P.J., Ennis, M. & Shelbourne, P.F. Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum Mol Genet* **12**, 3359-67 (2003).

7. Chong, S.S., McCall, A.E., Cota, J., Subramony, S.H., Orr, H.T., Hughes, M.R. & Zoghbi, H.Y. Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* **10**, 344-50 (1995).

8. Hashida, H., Goto, J., Kurisaki, H., Mizusawa, H. & Kanazawa, I. Brain regional differences in the expansion of a CAG repeat in the spinocerebellar ataxias: dentatorubral-pallidoluysian atrophy, Machado-Joseph disease, and spinocerebellar ataxia type 1. *Ann Neurol* **41**, 505-11 (1997).

9. Lopes-Cendes, I., Maciel, P., Kish, S., Gaspar, C., Robitaille, Y., Clark, H.B., Koeppen, A.H., Nance, M., Schut, L., Silveira, I., Coutinho, P., Sequeiros, J. & Rouleau, G.A. Somatic mosaicism in the central nervous system in spinocerebellar ataxia type 1 and Machado-Joseph disease. *Ann Neurol* **40**, 199-206 (1996).

10. Lopez Castel, A., Nakamori, M., Tome, S., Chitayat, D., Gourdon, G., Thornton, C.A. & Pearson, C.E. Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues. *Hum Mol Genet* **20**, 1-15 (2011).

11. Ueno, S., Kondoh, K., Kotani, Y., Komure, O., Kuno, S., Kawai, J., Hazama, F. & Sano, A. Somatic mosaicism of CAG repeat in dentatorubral-pallidoluysian atrophy (DRPLA). *Hum Mol Genet* **4**, 663-6 (1995).

12. David, G., Abbas, N., Stevanin, G., Durr, A., Yvert, G., Cancel, G., Weber, C., Imbert, G., Saudou, F., Antoniou, E., Drabkin, H., Gemmill, R., Giunti, P., Benomar, A., Wood, N., Ruberg, M., Agid, Y., Mandel, J.L. & Brice, A. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat Genet* **17**, 65-70 (1997).

13. Stevanin, G., Durr, A. & Brice, A. Clinical and molecular advances in autosomal dominant cerebellar ataxias: from genotype to phenotype and physiopathology. *Eur J Hum Genet* **8**, 4-18 (2000).

14. Lebre, A.S. & Brice, A. Spinocerebellar ataxia 7 (SCA7). *Cytogenet Genome Res* **100**, 154-63 (2003).

15. Palhan, V.B., Chen, S., Peng, G.H., Tjernberg, A., Gamper, A.M., Fan, Y., Chait, B.T., La Spada, A.R. & Roeder, R.G. Polyglutamine-expanded ataxin-7 inhibits STAGA histone acetyltransferase activity to produce retinal degeneration. *Proc Natl Acad Sci U S A* **102**, 8472-7 (2005).

16. Helmlinger, D., Hardy, S., Sasorith, S., Klein, F., Robert, F., Weber, C., Miguet, L., Potier, N., Van-Dorsselaer, A., Wurtz, J.M., Mandel, J.L., Tora, L. & Devys, D. Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. *Hum Mol Genet* **13**, 1257-65 (2004).

17. Martinez, E., Palhan, V.B., Tjernberg, A., Lymar, E.S., Gamper, A.M., Kundu, T.K., Chait, B.T. & Roeder, R.G. Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors in vivo. *Mol Cell Biol* **21**, 6782-95 (2001).

18. Timmers, H.T. & Tora, L. SAGA unveiled. *Trends Biochem Sci* **30**, 7-10 (2005).

19. Kohler, A., Schneider, M., Cabal, G.G., Nehrbass, U. & Hurt, E. Yeast Ataxin-7 links histone deubiquitination with gene gating and mRNA export. *Nat Cell Biol* **10**, 707-15 (2008).

20. Lee, K.K., Swanson, S.K., Florens, L., Washburn, M.P. & Workman, J.L. Yeast Sgf73/Ataxin-7 serves to anchor the deubiquitination module into both SAGA and Slik(SALSA) HAT complexes. *Epigenetics Chromatin* **2**, 2 (2009).

21. Zhao, Y., Lang, G., Ito, S., Bonnet, J., Metzger, E., Sawatsubashi, S., Suzuki, E., Le Guezennec, X., Stunnenberg, H.G., Krasnov, A., Georgieva, S.G., Schule, R., Takeyama, K., Kato, S., Tora, L. & Devys, D. A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. *Mol Cell* **29**, 92-101 (2008).

22. Mortimer, R.K. & Johnston, J.R. Life span of individual yeast cells. *Nature* **183**, 1751-1752 (1959).

23. Outeiro, T.F. & Giorgini, F. Yeast as a drug discovery platform in Huntington's and Parkinson's diseases. *Biotechnol J* **1**, 258-69 (2006).

24. Bonnet, J., Wang, Y.H., Spedale, G., Atkinson, R.A., Romier, C., Hamiche, A., Pijnappel, W.W., Timmers, H.T., Tora, L., Devys, D. & Kieffer, B. The structural plasticity of SCA7 domains defines their differential nucleosome-binding properties. *EMBO Rep* **11**, 612-8 (2010).

25. Kharchenko, P.V., Tolstorukov, M.Y. & Park, P.J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* **26**, 1351-9 (2008).

Chapter 1: Expression levels of DNA replication and repair genes predict regional somatic repeat instability in the brain but are not altered by polyglutamine disease protein expression or age

Abstract:

Expansion of CAG/CTG trinucleotide repeats cause numerous inherited neurological disorders, including Huntington's disease (HD), several spinocerebellar ataxias, and myotonic dystrophy type 1. Expanded repeats are genetically unstable with a propensity to further expand when transmitted from parents to offspring. For many alleles with expanded repeats, extensive somatic mosaicism has been documented. For CAG repeat diseases, dramatic instability has been documented in the striatum, with larger expansions noted with advancing age. By contrast, only modest instability occurs in the cerebellum. Using microarray expression analysis, we sought to identify the genetic basis of these regional instability differences by comparing gene expression in the striatum and cerebellum of aged wild-type C57BL/6J mice. We identified eight candidate genes enriched in cerebellum, and validated four – *Pcna, Rpa1, Msh6*, and *Fen1* – along with a highly associated interactor, *Lig1*. We also explored whether expression levels of mismatch repair (MMR) proteins are altered in a line of HD transgenic mice, R6/2, that is known to show pronounced regional repeat instability. Compared to wild-type littermates, MMR expression levels were not significantly altered in R6/2 mice regardless of age. Interestingly, expression levels of these candidates were significantly increased in the cerebellum of control and HD human samples in comparison to striatum. Together,

our data suggest that elevated expression levels of DNA replication and repair proteins in cerebellum may act as a safeguard against repeat instability, and may account for the dramatically reduced somatic instability present in this brain region, compared to the marked instability observed in the striatum.

Introduction:

Expansions of CAG/CTG trinucleotide repeat sequences (TNRs) cause numerous inherited neurological disorders, including Huntington's disease (HD), myotonic dystrophy (DM1), and several spinocerebellar ataxias (SCAs) [1,2]. Longer CAG repeat sequences inversely correlate with worsening disease severity and earlier age of onset. TNR tracts less than 35 units are generally stable, while tracts greater than 35 units, referred to as expanded trinucleotide repeats, become unstable and have the propensity to expand further. The molecular events that underlie trinucleotide repeat instability are still poorly understood [3]. For some expanded trinucleotide repeat alleles, extensive 'somatic mosaicism' has been documented, wherein variation in repeat length (both between and within tissues) exists [4,5]. Somatic mosaicism has been shown to be age-dependent, highly tissue-specific, and associated with disease progression [2,6].

The CAG repeat diseases display significant somatic mosaicism, with HD being the first CAG repeat disease documented to show this phenomenon [7]. One striking feature of repeat mosaicism in HD CNS is the non-random differential pattern of expanded repeats present between different central nervous system (CNS) regions. Analysis of autopsied HD patient brains and transgenic mice indicates that the

striatum displays the largest range of CAG repeats, while the cerebellum contains CAG repeats that are much smaller [8]. Thus, instability is greater in the striatum than in the cerebellum in HD. Studies of dentatorubral-pallidoluysian atrophy (DRPLA), DM1, SBMA, SCA1, and SCA3 have also demonstrated that the smallest expansions occur in the cerebellum [1,9-14]. These profiles parallel the patterns observed in CNS tissues in HD, despite the fact that different populations of neurons degenerate in each of these six diseases. This suggests that CNS somatic mosaicism patterns may be independent of the disease process or repeat locus and that instead, the cellular characteristics of a tissue type dictate differences in the degree of repeat instability.

Mouse models of CAG/CTG repeat instability accurately recapitulate somatic instability patterns documented in human patients, with the extent of mosaicism observed in different somatic tissues and brain regions mimicking that seen in human patients, although they do not accurately model the dramatic intergenerational expansions seen in human disease [15-19]. PCR analysis of repeat length variation in striata from HD and SCA1 mouse models revealed dramatic instability - with rare changes of >150 CAG repeats, whereas cerebellar repeat variation was modest by comparison [19,20]. In these disease models and a SCA7 mouse model, age correlates with somatic instability, with extensive variation in repeat length size only clearly emerging once the mice are well into adulthood, with prominent repeat instability noted both in the CNS (cortex, striatum) and periphery (kidney, liver) [15,19,20]. It is well known that CAG repeat-expanded RNAs and polyglutamine-expanded proteins result in altered splicing and expression of many genes [21-24]. Yet the presence of cell-type and age-specific repeat instability in the cerebellum and striatum in multiple different

diseases suggests that tissue-specific differences in the expression of *trans*-acting protein factors may account for differences in somatic instability between these two brain regions. In this work, we considered repeat instability mechanisms operating in different regions of the brain, as most CAG/CTG repeat diseases are characterized by progressive neurodegeneration accompanied by differences in expanded repeat lengths in various CNS regions. The fact that somatic mosaicism of expanded repeats appears invariant in certain brain regions led us to ask whether comparison of a highly unstable brain region (striatum) with a highly stable brain region (cerebellum) might yield insights into the *trans*-acting factors that either promote instability or prevent it.

Results:

Microarray expression analysis of cerebellum and striatum

To identify differentially regulated transcripts that might underlie somatic instability differences between the cerebellum and striatum, we performed a microarray expression analysis. Somatic instability increases dramatically with aging; thus, we studied wild-type C57BL/6J mice at 40 weeks of age, as marked regional instability differences emerge in most repeat disease mouse models by this time. For this study, we dissected out the striatum and cerebellum of six mice, three males and three females, and isolated total RNA. We then analyzed these samples using the Affymetrix Murine 430A GeneChip array, containing ~14,000 annotated mouse genes. To assess the validity of this data set, we generated a microarray correlation heat map of the transformed sample intensities. Importantly, we found that the transcriptional signatures of the six individual cerebellar samples were highly

correlative and the transcriptional signatures of the six striatal samples were also highly correlative, but they were not correlative with each other (Figure 1-1A). In addition we found that there was no profile clustering based on mouse gender (Figure 1-1A). This indicates that the samples represent reliable transcriptional profile replicates within each tissue, while there are marked differences between the two tissue types, as expected. By referencing a C57BL/6J *in situ* hybridization expression library (www.alleninstitute.org; http://mouse.brain-map.org; [25], we further validated the reliability of our microarray findings by examining the genes with the highest expression in each tissue from our microarray, and observing that these genes were respectively highly expressed cerebellar genes (Supplementary Figure 1-1A and B; Supplementary Table 1-1), and highly expressed striatal genes (Supplementary Figure 1-1C and D; Supplementary Table 1-2). Analysis of the microarray expression array data set was performed using Affymetrix expression console software. By applying Robust Multichip Average (RMA) analysis [26-28], we obtained intensity values for each probe. Probe values for the cerebellar data set were then compared to probe values for the striatal data set, using a paired t-test to identify probes that are differentially expressed among the tissue types. In this way, we identified 2,655 differentially expressed genes with a p-value of <0.002.

Identification of differentially regulated pathways

To investigate which pathways are down regulated in the striatum in comparison to cerebellum, we used DAVID (Database for Annotation, Visualization and Integrated Discovery) to analyze all significantly regulated genes [29,30], with a

threshold significance of p <0.002 and a fold difference of at least 1.4, which yielded 929 genes (Supplementary Table 1-3). KEGG (Kyoto Encyclopedia of Genes and Genomes) Pathway analysis output [31,32] highlighted a number of pathways (Supplementary Table 1-4), with two pathways emerging as enriched in our data set: 1) DNA replication (*Prim1*, *Mcm7*, *Rpa1*, *Pcna, Rfc4*, *Fen1*, and *Pold2*) genes, and 2) mismatch repair (*Msh6*, *Pold2*, *Pcna*, *Rpa1*, and *RFfc4*) genes (Table 1-1). These pathways have previously been implicated as key modulators of repeat instability [1,2,33-35]. Of the genes listed in Table 1-1, we selected a subset (*Pcna*, *Rpa1*, and *Msh6*) with well-established roles in repeat instability [3,36] to study in detail.

Pathway analysis underscores importance of DNA replication and DNA repair in somatic instability differences

Using the 'Search Tool for the Retrieval of INteracting Genes/proteins' (STRING v9.0) [37] analysis to interrogate the protein interaction network centered at PCNA, RPA1, and MSH6, we noted that the resulting network highlights essential components of DNA replication and repair (Figure 1-1B). Our STRING analysis yielded a network of predicted functional partners of PCNA, RPA1, and MSH6, including LIG1 and FEN1. The most highly associated interacting genes/proteins all exhibited higher expression levels in the cerebellum in comparison to the striatum in our data set (Supplementary Table 1-5). The majority of the genes revealed by STRING analysis belong to the KEGG Pathway for the 'DNA replication complex'. The DNA replication complex is an enriched pathway in the cerebellum compared to the striatum in the microarray data, revealing that a substantial number of key players

in the pathway are differentially regulated between the two tissue types. Proteins involved in DNA replication are critically important for maintaining DNA integrity, when DNA damage occurs [1]. Our data thus suggest that increased expression of DNA metabolism genes in the cerebellum may contribute to the stability of CAG repeat expansions in this brain region.

Confirmation of RNA expression level changes for *Pcna*, *Rpa1*, *Msh6, Lig1, and Fen1*

To confirm that the DNA replication complex is enriched in the cerebellum in comparison to the striatum, we performed real-time RT-PCR analysis on *Pcna*, *Rpa1*, *Msh6, Lig1, and Fen1* using the TaqMan approach [38]. Significant enrichment ($p < 0.01$) was detected for *Pcna*, *Rpa1*, *Msh6, Lig1, and Fen1* in the cerebellum compared to the striatum in 40 week-old C57BL/6J mice (Figure 1-1C). These results confirm the alterations reported for the microarray expression comparisons, and further validate the findings of this microarray study.

PCNA, RPA1, and mismatch repair proteins are markedly enriched in cerebellum

As gene transcript expression levels may not correlate with protein expression levels [39], we chose to analyze cerebellar and striatal protein levels of PCNA, RPA1, and the three principal mismatch repair (MMR) proteins, MSH2, MSH3, and MSH6, as they form a set of multimeric complexes: MutSa (MSH2-MSH6) and MutSb (MSH2-MSH3). Protein lysates were obtained from the cerebellum and striatum of

six 40 week-old wild-type C57BL/6J mice, and MMR protein levels were measured by simultaneous Western blot analysis, an established and previously published for sensitive detection of MMR expression level variation [39-42]. Using this approach, we found that all three MMR proteins were markedly enriched in the cerebellum compared to the striatum (Figure 1-2A). MSH6 protein expression was enriched by 2.3-fold, MSH3 protein expression was enriched by 2.3-fold, and MSH2 protein expression was enriched by 4.4-fold in the cerebellum vs. the striatum (Figure 1-2A). Western blot analysis of RPA1 and PCNA also revealed higher protein expression levels in the cerebellum (Figure 1-2B and C). Cerebellar RPA1 protein expression was enriched by 3.6-fold and cerebellar PCNA protein expression was enriched by ~13-fold (Figure 1-2B and C).

Brain region-specific differences in *Pcna*, *Rpa1*, *Msh6, Lig1, and Fen1* are age-independent

The cerebellum and striatum have very different cellular compositions; hence, expression differences in the two tissues could be due to age or to the intrinsic nature of the tissues. We therefore dissected the striatum and cerebellum from six 8-week-old C57BL/6J mice, three males and three females, using half of the tissue for RNA isolation and the other half for protein extraction. As in the aged tissues, we observed significantly increased RNA expression levels ($p < 0.005$) for *Pcna*, *Rpa1*, *Msh6, Lig1, and Fen1* in the cerebellum in comparison to the striatum in 8 week-old mice (Figure 1-3A). Western blot analysis of RPA and PCNA similarly revealed considerably higher protein expression levels in cerebellum (Figure 1-3B). We also

found that MSH2 and MSH3 proteins are markedly enriched in the cerebellum compared to the striatum, while MSH6 expression levels were too low to permit quantitative comparison (Figure 1-3B).

MMR expression levels are not affected by polyglutamine disease protein expression or age

Although the expression levels of MMR genes did not change within a tissue with age, we wondered if the presence of an expressed CAG repeat expansion might alter the levels of these protein factors, especially since MMR proteins have been directly implicated in somatic repeat instability [17]. To address this, we measured the expression levels of *Msh2, Msh3 and Msh6* in the striatum and cerebellum of wild-type (WT) control mice and in HD R6/2 transgenic mice, a line that carries a 250 CAG repeat and shows marked age-dependent instability in CAG repeat length [16]. We found that the RNA expression levels of *Msh2*, *Msh3* and *Msh6* do not significantly change in R6/2 mice at two different ages, 13 weeks (pre-symptomatic) and 42 weeks (symptomatic with documented somatic mosaicism), and noted that MMR RNA expression levels in HD R6/2 mice were similar to the levels measured in WT control mice (Figure 1-4). These findings indicate that MMR gene expression is not altered by polyglutamine-expanded protein expression or on-going disease. We also performed quantitative Western blot analysis, and did not observe any effect of the expressed polyglutamine-expanded protein on MSH2, MSH3 and MSH6 protein levels between 13 week-old or 42 week-old R6/2 mice and WT controls (Figure 1-5).

Tissue-specific patterns of CAG instability and MMR expression levels show no correlation

We next examined whether the expression of MMR protein levels might correlate with the levels or pattern of trinucleotide repeat instability in tissues of HD mice. Using simultaneous Western blot analysis to obtain the relative expression levels of MSH2, MSH3 and MSH6 proteins in different tissue samples, we found that MSH2, MSH3 and MSH6 protein levels vary widely between tissues from WT control mice, with MMR protein expression highest in thymus and spleen, intermediate in striatum and liver, and lowest in skeletal muscle and heart (Figure 1-6A). When we compared the level of MMR protein expression in different mouse tissues to the extent of somatic repeat instability documented for these same tissues in HD R6/2 mice, we found that for certain tissues (cerebellum, spleen, thymus), high MMR expression correlates with reduced somatic mosaicism, but for other tissues (kidney, muscle), MMR expression level does not correlate with somatic repeat instability (Figure 1-6B). These results underscore the complexity of the repeat instability process, and indicate that MMR protein expression is but one factor in determining the extent of repeat instability in a given tissue.

Expression levels of *PCNA*, *RPA1*, *MSH6, LIG1, and FEN1* in human striatum and cerebellum parallel expression levels observed in mice

To determine if these findings are potentially relevant to human patients afflicted with CAG/CTG repeat expansion disorders, we evaluated the levels of candidate DNA metabolism genes in adult cerebellum and striatum RNA samples

isolated from unaffected human individuals as well as from HD patients. Real-time RT-PCR analysis of human cerebellum and striatum samples revealed increased expression of *RPA1*, *MSH6, LIG1, and FEN1* in the cerebellum in comparison to the striatum for these DNA metabolism and repair genes in both unaffected controls and in affected HD patients (Figure 1-7A and B). Interestingly, although *PCNA* was more highly expressed in the cerebellum of unaffected controls in comparison to striatum, *PCNA* expression levels in HD cerebellum and striatum were comparable.

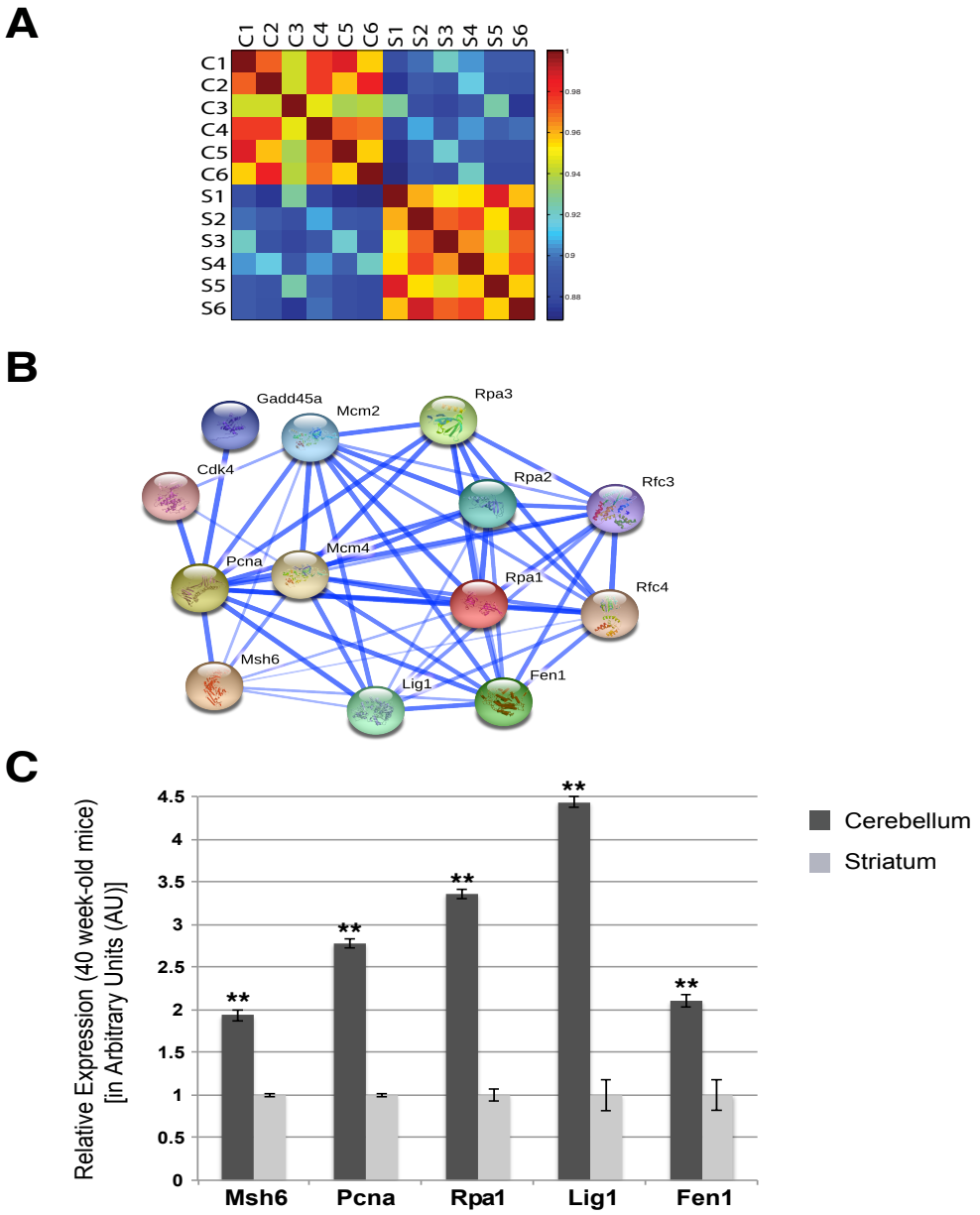


Figure 1-1: DNA replication and repair pathway genes are differentially expressed in cerebellum and striatum.

A) Microarray correlation heat map of transformed intensities from the six different cerebellar samples (C1-C6) and six striatal samples (S1-S6). Strong correlations are seen between samples from the same tissue type, while there is a weak correlation between the different tissue types.

B) STRING protein interaction network centered on PCNA, RPA1, and MSH6, which are differentially expressed in cerebellum in comparison to the striatum, highlights key players in DNA replication and repair.

C) Real-time RT-PCR confirmation of microarray expression differences between *Rpa1, Pcna*, *Msh6, Lig1, and Fen1* in 40 week-old mice (n = 6; mean ± s.e.m., 3 independent experiments; $^{**}P < 0.01$; t-test).

Table 1-1: Candidate Trans Acting Factors of DNA Replication, and Mismatch Repair

Gene Name	**Symbol**	**Probe**	**Fold Change (Cb/Str)**	**P-Value**
Replication Protein A1	*Rpa1*	1437309_a_at	↑1.99	6.6E-5
MutS homolog 6	*Msh6*	1416915_at	↑1.84	0.0017
Proliferating cell nuclear antigen	*PCNA*	1417947_at	↑1.50	1.9E-5
Flap structure specific endonuclease 1	*Fen1*	1421731_a_at	↑1.68	0.0005
Minichromosome maintenance deficient 7 (S. cerevisiae)	*Mcm7*	1438320_s_at	↑1.53	2.2E-5
Replication factor C (activator 1) 4	*Rfc4*	1438161_s_at	↑1.55	0.0001
Polymerase (DNA directed), delta 2, regulatory subunit	*Pold2*	1448277_at	↑1.42	0.0001
DNA primase, p49 subunit	*Prim1*	1418369_at	↑1.42	6.4E-5

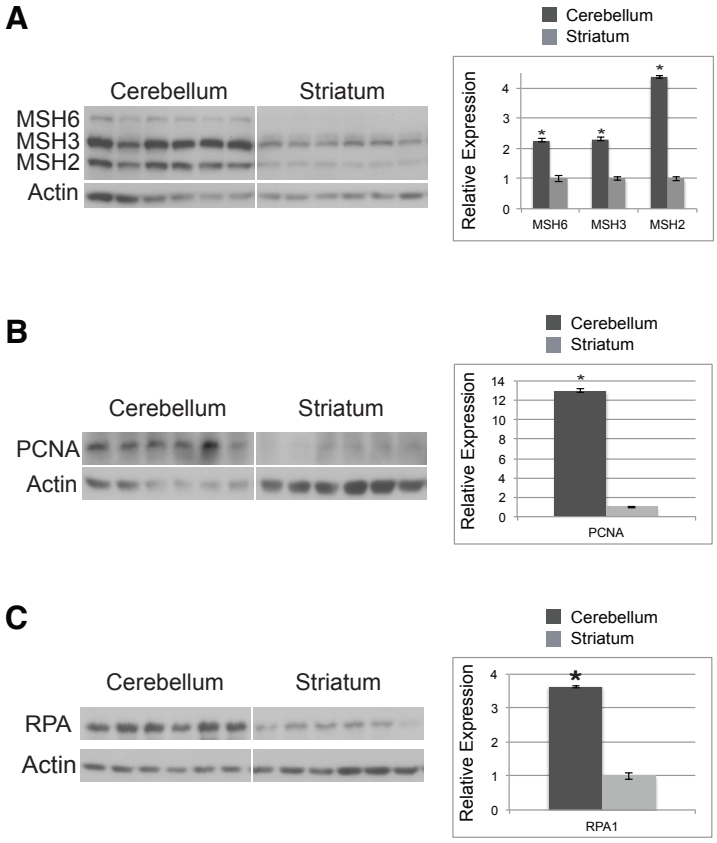


Figure 1-2: DNA replication and repair protein expression levels are significantly increased in the cerebellum.

A) Mismatch repair (MMR) protein expression levels in the cerebellum and the striatum from six age-matched C57BL/6 mice were analyzed by the simo-blot technique. MMR protein levels were quantified by Image J, normalized to actin, and presented to the right in arbitrary units, with striatal expression level for each MMR protein arbitrarily set to 1. Cerebellar MMR protein expression is significantly higher for each MMR protein in comparison to striatum MMR protein expression (mean ± s.e.m., 3 independent experiments; $*P < 0.05$; t-test).

B) Proliferating cell nuclear antigen (PCNA) protein expression levels in the cerebellum and the striatum from six age-matched C57BL/6 mice were analyzed by Western blot analysis. PCNA protein levels were quantified by Image J, normalized to actin, and presented to the right in arbitrary units, with striatal expression level arbitrarily set to 1. Cerebellar PCNA protein expression is significantly higher in comparison to striatum PCNA protein expression (mean ± s.e.m., 3 independent experiments; $*P < 0.05$; t-test).

C) Replication protein A1 (RPA1) protein expression levels in the cerebellum and the striatum from six age-matched C57BL/6 mice were analyzed by Western blot analysis. RPA protein levels were quantified by Image J, normalized to actin, and presented to the right in arbitrary units, with striatal expression level arbitrarily set to 1. Cerebellar RPA protein expression is significantly higher in comparison to striatum RPA protein expression (mean ± s.e.m., 3 independent experiments; $*P < 0.05$; t-test).

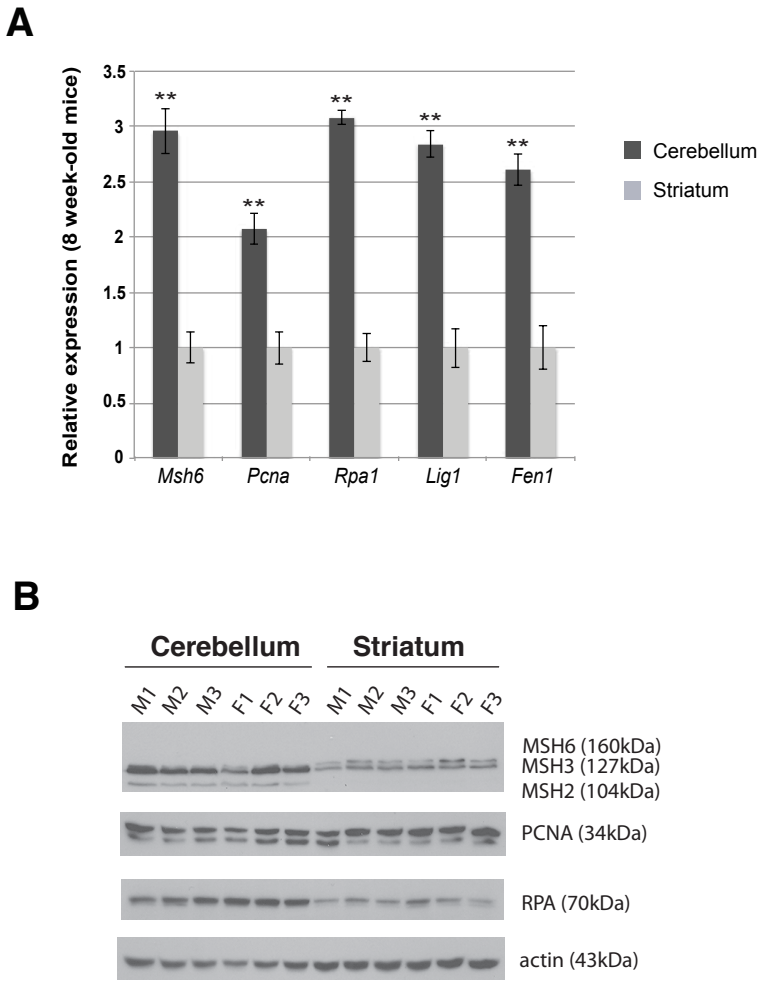


Figure 1-3: DNA replication and repair genes display elevated RNA and protein expression in the cerebellum in young mice.

A) The RNA expression levels for *Rpa1, Pcna*, *Msh6, Lig1 and Fen1* were quantified by real-time RT- PCR analysis for sets of cerebellum and striatum samples (n = 6 / group) for 8 week-old mice. Significant mean elevations for all five genes were detected in the cerebellum as compared to the striatum (mean ± s.e.m., 3 independent experiments; $^{**}P < 0.01$; t-test). b-actin served as an internal normalization control.

B) Western blot analysis of MSH2, MSH3, MSH6, PCNA, and RPA1 proteins in the cerebellum of 8 week-old C57BL/6J mice confirmed elevated RNA expression levels.

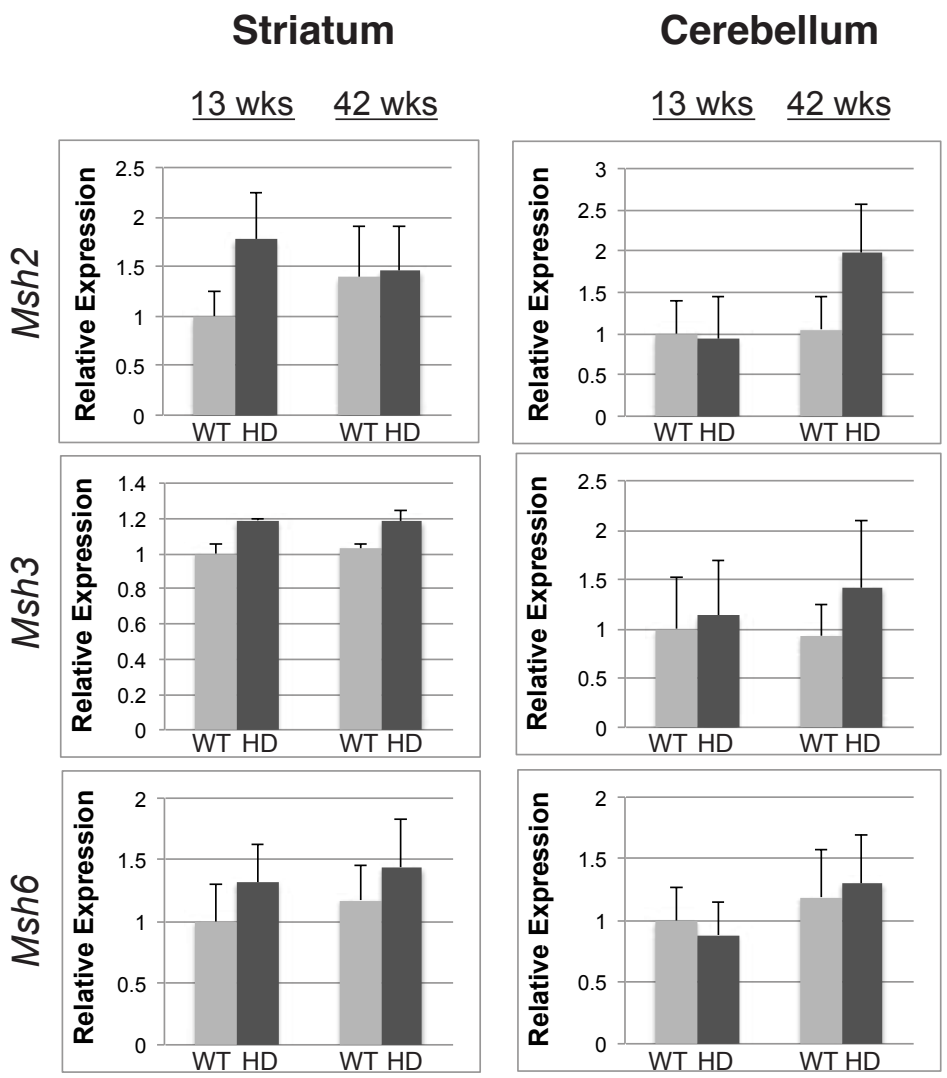


Figure 1-4: Polyglutamine-expanded huntingtin does not alter MMR RNA expression levels in cerebellum and striatum, even in aged mice.
Real-time RT-PCR analysis of *Msh2, Msh3,* and *Msh6* in the striatum and cerebellum of HD R6/2 mice (HD), carrying a hyper-expanded CAG repeat expansion (255 – 258 units at 13 weeks; 347 – 350 units at 43 weeks), and for littermate non-transgenic wild-type (WT) controls (n = 3 / group). RNA expression levels for each tissue sample set for a given target (i.e. each "box") are given relative to the 13 week-old WT result, which was arbitrarily set to 1 (mean ± s.e.m., 3 independent experiments*; P* = n.s.; t-test). 18S RNA served as the internal control.

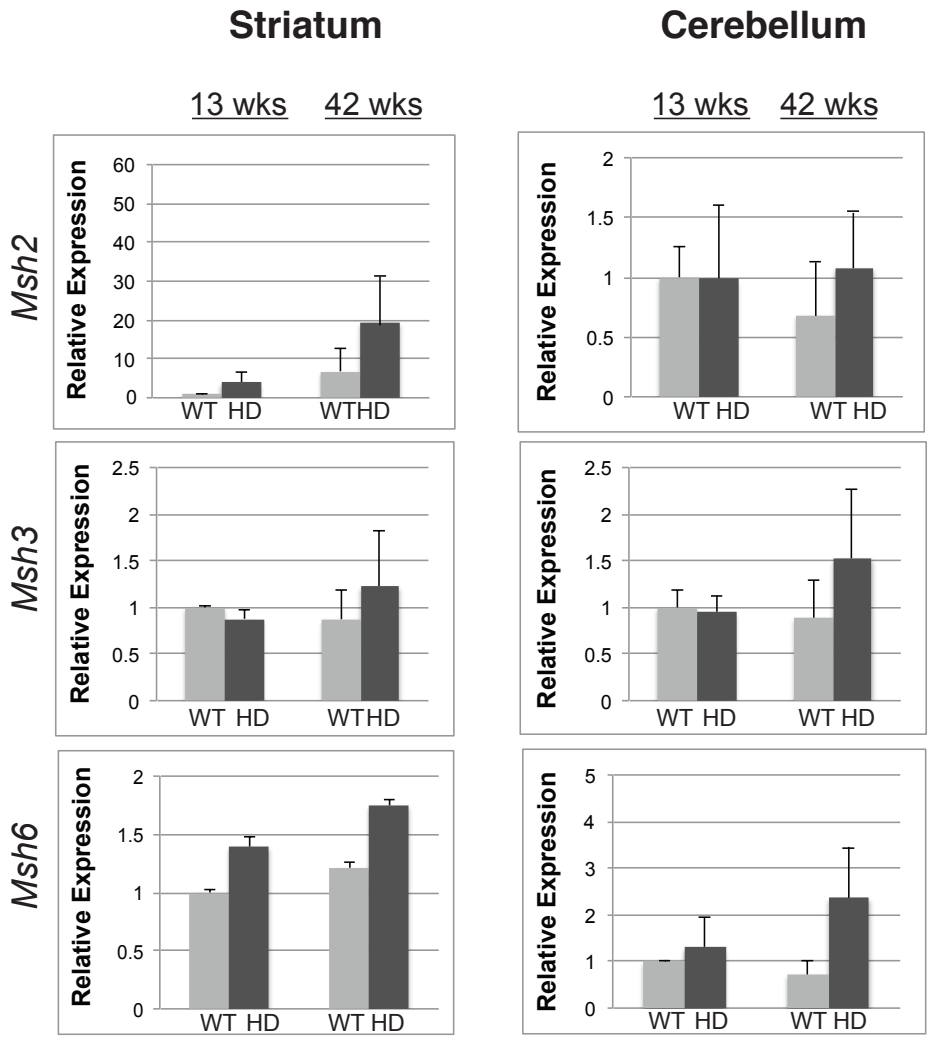


Figure 1-5: Polyglutamine-expanded huntingtin does not alter MMR protein expression levels in cerebellum and striatum, even in aged mice.
Western blot analysis of MSH2, MSH3, and MSH6 proteins in the striatum and cerebellum for HD R6/2 mice (HD), carrying a hyper-expanded CAG repeat expansion (255 – 258 units at 13 weeks; 347 – 350 units at 43 weeks), and for littermate non-transgenic wild-type (WT) controls (n = 3 / group). b-actin was used as a loading control for expression level normalization, and protein expression levels for each tissue sample set for a given target (i.e. each "box") are given relative to the 13 week-old WT result, which was arbitrarily set to 1 (mean ± s.e.m., 3 independent experiments; P = n.s.; t-test).

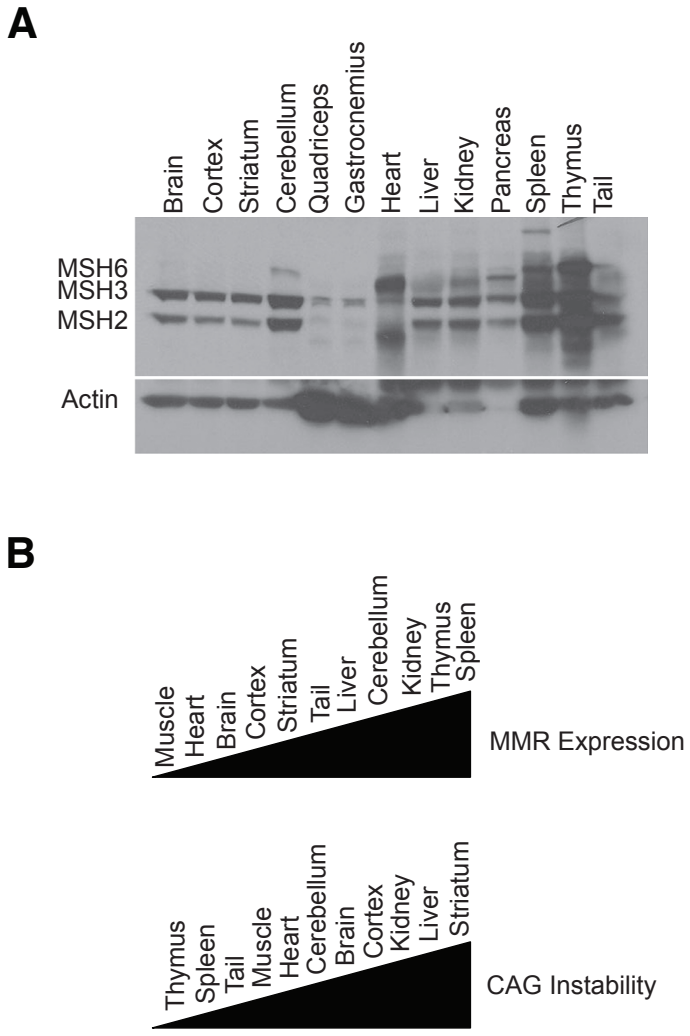


Figure 1-6: MMR protein levels in various tissues and their relationship to CAG instability

A) Simultaneous Western blot analysis of MMR protein levels in tissues from a 14 week-old WT CBA x C57BL/6J mouse. b-actin was used as the loading control.

B) Illustration of the rank order of MMR expression levels in 13 tissues from 14 week-old WT CBA x C57BL/6J mouse and the extent of CAG repeat instability in the same tissues of 13 week-old R6/2 transgenic mice [16]. Muscle corresponds to quadriceps and gastrocnemius, which were comparable.

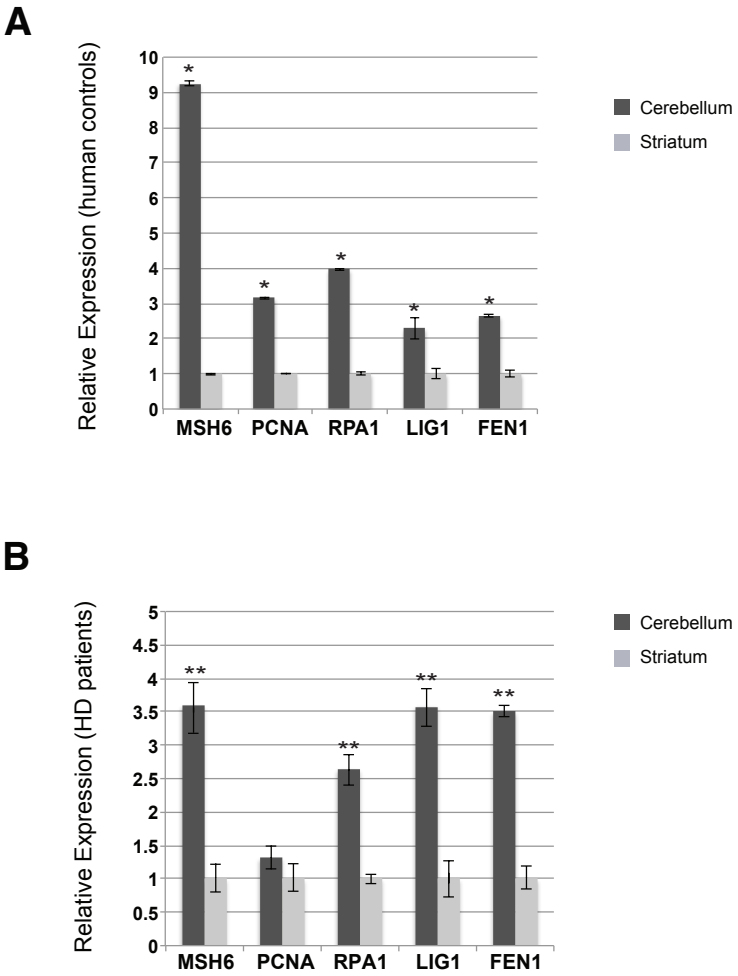


Figure 1-7: DNA replication and repair gene expression is elevated in the cerebellum of human HD patients and unaffected controls.
(A) Real-time RT-PCR analysis of *MSH6, PCNA*, *RPA1, LIG1,* and *FEN1* expression levels in the striatum and cerebellum of unaffected human adults (n = 2; mean $\pm$ s.e.m., 3 independent experiments*;* **P* < 0.05; t-test). b-actin served as the normalization control.
(B) Real-time RT-PCR analysis of *MSH6, PCNA*, *RPA1, LIG1,* and *FEN1* expression levels in the striatum and cerebellum of human HD patients (n = 2; mean $\pm$ s.e.m., 3 independent experiments*;* ***P* < 0.01; t-test). b-actin served as the normalization control.

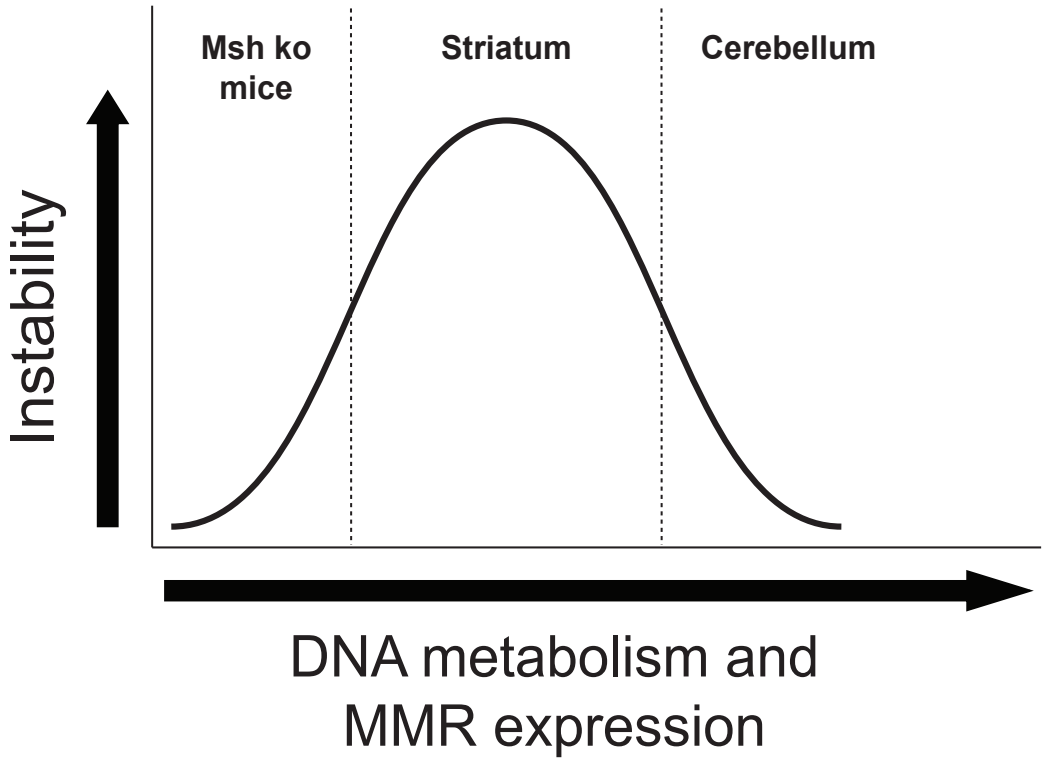


Figure 1-8: Model for biphasic MMR and DNA metabolism pathway regulation of repeat instability.
According to this model, a threshold level of DNA metabolism gene expression and MMR protein expression is required for pronounced repeat instability to occur, which is the situation in the striatum. However, when DNA replication and MMR expression levels are below this threshold, repeat instability does not occur, as is the case for repeat instability mice crossed onto a *Msh* knock-out background. Similarly, when elevated levels of expression for both DNA replication and MMR proteins are achieved, as observed for the cerebellum, repeat instability is also suppressed.

Discussion:

Despite years of intensive study, the molecular basis of repeat instability in TNR disease remains enigmatic – particularly with regard to tissue-selectivity. Investigations into the mechanistic basis of the repeat instability process have spanned a wide range of model organisms, *in vitro* systems, and cell culture approaches. These studies have shown that expanded repeat tracts have the propensity to adopt aberrant structures at the DNA/RNA level, including hairpins, slipped-DNAs, triplexes, and R-loops, and that the predisposition to adopt such altered DNA/RNA conformations is what renders expanded repeats susceptible to high rates of germ line and somatic instability [43-50]. In light of the importance of DNA structures and biology for repeat instability, pathways of DNA metabolism and repair emerged as likely candidates for involvement in the molecular basis of repeat instability. DNA replication and mismatch repair (MMR) are thus viewed as central to the instability process, although the exact nature of how these processes promote repeat instability and whether they act independently or in concert to yield repeat length change remains uncertain [2,34,51-54]. Experiments in both mouse models and human patients have shown that patterns of repeat instability in the different CAG/CTG repeat diseases are constant across specific brain regions, despite the fact that the different diseases exhibit distinct regional vulnerabilities and involve unique gene loci distributed throughout the genome. In this study, we applied microarray analysis to identify gene expression differences between the striatum and the cerebellum to determine potential molecular explanations for these observed repeat instability differences. We identified eight candidate factors, and then verified the differential expression of the leading

candidates. We found that their expression did not change with advanced age or upon polyglutamine-expanded disease protein expression. Our findings provide insight into why repeat instability differences may exist between the striatum, a site of marked somatic mosaicism, and the cerebellum, a brain region with limited repeat length variation.

To delineate the molecular processes underlying repeat instability, we performed a microarray expression analysis comparison of striatum and cerebellum RNAs, and chose to compare aged tissues, since repeat instability increases with time in somatic tissues. Surprisingly, we found that certain DNA replication and DNA repair genes and proteins are more highly expressed in the cerebellum, where somatic mosaicism is limited. This was unexpected, as numerous repeat instability studies in mice have shown that MMR gene expression is required for enhanced somatic repeat expansion [17,35,55-63], [64]. These prior mouse studies demonstrated that contractions arise and expansions are suppressed in the absence of MMR, suggesting that MMR is required to promote CAG/CTG repeat expansion. In light of previous work, we conclude that a baseline level of MMR protein expression is required for dramatic expansion-biased repeat instability to occur, as is the case in the striatum, but when DNA metabolism enzymes and MMR proteins are expressed at much higher levels, as in the cerebellum, pronounced repeat instability does not occur. A model for this biphasic MMR and DNA metabolism pathway regulation of TNR instability is presented in Figure 8. According to this model, high levels of MMR proteins and DNA metabolism factors in the cerebellum favor conditions that stabilize expanded CAG repeats and prevent their expansion, although the mechanisms for reduced repeat instability remain

unclear. One possible explanation is that excessively high expression levels yield impaired enzymatic function, as direct modulation of MSH3 levels in a prior study demonstrated that mismatch repair function is impaired when MSH3 is very highly expressed [39]. Consequently, a threshold level of DNA metabolism and MMR gene expression might promote repeat instability, as it would yield initiation of a robust repair process destined to be unsuccessful, and instead produce CAG repeat expansions.

Another recent investigation into the role of *trans*-acting factors in somatic instability reported that differences in MMR protein expression or DNA metabolism pathways do not account for brain region variations in somatic mosaicism in HD model mice [65]. However, in this work, Western blot analysis of MSH2 revealed a dramatic increase in MSH2 protein levels in the cerebellum in comparison to striatum, though dual comparison of sets of stable and unstable tissues did not support a correlation between *MSH2* levels and instability, when non-CNS tissues were included [65]. Interestingly, of 74 down-regulated genes with Pearson coefficient correlations that were weak to moderate, 63 fell within the DNA metabolism gene class, suggesting that DNA metabolism gene expression patterns may correspond with regional instability differences [65]. It is also noteworthy that MMR gene expression is markedly down-regulated in human embryonic stem cells derived from DM1 parents with expanded CTG repeats, a phenomenon coincident with the loss of spontaneous CTG instability [42]. Similar to MSH2 and MSH3, perturbation of DNA Ligase 1, FEN1, XPA and other DNA repair proteins has been reported to alter CAG instability [66-70], and a recent screen for genes affecting GAA instability in non-dividing yeast cells also

yielded DNA metabolism factors, including *POL30 (Pcna), MCM7*, and *RAD27 (Fen1)* [71], all of which were identified in our microarray expression comparison. Another approach for modeling repeat instability has been to develop induced pluripotent stem cell (iPSC) models, and this strategy has yielded pronounced expansion-biased instability in iPSCs and iPSC derivatives from Friedreich's ataxia (FRDA) patients [72]. Directed studies of MMR using the FRDA iPSC model system found that reduced expression of *MSH2* or *MSH6* could diminish GAA repeat instability [72], which is consistent with the threshold model proposed here (Figure 1-8).

To clarify the nature of MMR protein expression change as a function of age and in the face of polyglutamine neurodegeneration, we measured MMR expression levels in young mice and in the R6/2 HD mouse model. In both cases, we observed increased MMR expression in the cerebellum compared with the striatum, and did not detect altered MMR expression in young mice or in the brains of the HD R6/2 mice. This is consistent with a recent study that examined the role of *trans*-acting factors in regional instability differences in HD knock-in mice [73]. This previous work also evaluated the effect of neurodegeneration in the cerebellum by crossing a HD knock-in mouse model with the Harlequin (*Hq*) model of cerebellum degeneration, and reported no increase in somatic repeat instability in the cerebella of HD knock-in mice carrying the *Hq* mutation [73]. Hence, it appears that the cellular changes that accompany neurodegeneration do not further enhance the pathways that promote repeat instability in somatic tissues in the brain.

Although our study reinforces a role for DNA replication and MMR in regulating repeat instability, we found that differences in MMR protein expression can

not fully account for repeat instability differences resulting from advancing age or for somatic mosaicism differences between certain tissues. Our findings indicate that factors other than just MMR protein expression level must contribute to the complex process of repeat expansion. One possible explanation is that with age or in certain tissues, the ability for alternative structures to form increases. There are many reasons why DNA conformation change could be favored with age or in certain tissue milieus, including increasing oxidative damage to DNA or accumulation of epigenetic alterations at the level of histone modification status in chromatin [74]. Hence, future studies of repeat instability will need to consider a variety of processes influencing DNA metabolism and repair, including especially epigenetic regulatory pathways, as epigenetic processes are emerging as powerful regulators of MMR [75], and repeat instability differences due to parent-of-origin effects and aging likely reflect fundamentally distinct epigenetic processes at work in these varying milieus.

Instead of only relying upon mouse models to examine the molecular basis of TNR instability, we also obtained human cerebellum and striatum RNA samples to determine whether the changes identified in mouse brain also occur in the human brain. We confirmed that in HD patients, as well as in normal human controls, both DNA metabolism genes and MMR genes are more highly expressed in the cerebellum than in the striatum, with one exception – *PCNA*. These findings thus support a role for DNA metabolism and MMR gene expression in the somatic instability differences documented between the striatum and the cerebellum in post-mortem materials obtained from human patients afflicted with CAG/CTG repeat disease [8].

Certainly, one fascinating aspect of the TNR disease field is the uniqueness of these diseases in the human, especially when one considers that the dramatic expansion-biased repeat instability seen in human disease does not occur naturally in related mammalian species or model organisms. While a set of examples of repeat-associated pathology has been reported in plants and in a canine species, neither of these repeats demonstrates the dramatic repeat instability seen in human patients [2,34]. Hence, understanding the biology of repeat instability holds the potential for unlocking one of the perplexing mysteries of this uniquely human disease category, and perhaps may shed light on the evolutionary processes that promoted the relatively rapid rise of the human species from a small founder population more than 150,000 years ago.

Materials and Methods:

Mice :

WT C57BL/6J and CBA x C57BL/6J mice were obtained from the Jackson Laboratories. All animal experiments adhered to National Institute of Health (NIH) guidelines and were approved by the University of California, San Diego Institutional Animal Care and Use Committee (IACUC). HD and WT mice bred on a CBA x C57BL/6 F1 background were taken from an allelic series colony established at the University of Cambridge as previously described [76]. Genotyping and repeat length measurement were performed by Laragen, Los Angeles, USA. All studies were carried out in accordance with the UK Animals (Scientific Procedures) Act 1986. In the present study, we used R6/2 mice carrying a repeat length of either ~256 or ~346 CAG repeats.

Nucleic acid studies: Total RNA was extracted from striatum and cerebellum frozen tissues of 13 and 42 week-old WT and HD (R6/2) mice, using RNeasy mini kit from Qiagen. Reverse transcription was performed with 500 ng of total RNA and SuperScript™ II Reverse Transcriptase (Invitrogen), using hexamer primers. Total RNA was extracted from fresh striatum and cerebellum tissues of 8 and 40 week-old WT C57BL/6 mice, using TRIzol reagent from Invitrogen using their standard protocol. Reverse transcription was done using 1µg of total RNA and MultiScribe™ Reverse Transcriptase (Life Technologies), using random primers. PCNA (Mm_00448100_g1), RPA1 (Mm_01253368_m1), Msh6 (Mm_00487761_m1), Lig1 (Mm_00495331_m1), and Fen1 (Mm_ 01700195_m1) real-time PCR experiments on C57BL/6J WT mice (Taqman (life sciences/ABI)) were performed using a 7500 real

time PCR system (Applied Biosystems). Experimental samples were diluted 50 times to analyse the expression of all genes of interest and β-actin (Mm_01205647_g1) as control. The PCR cycling parameters were 50°C for 2 min, 95°C for 10min, and 40 cycles of 95°C for 15 s, 60°C for 1min.

MMR Real-time PCR experiments on HD and CBA WT (SybR green (Roche) with dissociation curve) were performed using a Mx3005P cycler (Strategen). Specific primers for each tested genes and for 18S mouse endogenous control were designed using NCBI primer-Blast (below). Standard curves were generated for each gene of interest using serial dilutions of testis mouse cDNAs. Experimental samples diluted 20 times to analyze MMR expression and 2000 times to analyze 18S rRNA expression were all run in triplicate. The PCR cycling parameters were 95 °C for 10 min, and 40 cycles of 95 °C for 30 s, 60 °C for 1 min and 72 °C for 1 min. Student's t test to determine exact p-values, was used to determine differences in expression between wild type (WT) and HD transgenic mice (biological triplicate and experimental duplicate for each tested gene). The pattern of MMR expression is reproducible between experiments for each tested gene.

Gene Specific Primers are as follows:

PCR	**Name**	**Sequence**	**bp**
Msh2	Msh2ex14STA	AGCGCTCACTACTGAGGAGACCC	23
	Msh2ex15STB	GCGCACGCTATCACGTGCCTC	21
Msh3	Msh3ex20STA	ATGGCTCAGATTGGCTCCTACG	22
	Msh3ex21STB	TTCCGCTGTGTCCGTCAGTTCTTC	24
Msh6	Msh6ex3STA	AGGCTGCAGCTGGCAGTGTG	20
	Msh6ex4STB	AGGCCCCTGAACACTGGGCT	20
18S	18S3F	CAGTGAAACTGCGAATGG	18
	18S3R	CGGGTTGGTTTTGATCTG	18

Heat map generation:

The heat map of the correlation of the log-transformed intensities was generated using matlab with the following script:

```
>> [x,y,z]=xlsread('C:file');
>> c=corrcoef(x);
>> imagesc(c);
```

Western blot analysis:

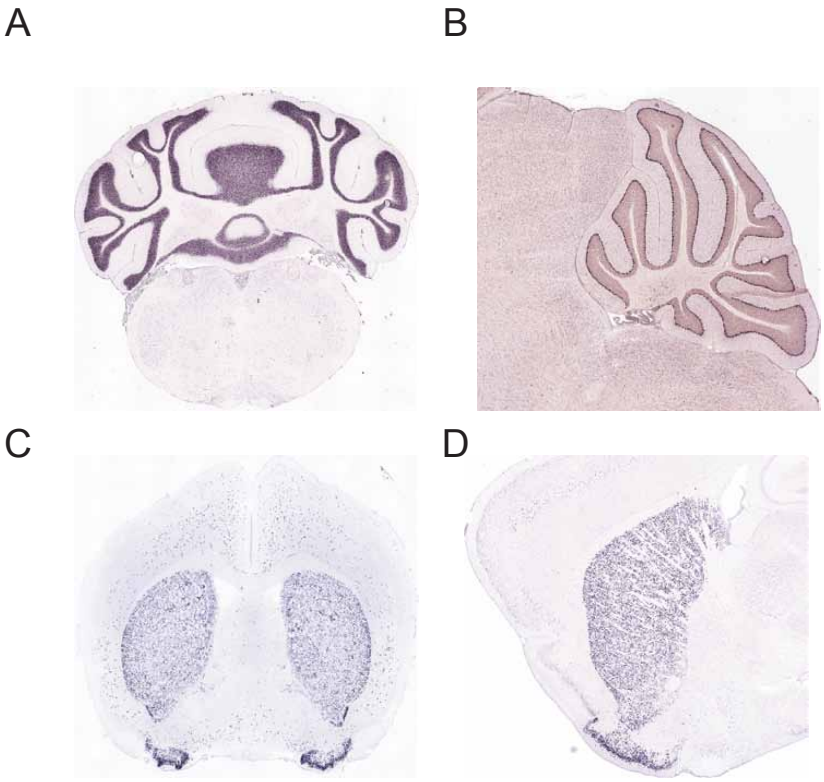
We determined the MSH2, MSH3 and MSH6 protein expression in WT and HD mice (CBAxC57Bl/6) by simultaneously Western blotting of MSH2, MSH3, MSH6 and Actin (as a loading control) [40]. The different tissues were collected from WT C57BL/6 mice killed at 8 or 40weeks or R6/2 and wild type littermate control mice killed at either 13.7 or 42 weeks of age. Proteins were extracted by mechanical homogenisation in lysis buffer (0.125 M Tris-HCl pH 6.8, 4% SDS, 10% glycerol) containing complete Mini 7x protease inhibitor cocktail (Roche). Protein concentration was determined using the Pierce BCA protein assay kit. 40 μg of protein was denatured for 5 min at 95C° resolved by electrophoresis on a 8% polyacrylamide SDS-PAGE gel and electroblotted in transfer buffer (25 mM Tris-HCl pH 8.0, 192 mM glycine, 20% methanol and 0.1% SDS) at 300 mA at 4°C. Membranes were blocked for one hour at room temperature in 5% milk in PBST 1X then incubated overnight at 4°C in simultaneous primary antibodies. The membranes were washed 3 times for 20 min each in PBST, incubated for 1 hr in secondary antibody (GE healthcare, α-mouse-HRP, 1:5000) at room temperature for MSH2, MSH3, MSH6,

PCNA, RPA and for actin, and washed 3 times for 20 min each. Antibody binding was visualized using ECL plus Western blotting detection system (Amersham). MSH2, MSH6, MSH3 and actin were detected using antibodies mouse anti-MSH2 (Calbiochem, 1:200), mouse anti-MSH6 (BD Laboratories, 1:200), MSH3 antibody, clone 2F11 (from Glenn Morris's lab, 1:750) [77], monoclonal mouse anti-PCNA (PC10) (Santa Cruz, cat. # Sc-56, dilution: 1/1000), monoclonal mouse anti-RPA 70 kDa subunit (B-6) (Santa Cruz , cat. no Sc-28304, dilution 1/1000), and Actin (BD Laboratories, 1:5000). Each experiment was reproduced at least three times for each protein tested.

Acknowledgments

This work was supported by the NIH (R01 GM059356 and R01 AG033082 to A.R.L, T32 award GM008666-14 to A.G.M., and P30-HD02274 to the UW Center on Human Development & Disability) and by funding from the Canadian Institutes of Health Research to C.E.P.

Chapter 1, in full, is a reprint of material that has been published in the journal Human Molecular Genetics. Amanda G. Mason, Stephanie Tomé, Jodie Simard, Randell T. Libby, Theodor K. Bammler, Richard P. Beyer, A. Jennifer Morton, Christopher E. Pearson & Albert R. La Spada, 2013. The thesis author was the primary investigator and author of this paper.



Supplementary Figure 1-1: Validation of microarray expression data in mouse brain. *In situ* hybridization staining on 56-day-old male C57BL/6J mice: ALLEN mouse brain collection (http://mouse.brain-map.org) [25]. Characteristic staining of the most highly enriched cerebellar genes from our data set: (A) coronal section Gabra6 (http://mouse.brain-map.org/experiment/show/75551467); (B) sagittal section Pcp2 (http://mouse.brain-map.org/experiment/show/74357592). Characteristic straining of the most highly enriched striatal genes in our data set: (C) coronal section Tac1 (http://mouse.brain-map.org/experiment/show/1038); (D) sagittal section Gpr88 (http://mouse.brain-map.org/experiment/show/520).

Supplementary Table 1-1: Most highly expressed cerebellar genes.
After a significance cut off of <0.002 the following 10 genes had the highest fold enrichment in the cerbellum.

Gene Name	**Symbol**	**Probe ID**	**Fold Change (Cb/Str)**	**P-Value**
Gamma-aminobutyric acid (GABA) A receptor, subunit alpha 6	Gabra6	1417121_at, 1451706_a_at	↑67.07 ↑28.37	2.2E-05 1.0E-05
Cerebellin 1 precursor protein	Cbln1	1423287_at, 1423286_at	↑51.33 ↑17.34	5.6E-05 1.2E-05
Carbonic anhydrase 8	Car8	1427482_a_at	↑44.80	2.4E-06
Neurogenic differentiation 1	Neurod1	1426412_at	↑40.06	1.2E-05
Purkinje cell protein 2 (L7)	Pcp2	1419085_at, 1419084_a_at	↑39.96 ↑27.57	1.5E-07 8.5E-07
G substrate	Gsbs	1449240_at	↑25.15	1.3E-05
Carbonic anhydrase 8	Car8	1424958_at	↑22.69	8.4E-06
Solute carrier family 1 (high affinity aspartate/glutamate transporter), member 6	Slc1a6	1418933_at	↑17.50	1.3E-05
Parvalbumin	Pvalb	1417653_at	↑15.63	5.3E-06
Cerebellin 3 precursor protein	Cbln3	1422911_at	↑13.222	1.1E-06

Supplementary Table 1-2: Most highly expressed striatal genes.
After a significance cut off of <0.002 the following 10 genes had the highest fold enrichment in the striatum.

Gene Name	**Symbol**	**Probe ID**	**Fold Change (Str/Cb)**	**P-value**
Tachykinin 1	Tac1	1416783_at	↑23.81	2.0E-05
B-cell leukemia/lymphoma 11B	Bcl11b	1450339_a_at	↑22.07	2.8E-06
Cyclic AMP-regulated phosphoprotein, 21	Arpp21	1451280_at	↑21.03	7.3E-06
G-protein coupled receptor 88	Gpr88	1423171_at	↑19.80	1.8E-04
Neuropilin (NRP) and tolloid (TLL)-like 1	Neto1	1456283_at	↑16.22	3.6E-05
Meis homeobox 2	Meis2	1417129_a_at	↑15.69	7.9E-06
CaM kinase-like vesicle-associated	Camkv	1423802_at	↑15.40	1.1E-05
Cyclic AMP-regulated phosphoprotein, 21	Arpp21	1424248_at	↑15.33	2.0E-05
Neuronal guanine nucleotide exchange factor	Ngef	1448978_at	↑15.23	4.3E-06
Kv channel-interacting protein 2	Kcnip2	1425870_a_at	↑15.0	1.6E-05

Supplementary Table 1-4: Significantly up-regulated pathways in the Cerebellum

KEGG Pathway Term	**Genes in Pathway**	**P-Value**
Phosphatidylinositol signaling system	12	4.2E-3
Aldosterone-regulated sodium reabsorption	9	7.4E-3
Long-term depression	11	2.2E-2
Axon guidance	16	2.5E-2
DNA replication	7	3.8E-2
Inositol phosphate metabolism	8	5.5E-2
Mismatch repair	5	8.5E-2

Supplementary Table 1-5: Highly associated interacting genes/proteins of MSH6, PCNA, and RPA1 predicted from STRING in our data set

Gene Name	**Symbol**	**Probe ID**	**Fold Change (Cb/Str)**	**P-Value**
Flap structure specific endonuclease 1	Fen1	1421731_a_at	↑1.68	0.0005
Growth arrest and DNA-damage-inducible 45 alpha	Gadd45A	1449519_at	↑1.60	0.0169
Replication factor C (activator 1) 4	Rfc4	1438161_s_at	↑1.55	0.0001
Cyclin-dependent kinase 4	Cdk4	1422441_x_at, 1422440_at, 142239_a_at	↑1.06	0.47 0.12 0.33
Minichromosome maintenance deficient 2 mitotin (S. cerevisiae)	Mcm2	1448777_at	↑1.25	0.02
Minichromosome maintenance deficient 4 homolog (S. cerevisiae)	Mcm4	1436708_x_at, 1416214_at	↑1.165	0.16 0.005
Ligase I, DNA, ATP-dependent	Lig1	1416641_at	↑1.14	0.03
Replication protein A3	Rpa3	1448938_at	↑1.26	0.006
Replication protein A2	Rpa2	1454011_a_at	↑1.12	0.02
Replication factor C (activator 1) 3	Rfc3	1432538_a_at	↑1.19	0.004

References

1. Lopez Castel, A., Cleary, J.D. & Pearson, C.E. Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat Rev Mol Cell Biol* **11**, 165-70 (2010).

2. Pearson, C.E., Nichol Edamura, K. & Cleary, J.D. Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet* **6**, 729-42 (2005).

3. La Spada, A.R., Richards, R.I. & Wieringa, B. Dynamic mutations on the move in Banff. *Nat Genet* **36**, 667-70 (2004).

4. Cleary, J.D. & Pearson, C.E. The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence. *Cytogenet Genome Res* **100**, 25-55 (2003).

5. La Spada, A.R. Trinucleotide repeat instability: genetic features and molecular mechanisms. *Brain Pathol* **7**, 943-63 (1997).

6. Cleary, J.D., Tome, S., Lopez Castel, A., Panigrahi, G.B., Foiry, L., Hagerman, K.A., Sroka, H., Chitayat, D., Gourdon, G. & Pearson, C.E. Tissue- and age-specific DNA replication patterns at the CTG/CAG-expanded human myotonic dystrophy type 1 locus. *Nat Struct Mol Biol* **17**, 1079-87 (2010).

7. Telenius, H., Kremer, B., Goldberg, Y.P., Theilmann, J., Andrew, S.E., Zeisler, J., Adam, S., Greenberg, C., Ives, E.J., Clarke, L.A. & et al. Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat Genet* **6**, 409-14 (1994).

8. Ishiguro, H., Yamada, K., Sawada, H., Nishii, K., Ichino, N., Sawada, M., Kurosawa, Y., Matsushita, N., Kobayashi, K., Goto, J., Hashida, H., Masuda, N., Kanazawa, I. & Nagatsu, T. Age-dependent and tissue-specific CAG repeat instability occurs in mouse knock-in for a mutant Huntington's disease gene. *J Neurosci Res* **65**, 289-97 (2001).

9. Aoki, M., Abe, K., Tobita, M., Kameya, T., Watanabe, M. & Itoyama, Y. Reduction of CAG expansions in cerebellar cortex and spinal cord of DRPLA. *Clin Genet* **50**, 199-201 (1996).

10. Chong, S.S., McCall, A.E., Cota, J., Subramony, S.H., Orr, H.T., Hughes, M.R. & Zoghbi, H.Y. Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* **10**, 344-50 (1995).

11. Hashida, H., Goto, J., Kurisaki, H., Mizusawa, H. & Kanazawa, I. Brain regional differences in the expansion of a CAG repeat in the spinocerebellar ataxias: dentatorubral-pallidoluysian atrophy, Machado-Joseph disease, and spinocerebellar ataxia type 1. *Ann Neurol* **41**, 505-11 (1997).

12. Lopes-Cendes, I., Maciel, P., Kish, S., Gaspar, C., Robitaille, Y., Clark, H.B., Koeppen, A.H., Nance, M., Schut, L., Silveira, I., Coutinho, P., Sequeiros, J. & Rouleau, G.A. Somatic mosaicism in the central nervous system in spinocerebellar ataxia type 1 and Machado-Joseph disease. *Ann Neurol* **40**, 199-206 (1996).

13. Takano, H., Onodera, O., Takahashi, H., Igarashi, S., Yamada, M., Oyake, M., Ikeuchi, T., Koide, R., Tanaka, H., Iwabuchi, K. & Tsuji, S. Somatic mosaicism of expanded CAG repeats in brains of patients with dentatorubral-pallidoluysian atrophy: cellular population-dependent dynamics of mitotic instability. *Am J Hum Genet* **58**, 1212-22 (1996).

14. Ueno, S., Kondoh, K., Kotani, Y., Komure, O., Kuno, S., Kawai, J., Hazama, F. & Sano, A. Somatic mosaicism of CAG repeat in dentatorubral-pallidoluysian atrophy (DRPLA). *Hum Mol Genet* **4**, 663-6 (1995).

15. Libby, R.T., Monckton, D.G., Fu, Y.H., Martinez, R.A., McAbney, J.P., Lau, R., Einum, D.D., Nichol, K., Ware, C.B., Ptacek, L.J., Pearson, C.E. & La Spada, A.R. Genomic context drives SCA7 CAG repeat instability, while expressed SCA7 cDNAs are intergenerationally and somatically stable in transgenic mice. *Hum Mol Genet* **12**, 41-50 (2003).

16. Mangiarini, L., Sathasivam, K., Mahal, A., Mott, R., Seller, M. & Bates, G.P. Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nat Genet* **15**, 197-200 (1997).

17. Manley, K., Shirley, T.L., Flaherty, L. & Messer, A. Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat Genet* **23**, 471-3 (1999).

18. Richards, R.I. Dynamic mutations: a decade of unstable expanded repeats in human genetic disease. *Hum Mol Genet* **10**, 2187-94 (2001).

19. Watase, K., Venken, K.J., Sun, Y., Orr, H.T. & Zoghbi, H.Y. Regional differences of somatic CAG repeat instability do not account for selective neuronal vulnerability in a knock-in mouse model of SCA1. *Hum Mol Genet* **12**, 2789-95 (2003).

20. Kennedy, L. & Shelbourne, P.F. Dramatic mutation instability in HD mouse striatum: does polyglutamine load contribute to cell-specific vulnerability in Huntington's disease? *Hum Mol Genet* **9**, 2539-44 (2000).

21. Cha, J.H. Transcriptional signatures in Huntington's disease. *Prog Neurobiol* **83**, 228-48 (2007).

22. Hodges, A., Strand, A.D., Aragaki, A.K., Kuhn, A., Sengstag, T., Hughes, G., Elliston, L.A., Hartog, C., Goldstein, D.R., Thu, D., Hollingsworth, Z.R., Collin, F., Synek, B., Holmans, P.A., Young, A.B., Wexler, N.S., Delorenzi, M., Kooperberg, C., Augood, S.J., Faull, R.L., Olson, J.M., Jones, L. & Luthi-Carter, R. Regional and cellular gene expression changes in human Huntington's disease brain. *Hum Mol Genet* **15**, 965-77 (2006).

23. Luthi-Carter, R., Strand, A., Peters, N.L., Solano, S.M., Hollingsworth, Z.R., Menon, A.S., Frey, A.S., Spektor, B.S., Penney, E.B., Schilling, G., Ross, C.A., Borchelt, D.R., Tapscott, S.J., Young, A.B., Cha, J.H. & Olson, J.M. Decreased expression of striatal signaling genes in a mouse model of Huntington's disease. *Hum Mol Genet* **9**, 1259-71 (2000).

24. Luthi-Carter, R., Strand, A.D., Hanson, S.A., Kooperberg, C., Schilling, G., La Spada, A.R., Merry, D.E., Young, A.B., Ross, C.A., Borchelt, D.R. & Olson, J.M. Polyglutamine and transcription: gene expression changes shared by DRPLA and Huntington's disease mouse models reveal context-independent effects. *Hum Mol Genet* **11**, 1927-37 (2002).

25. Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., Chen, L., Chen, T.M., Chin, M.C., Chong, J., Crook, B.E., Czaplinska, A., Dang, C.N., Datta, S., Dee, N.R., Desaki, A.L., Desta, T., Diep, E., Dolbeare, T.A., Donelan, M.J., Dong, H.W., Dougherty, J.G., Duncan, B.J., Ebbert, A.J., Eichele, G., Estin, L.K., Faber, C., Facer, B.A., Fields, R., Fischer, S.R., Fliss, T.P., Frensley, C., Gates, S.N., Glattfelder, K.J., Halverson, K.R., Hart, M.R., Hohmann, J.G., Howell, M.P., Jeung, D.P., Johnson, R.A., Karr, P.T., Kawal, R., Kidney, J.M., Knapik, R.H., Kuan, C.L., Lake, J.H., Laramee, A.R., Larsen, K.D., Lau, C., Lemon, T.A., Liang, A.J., Liu, Y., Luong, L.T., Michaels, J., Morgan, J.J., Morgan, R.J., Mortrud, M.T., Mosqueda, N.F., Ng, L.L., Ng, R., Orta, G.J., Overly, C.C., Pak, T.H., Parry, S.E., Pathak, S.D., Pearson, O.C., Puchalski, R.B., Riley, Z.L., Rockett, H.R., Rowland, S.A., Royall, J.J., Ruiz, M.J., Sarno, N.R., Schaffnit, K., Shapovalova, N.V., Sivisay, T., Slaughterbeck, C.R., Smith, S.C., Smith, K.A., Smith, B.I., Sodt, A.J., Stewart, N.N., Stumpf, K.R., Sunkin, S.M., Sutram, M., Tam, A., Teemer, C.D., Thaller, C., Thompson, C.L., Varnam, L.R., Visel, A., Whitlock, R.M., Wohnoutka, P.E., Wolkey, C.K., Wong, V.Y., Wood, M., Yaylaoglu, M.B., Young, R.C., Youngstrom, B.L., Yuan, X.F., Zhang, B., Zwingman, T.A. & Jones, A.R. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168-76 (2007).

26. Bolstad, B.M., Irizarry, R.A., Astrand, M. & Speed, T.P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-93 (2003).

27. Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B. & Speed, T.P. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* **31**, e15 (2003).

28. Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U. & Speed, T.P. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-64 (2003).

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

30. Huang da, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).

31. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).

32. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40**, D109-14 (2012).

33. Liu, G. & Leffak, M. Instability of (CTG)n*(CAG)n trinucleotide repeats and DNA synthesis. *Cell Biosci* **2**, 7.

34. Mirkin, S.M. Expandable DNA repeats and human disease. *Nature* **447**, 932-40 (2007).

35. Slean, M.M., Panigrahi, G.B., Ranum, L.P. & Pearson, C.E. Mutagenic roles of DNA "repair" proteins in antibody diversity and disease-associated trinucleotide repeat instability. *DNA Repair (Amst)* **7**, 1135-54 (2008).

36. Lenzmeier, B.A. & Freudenreich, C.H. Trinucleotide repeat instability: a hairpin curve at the crossroads of replication, recombination, and repair. *Cytogenet Genome Res* **100**, 7-24 (2003).

37. Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., Jensen, L.J. & von Mering, C. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* **39**, D561-8 (2011).

38. Livak, K.J., Flood, S.J., Marmaro, J., Giusti, W. & Deetz, K. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl* **4**, 357-62 (1995).

39. Panigrahi, G.B., Slean, M.M., Simard, J.P., Gileadi, O. & Pearson, C.E. Isolated short CTG/CAG DNA slip-outs are repaired efficiently by hMutSbeta, but clustered slip-outs are poorly repaired. *Proc Natl Acad Sci U S A* **107**, 12593-8 (2010).

40. Seriola, A., Spits, C., Simard, J.P., Hilven, P., Haentjens, P., Pearson, C.E. & Sermon, K. Huntington's and myotonic dystrophy hESCs: down-regulated

trinucleotide repeat instability and mismatch repair machinery expression upon differentiation. *Hum Mol Genet* **20**, 176-85 (2011).

41. Tome, S., Simard, J.P., Slean, M.M., Holt, I., Morris, G.E., Wojciechowicz, K., te Riele, H. & Pearson, C.E. Tissue-specific mismatch repair protein expression: MSH3 is higher than MSH6 in multiple mouse tissues. *DNA Repair (Amst)* **12**, 46-52 (2013).

42. Mortimer, R.K. & Johnston, J.R. Life span of individual yeast cells. *Nature* **183**, 1751-1752 (1959).

43. Ireland, M.J., Reinke, S.S. & Livingston, D.M. The impact of lagging strand replication mutations on the stability of CAG repeat tracts in yeast. *Genetics* **155**, 1657-65 (2000).

44. Hartenstine, M.J., Goodman, M.F. & Petruska, J. Base stacking and even/odd behavior of hairpin loops in DNA triplet repeat slippage and expansion with DNA polymerase. *J Biol Chem* **275**, 18382-90 (2000).

45. Kunkel, T.A. Misalignment-mediated DNA synthesis errors. *Biochemistry* **29**, 8003-11 (1990).

46. Kunkel, T.A. Nucleotide repeats. Slippery DNA and diseases. *Nature* **365**, 207-8 (1993).

47. Schweitzer, J.K. & Livingston, D.M. The effect of DNA replication mutations on CAG tract stability in yeast. *Genetics* **152**, 953-63 (1999).

48. Zhang, X.Y., Varthi, M., Sykes, S.M., Phillips, C., Warzecha, C., Zhu, W., Wyce, A., Thorne, A.W., Berger, S.L. & McMahon, S.B. The putative cancer stem cell marker USP22 is a subunit of the human SAGA complex required for activated transcription and cell-cycle progression. *Mol Cell* **29**, 102-11 (2008).

49. Henry, K.W., Wyce, A., Lo, W.S., Duggan, L.J., Emre, N.C., Kao, C.F., Pillus, L., Shilatifard, A., Osley, M.A. & Berger, S.L. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev* **17**, 2648-63 (2003).

50. Scheel, H., Tomiuk, S. & Hofmann, K. Elucidation of ataxin-3 and ataxin-7 function by integrative bioinformatics. *Hum Mol Genet* **12**, 2845-52 (2003).

51. Gonitel, R., Moffitt, H., Sathasivam, K., Woodman, B., Detloff, P.J., Faull, R.L. & Bates, G.P. DNA instability in postmitotic neurons. *Proc Natl Acad Sci U S A* **105**, 3467-72 (2008).

52. Parniewski, P., Jaworski, A., Wells, R.D. & Bowater, R.P. Length of CTG.CAG repeats determines the influence of mismatch repair on genetic instability. *J Mol Biol* **299**, 865-74 (2000).

53. Wang, G. & Vasquez, K.M. Non-B DNA structure-induced genetic instability. *Mutat Res* **598**, 103-19 (2006).

54. Wang, G. & Vasquez, K.M. Models for chromosomal replication-independent non-B DNA structure-induced genetic instability. *Mol Carcinog* **48**, 286-98 (2009).

55. Dragileva, E., Hendricks, A., Teed, A., Gillis, T., Lopez, E.T., Friedberg, E.C., Kucherlapati, R., Edelmann, W., Lunetta, K.L., MacDonald, M.E. & Wheeler, V.C. Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. *Neurobiol Dis* **33**, 37-47 (2009).

56. Foiry, L., Dong, L., Savouret, C., Hubert, L., te Riele, H., Junien, C. & Gourdon, G. Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. *Hum Genet* **119**, 520-6 (2006).

57. Gomes-Pereira, M., Fortune, M.T., Ingram, L., McAbney, J.P. & Monckton, D.G. Pms2 is a genetic enhancer of trinucleotide CAG.CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Hum Mol Genet* **13**, 1815-25 (2004).

58. Kovtun, I.V., Thornhill, A.R. & McMurray, C.T. Somatic deletion events occur during early embryonic development and modify the extent of CAG expansion in subsequent generations. *Hum Mol Genet* **13**, 3057-68 (2004).

59. Savouret, C., Brisson, E., Essers, J., Kanaar, R., Pastink, A., te Riele, H., Junien, C. & Gourdon, G. CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J* **22**, 2264-73 (2003).

60. Savouret, C., Garcia-Cordier, C., Megret, J., te Riele, H., Junien, C. & Gourdon, G. MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice. *Mol Cell Biol* **24**, 629-37 (2004).

61. Tome, S., Holt, I., Edelmann, W., Morris, G.E., Munnich, A., Pearson, C.E. & Gourdon, G. MSH2 ATPase domain mutation affects CTG*CAG repeat instability in transgenic mice. *PLoS Genet* **5**, e1000482 (2009).

62. van den Broek, W.J., Nelen, M.R., Wansink, D.G., Coerwinkel, M.M., te Riele, H., Groenen, P.J. & Wieringa, B. Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum Mol Genet* **11**, 191-8 (2002).

63. Wheeler, V.C., Lebel, L.A., Vrbanac, V., Teed, A., te Riele, H. & MacDonald, M.E. Mismatch repair gene Msh2 modifies the timing of early disease in Hdh(Q111) striatum. *Hum Mol Genet* **12**, 273-81 (2003).

64. Kovalenko, M., Dragileva, E., St Claire, J., Gillis, T., Guide, J.R., New, J., Dong, H., Kucherlapati, R., Kucherlapati, M.H., Ehrlich, M.E., Lee, J.M. & Wheeler, V.C. Msh2 Acts in Medium-Spiny Striatal Neurons as an Enhancer of CAG Instability and Mutant Huntingtin Phenotypes in Huntington's Disease Knock-In Mice. *PLoS One* **7**, e44273 (2012).

65. Lee, J.M., Zhang, J., Su, A.I., Walker, J.R., Wiltshire, T., Kang, K., Dragileva, E., Gillis, T., Lopez, E.T., Boily, M.J., Cyr, M., Kohane, I., Gusella, J.F., MacDonald, M.E. & Wheeler, V.C. A novel approach to investigate tissue-specific trinucleotide repeat instability. *BMC Syst Biol* **4**, 29 (2010).

66. Goula, A.V., Berquist, B.R., Wilson, D.M., 3rd, Wheeler, V.C., Trottier, Y. & Merienne, K. Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice. *PLoS Genet* **5**, e1000749 (2009).

67. Goula, A.V., Pearson, C.E., Della Maria, J., Trottier, Y., Tomkinson, A.E., Wilson, D.M., 3rd & Merienne, K. The nucleotide sequence, DNA damage location, and protein stoichiometry influence the base excision repair outcome at CAG/CTG repeats. *Biochemistry* **51**, 3919-32 (2012).

68. Hubert, L., Jr., Lin, Y., Dion, V. & Wilson, J.H. Xpa deficiency reduces CAG trinucleotide repeat instability in neuronal tissues in a mouse model of SCA1. *Hum Mol Genet* **20**, 4822-30 (2011).

69. Lopez Castel, A., Tomkinson, A.E. & Pearson, C.E. CTG/CAG repeat instability is modulated by the levels of human DNA ligase I and its interaction with proliferating cell nuclear antigen: a distinction between replication and slipped-DNA repair. *J Biol Chem* **284**, 26631-45 (2009).

70. Yang, J. & Freudenreich, C.H. Haploinsufficiency of yeast FEN1 causes instability of expanded CAG/CTG tracts in a length-dependent manner. *Gene* **393**, 110-5 (2007).

71. Zhang, Y., Shishkin, A.A., Nishida, Y., Marcinkowski-Desmond, D., Saini, N., Volkov, K.V., Mirkin, S.M. & Lobachev, K.S. Genome-wide Screen Identifies Pathways that Govern GAA/TTC Repeat Fragility and Expansions in Dividing and Nondividing Yeast Cells. *Mol Cell* **48**, 254-65 (2012).

72. Du, J., Campau, E., Soragni, E., Ku, S., Puckett, J.W., Dervan, P.B. & Gottesfeld, J.M. Role of mismatch repair enzymes in GAA-TTC triplet-repeat expansion in Friedreich's ataxia induced pluripotent stem cells (iPSCs). *J Biol Chem* (2012).

73. Lee, J.M., Pinto, R.M., Gillis, T., St Claire, J.C. & Wheeler, V.C. Quantification of age-dependent somatic CAG repeat instability in Hdh CAG knock-in mice reveals different expansion dynamics in striatum and liver. *PLoS One* **6**, e23647 (2011).

74. Debacker, K., Frizzell, A., Gleeson, O., Kirkham-McCarthy, L., Mertz, T. & Lahue, R.S. Histone deacetylase complexes promote trinucleotide repeat expansions. *PLoS Biol* **10**, e1001257 (2012).

75. Kaidi, A. & Jackson, S.P. KAT5 tyrosine phosphorylation couples chromatin sensing to ATM signalling. *Nature* **498**, 70-4 (2013).

76. Duzdevich, D., Li, J., Whang, J., Takahashi, H., Takeyasu, K., Dryden, D.T., Morton, A.J. & Edwardson, J.M. Unusual structures are present in DNA fragments containing super-long Huntingtin CAG repeats. *PLoS One* **6**, e17119.

77. Holt, I., Thanh Lam, L., Tome, S., Wansink, D.G., Te Riele, H., Gourdon, G. & Morris, G.E. The mouse mismatch repair protein, MSH3, is a nucleoplasmic protein that aggregates into denser nuclear bodies under conditions of stress. *J Cell Biochem* **112**, 1612-21 (2011).

Chapter 2: The molecular basis of lifespan extension in *sgf73Δ* yeast

Introduction:

The budding yeast *Saccharomyces cerevisiae* has been and continues to be an instrumental tool in studying mammalian aging. By studying replicative life span (RLS) of yeast cells, defined by the number of daughter cells that one mother cell can produce though budding [1], we have garnered a better understanding of the mammalian aging process. One of the first major aging related findings from yeast was the involvement of *SIR2* in aging modulation. *SIR2* is a NAD+ dependent deacetylase and has been implicated in both lifespan regulation and caloric restriction in a variety of species, making it a central focus of aging research. The overexpression of Sir2, by integration of a second copy under the endogenous promoter has been shown to extend RLS by ~30% [2]. Sir2 is also required for silencing at telomeres, and represses rDNA recombination, rDNA recombination causes the accumulation of rDNA circles which is a cause of aging [3]. The overexpression of Sir2 in *Caenorhabditis elegans* [5], and *Drosophila melanogaster* [2,4] also results in life extension. In addition, the overexpression of *SIRT1*, the closest mammalian ortholog of *SIR2,* has varying effects on cellular senescence depending on the cell type, further implicating the important role of *SIR2* and its orthologs in aging [5,6]. Studies in mice have also demonstrated the importance of SIRT1 in aging, with one such study showing that overexpression of SIRT1 in the brain protects against aging [7].

Spinocerebellar ataxia type 7 (SCA7) is an autosomal dominant genetically inherited neurodegenerative disorder in which the causative mutation is a CAG/polyglutamine repeat expansion in the ataxin-7 gene[8]. There are currently nine known inherited disorders caused by a CAG/poly Q repeat track expansion at different loci: dentatorubral-pallidoluysian atrophy (DRPLA), Huntington's disease (HD), spinobulbar muscular atrophy (SBMA), and six forms of spinocerebellar ataxias (SCA1, 2, 3, 6, 7, and 17). The polyglutamine (polyQ) tract in ataxin-7 ranges in size from 4 - 35 glutamines in unaffected individuals, but is expanded to 37 or more glutamines in SCA7 affected patients [8,9]. In affected individuals the polyQ track is translated, resulting in a protein that is prone to aggregation. Ataxin-7 protein accumulation leads to the dysfunction and death of neurons in the retina, cerebellum, and cerebellar associated structures. This neuron loss results in blindness, progressive loss of coordination, and eventually premature death [10].

Ataxin-7 is a highly conserved member of the Spt3-Taf9-Gcn5 acetyltransferase (STAGA) complex, a major transcriptional coactivator complex in mammalian cells [11,12]. The STAGA complex harbors both histone acetyltransferase activity mediated by GCN5 and histone deubiquitination activity through USP22[13,14,15]. *SGF73* was identified as the *Saccharomyces cerevisiae* ortholog of ataxin-7 via sequence alignment[16,17]; the proteins share a region known as the SCA7 domain and an amino-terminal zinc finger (ZnF) domain[11,18] (Figure 2-1 a, b). Sgf73 is a member of the STAGA complex yeast equivalent, the Spt-Ada-Gcn5 acetyltransferase (SAGA) complex [11,19]. The SAGA complex harbors the same histone

acetyltransferase activity through Gcn5 and histone deubiquitination by Ubp8 (the yeast USP22 ortholog)[20] (reviewed [21]).

Further characterization identified ataxin-7 is a member of the USP22 histone deubiquitinase module. This is conserved in yeast as Sgf73 is a member of the Ubp8 histone deubiquitinase module (DUBm)[15,22,23]. Sgf73 has been shown to link the SAGA core components to the Ubp8 histone DUBm [15,22]. Sgf73 contains two highly conserved zinc-finger binding domains, one in the amino-terminal domain that interacts with the Ubp8 module, and a separate domain (aa 227-272) within the SCA7 box that binds nucleosomes and is within the region involved in SAGA binding [11,18,22] (Figure 2-1b). In addition, it has been demonstrated that Sgf73 is required for Ubp8 mediated histone H2B deubiquitination [22]. Deletion of *SGF73* causes a global increase in ubiquitinated histone H2B K123, similar to *UBP8* deletion [15,22,23].

Transcription of genes is dependent on chromatin structure, which is affected by multiple histone modifications. The SAGA complex affects multiple histone modification processes including H3 and H4 acetylation by Gcn5 [24,25], and histone H2B K123 deubiquitination by Ubp8 [20]. These histone modifications subsequently affect other histone modifications. H3K4 and H3K36 methylation are associated with an active transcription state and this methylation event is dependent on H2B ubiquitination [26-28]. Additionally the tri-methylation of H3K4, also associated with active transcription, is influenced by Gcn5 acetylation [29]. It is through coordination of these histone marks that SAGA is able to regulate the transcription of multiple genes after the complex is recruited to a genomic site.

In yeast, a second histone acetyltransferase complex, SAGA-Like (SLIK), is involved in transcriptional regulation and the retrograde response pathway and shares many of the same components of the SAGA complex [30,31]. Differences exist between the complexes as SLIK contains the protein Rtg2 and a cleaved form of protein Spt7, while lacking the protein Spt8 present in SAGA [30,31]. Spt7 is cleaved from its full length SAGA-associated form by the enzyme Pep4 into a c-terminal truncated SLIK-associated form that no longer contains an Spt8 binding domain [32]. Though utilizing these complex differences some regulatory processes in the cell can be defined as SAGA or SLIK dependent as the complexes have different regulatory roles [30,31,33].

In this study we examined the yeast strain *sgf73Δ* as it was identified to have drastic RLS extension. We found that the *sgf73Δ* RLS extension is SAGA and not SLIK dependent. The deletion of other components of the SAGA DUBm also resulted in RLS extension, but not to the degree seen in *sgf73Δ,* and the deletion of other non-DUBm SAGA components did not extend RLS. Additionally we found that the *sgf73Δ* RLS extension is Sir2 dependent and that Sgf73 and Sir2 physically interact.

Results:

Deletion of SAGA histone DUBm components results in extended RLS

Through the systematic screening of all non-essential single deletion mutants to explore the effects they have on yeast RLS, we discovered that the deletion of *SGF73* resulted in one of the longest single deletion RLS extensions seen to date. The *sgf73Δ* strain has a mean RLS, average number of daughter cells produced before senescence, of 40.7 in comparison to the WT mean RLS of 26.1, a 56% extension

(McCormick, Mason, & Guyenet et al submitted). Sgf73 is a member of the SAGA and SLIK transcriptional co-activating chromatin modifying complexes. The SAGA and SLIK complexes have both acetyltransferase activity conferred through Gcn5, and histone deubiquitinase activity though Ubp8. Sgf73 is a member of the four-protein DUBm comprised of Sgf73, Ubp8, Sgf11, and Sus1 where Sgf73 serves as the linking factor connecting the DUBm to the rest of SAGA/SLIK[23]. When *SGF73* is deleted it causes the SAGA/SLIK Ubp8 DUBm to be non-functional, and levels of ubiquitinated H2BK123 are elevated[22]. We therefore determined the RLS of the other three DUBm components, finding that the deletion of *SGF11* and *UBP8* increased RLS to a mean of 36.6 and 34.4 respectively, while strains lacking *SUS1* were not long-lived and had a mean RLS of 25.4 similar to the WT mean RLS of 24.9 (McCormick, Mason, & Guyenet et al submitted). *SUS1* may have other functional roles outside of involvement with SAGA/SLIK that results in the deletion not extending RLS. We also found that the double deletion strain *sgf73Δ ubp8Δ* had an RLS of 41.9 almost identical to the *sgf73Δ* RLS of 42.5 in this experiment (McCormick, Mason, & Guyenet et al submitted). This finding insinuates that both deletions increase RLS by a similar mechanism as the deletion of *UBP8* is not able to further extend *sgf73Δ* RLS. One explanation is that both deletions cause increased levels of ubiquitinated H2B [22], and thus may be a cause of RLS extension.

Deletion of non-essential non-DUBm SAGA/SLIK components are not long lived

To see if RLS extension is specific to deletion of DUBm components we explored the RLS of other non-essential SAGA/SLIK components. We found that

strains lacking the SAGA/SLIK components *HIF1*, *SPT7*, *ADA2*, and *SPT20* have reduced lifespan in comparison to WT with an RLS of 4.0, 4.3, 10.9 and 11.3 respectively. Strains lacking *NGG1*, *SPT3*, *RTG2*, *SGF29*, and *GCN5* have lifespans similar to WT with an RLS of 24.2, 24.9, 25.4, 25.4, 25.6, and 25.9 respectively (McCormick, Mason, & Guyenet et al submitted). *SPT8,* a SAGA specific protein influencing TBP-TATA interactions at promoters [34], and *CHD1* a chromodomain protein involved in maintaining chromatin structure during transcription by preventing histone exchange [35], have a slightly extended lifespans in comparison to WT with an RLS of 30.2 and 31.1 respectively. This RLS extension is not as dramatic as the RLS extension seen in DUBm mutants (McCormick, Mason, & Guyenet et al submitted). From these results we deduce that RLS extension in the DUBm mutants is a result of loss of DUBm function and is not a result of decreased core SAGA/SLIK function. However, RLS extension may also be a result of the uncoupling of SAGA/SLIK acetyltransferase activity from SAGA/SLIK deubiquitinase activity.

RLS extension is mediated by H2BK123 mono-ubiquitination

Since we found that RLS extension is associated with alterations to SAGA DUBm components we explored the primary target of Ubp8 deubiquitination, histone H2B K123. The E3 ubiquitin ligase Bre1, and the E2 ubiquitin-conjugating enzyme Rad6, form a heterodimeric complex that mono-ubiquitinates H2B K123 (reviewed [36]). Strains lacking *BRE1* have an RLS of 17.6 while strains lacking *RAD6* have an RLS of 16.0. thus the inability to mono-ubiquitinate H2B K123 causes a decreased RLS (McCormick, Mason, & Guyenet et al submitted). This result further supports

our hypothesis that RLS extension in *sgf73Δ* and *ubp8Δ* is mediated by H2B K123 mono-ubiquitination levels, with higher levels of ubiquitinated H2B K123 extending RLS. Rad6 also acts as the E2 ubiquitin ligase for other E3s in addition to Bre1, and therefore it's possible that reduced RLS in *rad6Δ* may have other contributing factors aside from lack of H2B K123 mono-ubiquitination. Ubp10 is another deubiquitinase in yeast that is capable of deubiquitinating histone H2B K123, but has been reported to do so at different genomic sites than Ubp8 [37]. To determine if global H2B K123 mono-ubiquitination or rather Ubp8 specific sites were the cause of RLS extension in *sgf73Δ* and *ubp8Δ,* we determined the RLS of *ubp10Δ*. The *ubp10Δ* strain exhibited a reduced RLS of 17.2 (McCormick, Mason, & Guyenet et al submitted), as opposed to the extended RLS seen in *sgf73Δ* and *ubp8Δ.* This suggests that increased H2B K123 mono-ubiquitination specifically at Ubp8 regulated sites influences RLS extension.

RLS extension in *sgf73Δand ubp8D* is SAGA-dependent but not SLIK-dependent

To determine the link between SAGA function and longevity, epistasis analysis was performed with non-DUBm components of the SAGA complex to determine if they are required for RLS extension associated with *sgf73Δ* and *ubp8Δ.* The SAGA components *GCN5* and *SAC3* are required for *sgf73Δ* RLS extension as the double mutants no longer have an extended RLS. *Sgf73Δgcn5Δ* has an RLS of 14.6 and *sgf73Δsac3Δ* has an RLS of 9.2 (McCormick, Mason, & Guyenet et al submitted). These results were specific for SAGA factors as the deletion of the nuclear transcription factors Gcn4 and Hap4 did not affect *sgf73Δ* RLS. The RLS of *sgf73Δ gcn4Δ* is 39.9 and *sgf73Δ hap4Δ* has an RLS of 40.6 (McCormick, Mason, & Guyenet

et al submitted). From these results we find SAGA must be otherwise intact for RLS extension in *sgf73Δ* and *ubp8Δ,* as loss of DUBm components shortens the RLS of strains lacking other non-DUBm SAGA components, rather than extending it.

Both SAGA and SLIK contain Gcn5 and Sac3. In order to distinguish between the two complexes, we specifically disrupted each complex in conjunction with deletion of *SGF73* or *UBP8*. Spt8 is a component of SAGA but not SLIK. We found that Spt8 is required for RLS extension in *sgf73Δ* or *ubp8Δ,* as *sgf73Δ spt8Δ* has a reduced RLS of 14.6 and *ubp8Δ spt8Δ* has a reduced RLS of 23.3 (McCormick, Mason, & Guyenet et al submitted). To explore the SLIK complex, generated *PEP4* deletion strains were generated. PEP4 is an endopeptidase responsible for cleaving full length SAGA associated Spt7 into a truncated SLIK associated form [32]. We found that both *sgf73Δ pep4Δ* and *ubp8Δ pep4Δ* strains retained a long lifespan with an RLS of 34.8 and 27.9 respectively (McCormick, Mason, & Guyenet et al submitted). These findings suggest that RLS extension by deletion of DUBm components requires an otherwise intact SAGA but not SLIK complex. To further confirm these results we examined the effects of deleting *RTG2*, another SLIK-specific component linked to the retrograde response. As with the deletion of *PEP4*, *sgf73Δ rtg2Δ* has an extended RLS of 36.2 (McCormick, Mason, & Guyenet et al submitted). As previously reported, *rtg2Δ* alone was slightly short-lived with an RLS of 22.6 [38]. We also explored two other components of retrograde signaling, *RTG1* and *RTG3*, and found that they are not required for lifespan extension by *sgf73Δ*. The *sgf73Δ rtg1Δ* mutant has an RLS of 34.0 and *sgf73Δ rtg3Δ* has an RLS of 31.9 (McCormick, Mason, & Guyenet et al submitted). These findings further implicate that it is the integrity of the

SAGA complex but not the SLIK complex that is required for lifespan extension by loss of DUBm components.

sgf73Δ and *ubp8D* RLS extension is *SIR2* and/or *FOB1* dependent

To further unravel the mechanism of RLS extension in *sgf73Δ* and *ubp8D,* we performed a set of RLS epistasis experiments with the known lifespan modulating factor Sir2. USP22, the mammalian ortholog of *UBP8*, has been reported to interact with SIRT1, the mammalian ortholog of *SIR2,* and promote SIRT1-dependent deacetylation of STAGA components [39-41]. This provides a basis to believe a connection exists between Sir2 function and SAGA activity. A caveat to these experiments is that the *sir2Δ* strain is short lived and modulations of other known RLS extending factors have no effect on this genetic background [42,43]. Indeed we find that the RLS of *sgf73Δ sir2Δ* is 11.3, and the RLS of *ubp8Δ sir2Δ* is 14.3, making these strains short lived like *sir2Δ* (McCormick, Mason, & Guyenet et al submitted). One of the reasons the *sir2Δ* strain is short lived is due to the accumulation of extra ribosomal circles (ERCs), however the deletion of *FOB1* suppresses the accumulation of ERCs giving the *sir2Δ fob1Δ* strain an RLS similar to WT [2,44]. Known RLS extension mutations that were ineffective in *sir2Δ* strains, have been shown to extend RLS in the *sir2Δ fob1Δ* strain background [42,43]. Therefore we explored the possibility that on the *sir2Δ fob1Δ* background, *SGF73* or *UBP8* deletion might cause RLS extension. Our experiments showed that this is not the case. The RLS of *sgf73Δ sir2Δ fob1Δ* is 22.9 compared to 23.6 in *sir2Δ fob1Δ*, and the RLS of *ubp8Δ sir2Δ fob1Δ* is 28.0 compared to 27.1 in *sir2Δ fob1Δ* (McCormick, Mason, & Guyenet et al submitted). These

findings suggest that *sgf73Δ* and *ubp8Δ* RLS extension is *SIR2* and/or *FOB1* dependent. They further suggest that upon *SGF73* or *UBP8* deletion, Sir2 function may be enhanced as in Sir2 overexpression (Sir2OE) strains. It is also possible that *FOB1* function may be inhibited or it could be a combination of both *SIR2* and *FOB1* function that partially contributes to *sgf73Δ* and *ubp8Δ* RLS extension.

Deletion of *SGF73* causes altered Sir2 dependent telomere silencing and rDNA recombination

To further delineate the connection between Sgf73 and Sir2 we chose to explore Sir2-dependent properties associated with longevity in the *sgf73Δ* yeast. The increased expression of Sir2 is known to increase RLS[2], thus a simple explanation would be that the *sgf73Δ* yeast express a higher lever of Sir2. However this is not the case as Sir2 levels in the *sgf73Δ* yeast are unchanged from those in WT yeast (McCormick, Mason, & Guyenet et al submitted).

There are two regions in the genome, rDNA and telomeres, where Sir2 function is known to influence longevity [2,45]. The rDNA locus is a region of consisting of 100-200 nine-kb repeats where recombination is known to take place and create ERCs, which have an ARS (autonomously replicating sequence**)** allowing them to replicate. Although ERCs can replicate they do not have a centromere so are not passed into daughter cells during budding. Their accumulation in the mother cell promotes aging [46]. Sir2 acts to repress rDNA recombination thus limiting the production of ERCs and promoting longevity [2]. Additionally, Sir2 is responsible for maintaining silencing at telomeres, and silences mating type loci by removing H4K16

acetylation and recruiting other silencing proteins (reviewed [47]). Previous studies have shown that Sir2 contributes to RLS extension partially through regulation of histones near telomeres resulting in subtelomeric gene silencing [45]. Interestingly, the loss of the SAGA component *ADA2* causes reduced rDNA and telomere silencing [48], as is seen in *sir2Δ* yeast. To test whether *sgf73Δ* yeast have reduced rDNA recombination associated with increased Sir2 function, we have performed an rDNA recombination assay. In this assay, the *ADE2* locus is inserted in one rDNA repeat, if the yeast cells have the *ADE2* locus they will appear red in color, but if the *ADE2* locus is lost due to recombination into an ERC, yeast cells appear white. rDNA recombination is measured by the frequency at which *ADE2* is lost from the rDNA array. This appears as red/white half-sectored colonies, as daughter cells do not receive the *ADE2* locus when it is contained on an ERC. In the *sgf73Δ* yeast, rDNA recombination was drastically reduced compared to WT, a phenotype consistent with increased Sir2 function (McCormick, Mason, & Guyenet et al submitted). Additional assays wherein *URA3* was integrated at telomeric loci were carried out to test for telomeric silencing levels in *sgf73Δ* yeast. If *URA3* is silenced, strains will not grow on media lacking uracil but will grow on media containing 5-FOA where *URA3* expression is toxic. Conversely, if *URA3* is not silenced *URA3* will be expressed allowing strains to grow on media lacking uracil and the strains will not grow on media containing 5-FOA as the *URA3* expression will be toxic. Consistent with our rDNA silencing findings, silencing of telomeric *URA3* is increased in *sgf73Δ* yeast (McCormick, Mason, & Guyenet et al submitted). These findings are consistent with increased Sir2 function in the *sgf73Δ* yeast. One explanation is that Sgf73 directs Sir2 to a number of targets and

when *SGF73* is deleted, Sir2 is no longer directed to this set of targets. This frees up the available cellular Sir2 to act on other targets, resulting in increased rDNA and telomeric silencing.

Sgf73 interacts with Sir2

Based upon our findings of altered Sir2-dependent properties in the *sgf73Δ* yeast as well as the dependence on Sir2 for *sgf73Δ* RLS extension, we chose to explore additional potential connections between Sgf73 and Sir2. Evidence exists in mammalian cells for an interaction between the ortholog of *SIR2*, SIRT1, and the ortholog of *UBP8*, USP22 [40,41]. This suggests that Sir2 may interact with the Ubp8 DUBm components in yeast, and that the functional interactions between deacetylase and deubiquitinase components may be conserved. We tested whether Sir2 interacts with Sgf73 in yeast using two different assays. First, we examined the interaction *in vitro* by mixing yeast lysates expressing a GFP-tagged Sgf73 with purified recombinant GST-tagged Sir2 protein. Immunoprecipitation of Sir2 by glutathione sepharose was sufficient to recover GFP-Sgf73, as shown by immunoblot analysis of the precipitated material (Figure 2-2 a). An enzymatically dead Sir2-R139K mutant also retained the capacity to interact with Sgf73-GFP, indicating that the deacetylase activity of Sir2 is not required for interaction. In a second experiment, we detected an *in vivo* interaction between Sgf73-GFP and endogenous Sir2 by co-immunoprecipitation of Sgf73-GFP (Figure 2-2 b). The *in vivo* interaction was also seen between Sir2-GFP and Sgf73-13Myc by co-immunoprecipitation of Sir2-GFP (Figure 2-2 c). Together, these experiments indicate an interaction between Sir2 and

the DUBm of SAGA in yeast, and provide evidence for the highly conserved nature of this interaction, since it has also been detected in mammalian cells.

Expression of human 10Q and 113Q ataxin-7 in *sgf73Δ* does not rescue RLS

As ataxin-7 is the human ortholog of Sgf73 we decided to explore whether the human protein could function in place of the yeast protein to rescue the *sgf73Δ* RLS phenotype. A full length cDNA for both normal ataxin-7 with a CAG repeat length of 10 (SCA7Q10) as well as expanded ataxin-7 with a CAG repeat of 113 (SCA7Q113) were integrated into WT and *sgf73Δ* yeast (Figure 2-3 a). Integrated ataxin-7 is listed as SCA7Q10 or SCA7Q113 in all figures. We also ensured that the human ataxin-7 protein was expressed in the yeast and therefore would be present to substitute for Sgf73 (Figure 2-3 b). Despite adequate ataxin-7 protein expression, there is no rescue of the *sgf73Δ* RLS phenotype. Both *sgf73Δ* SCA7Q10 and *sgf73Δ* SCA7Q113 yeast have an extended RLS (Figure 2-4). Surprisingly, their RLS was even slightly longer than the *sgf73Δ*. In an unexpected finding, we found that the expression of either SCA7Q10 or SCA7Q113 in the WT background resulted in life extension (Figure 2-4). This suggests that while the human ataxin-7 protein is not able to substitute for yeast Sgf73 it may still disrupt SAGA complex formation by blocking correct Sgf73 SAGA DUBm interactions or other Sgf73 endogenous functions. We also tested the growth of the ataxin-7 integrated strains on 50ng/mL cycloheximide, which causes slow growth in the *sgf73Δ* yeast [49]. In concert with the RLS data, WT strains containing either SCA7Q10 or SCA7Q113 showed a phenotype very similar to the *sgf73Δ* yeast with delayed growth on cycloheximide (Figure 2-5 a). We then chose to

exacerbate this phenotype by growing strains on higher concentrations of cycloheximide, 75ng/mL and 100ng/mL. At high concentrations, the ataxin-7 integrated strains appear to have an intermediate level of sensitivity between the WT and *sgf73Δ* yeast (Figure 2-5 b). This data supports the finding that while being longer lived than WT the RLS in the WT ataxin-7 integrant strains is not as long as in *sgf73Δ* yeast. The *sgf73Δ* yeast integrated with either SCA7Q10 or SCA7Q113 still show sensitivity to cycloheximide as is expected from the inability of the human ataxin-7 expression to rescue the *sgf73Δ* RLS phenotype (Figure 2-5 c). In fact the ataxin-7 integrated strains have a higher sensitivity to cycloheximide than *sgf73Δ*.

Sir2 over expression and deletion of *UBP8* recapitulate the RLS extension in *sgf73Δ* yeast

Together our findings suggest that *sgf73Δ* yeast have extended RLS that is dependent on Sir2 and/or Fob1 and also on elevated levels of mono-ubiquitinated H2B K123 at Ubp8-specific sites. This study has shown that while *ubp8Δ* yeast have extended RLS it is not to the same extent as in *sgf73Δ* yeast despite both deletions disrupting SAGA DUB activity. Additionally, we find that there is a dependence on Sir2 for *sgf73Δ* RLS extension, and *sgf73Δ* strains have alterations in Sir2-dependent activities. Despite these finding the RLS extension in Sir2OE yeast is much lower than in either *sgf73Δ* or *ubp8Δ* yeast. This suggests that the extremely long RLS extension seen in *sgf73Δ* yeast may be a result of both altered Sir2 function and ubiquitinated H2B K123 levels. Indeed when we overexpress Sir2 in the *ubp8Δ* background the RLS is nearly identical to the *sgf73Δ* strain, with a mean RLS of 39.3 compared to

40.7 (McCormick, Mason, & Guyenet et al submitted). These results support a model in which *SGF73* is the main coordinator between the activity of two chromatin remodeling complexes, SAGA DUB activity and Sir2 deacetylase activity, thereby resulting in extended lifespan.

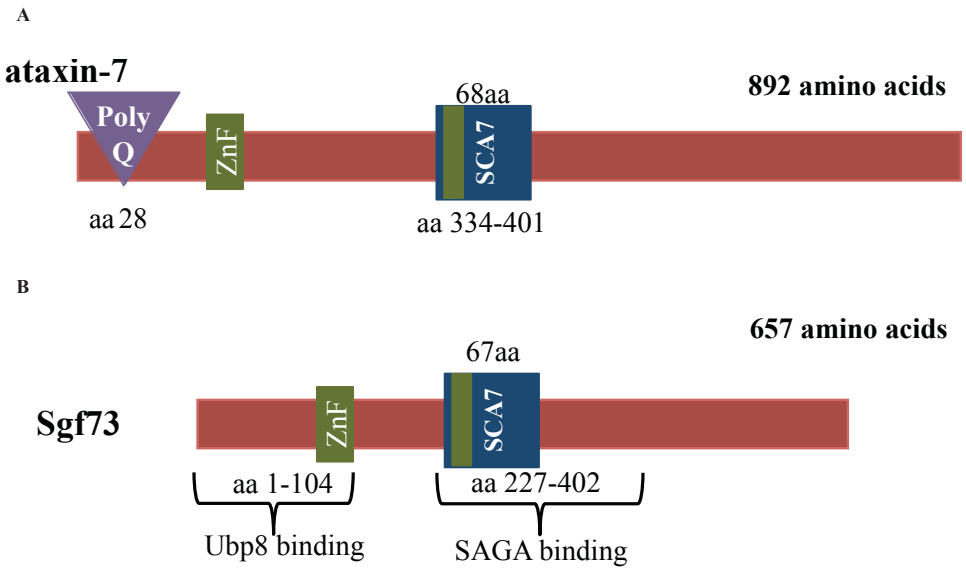


Figure 2-1: Domains of ataxin-7 and Sgf73
A) Schematic of ataxin-7 depicting the N-terminal polyQ repeat region, N-terminal zinc-finger binding domain, and the conserved SCA7 box domain containing the second zinc-finger binding domain. B) Schematic of Sgf73 depicting the N-terminal zinc-finger binding domain, the conserved SCA7 box domain containing the second zinc-finger binding domain, as well as outlining the regions known to be essential for Sgf73 binding to Ubp8 and SAGA.

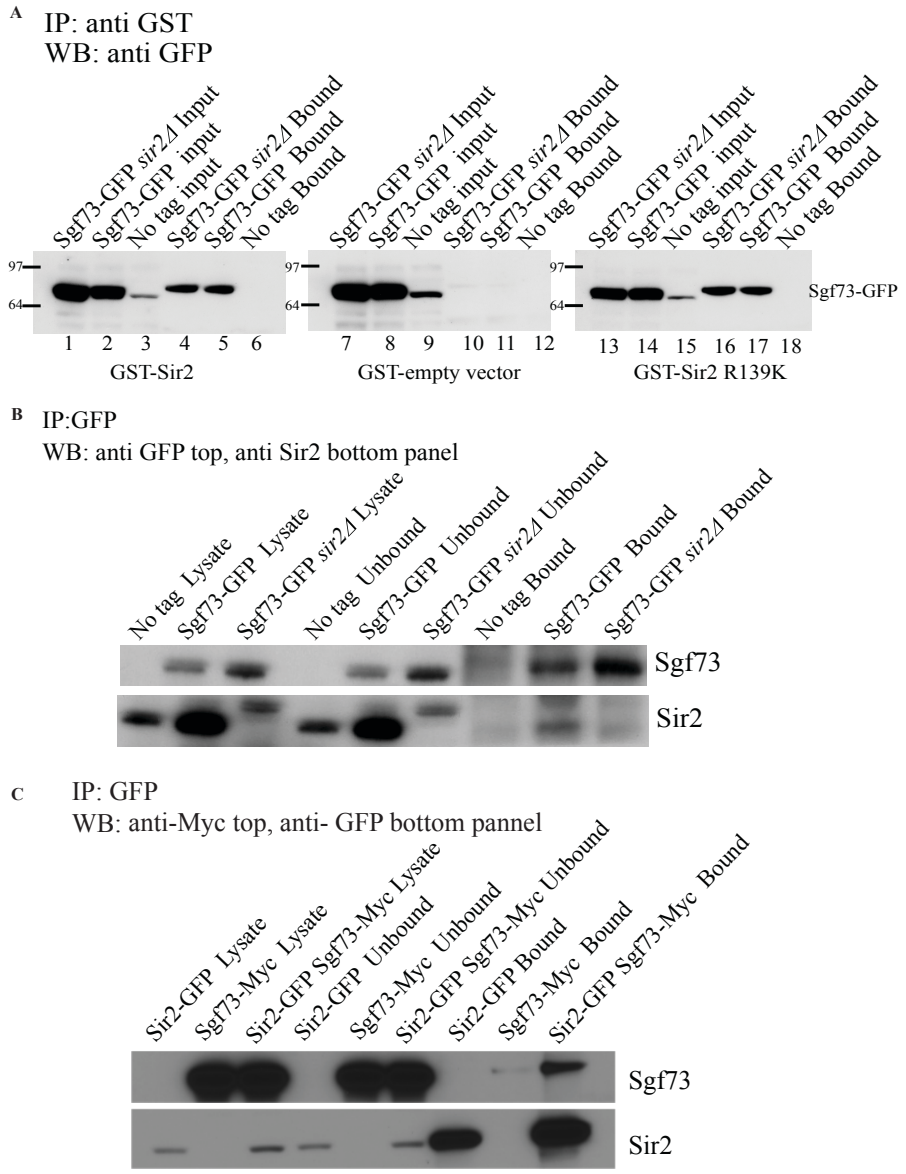


Figure 2-2: Sgf73 interaction with Sir2.
A) Interaction of recombinant Sir2-GST and Sgf73-GFP in the presence and absence of Sir2. Immunoprecipitation was done with glutathione beads coupled to either GST tagged Sir2, GST tagged enzymatically dead Sir2R319K, or GST, and western blot was probed with anti-GFP. B) Interaction of endogenous Sir2 and Sgf73-GFP. Immunoprecipitation was done using anti-GFP antibody and western blots were probed using anti-GFP (top panel) and anti-Sir2 (bottom panel) antibody. C) Interaction of Sir2-GFP and Sgf73-13Myc. Immunoprecipitation was done using anti-GFP antibody and western blots were probed using anti-Myc (top panel) and anti-GFP (bottom panel) antibodies.

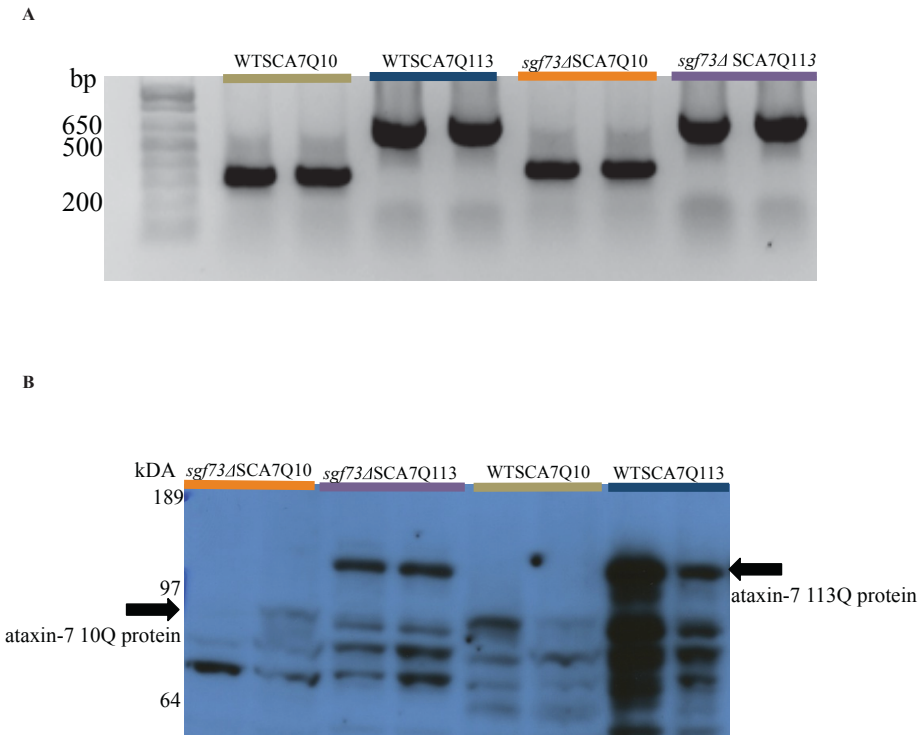


Figure 2-3: Integration of ataxin-7 10Q and 113Q into *sgf73Δ* yeast
A) PCR on genomic DNA from *sgf73Δ* strains post integration of SCA7Q10 or SCA7Q113 into the *URA3* locus. B) Western blot on whole cell lysates from *sgf73Δ* strains post integration of SCA7Q10 or SCA7Q113 into the *URA3* locus. Blot is probed with anti-ataxin-7 antibody K (La Spada lab).

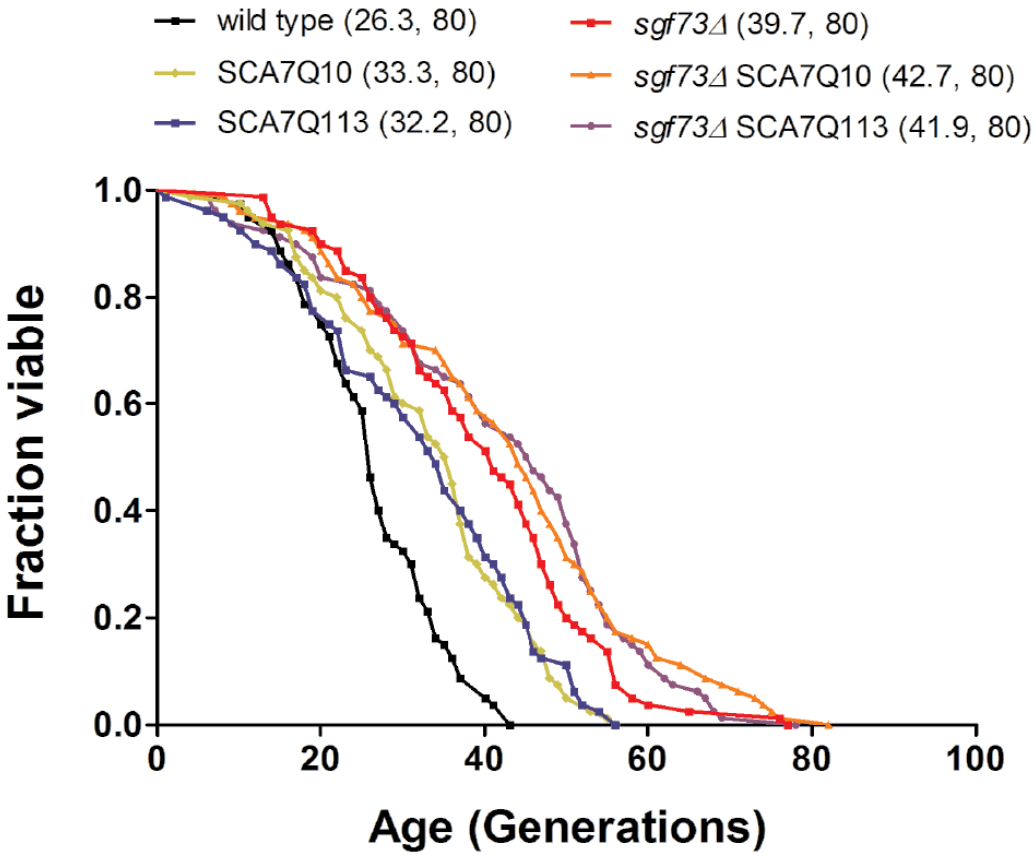


Figure 2-4: Replicative life span (RLS) plots on SCA7Q10 and SCA7Q113 integrated strains

Plot shows the effect of human ataxin-7 10Q (SCA7Q10) or 113Q (SCA7Q113) expression on the RLS (number of daughter cells produced) in either the WT or *sgf73Δ* background. There is an extension of RLS regardless of strain background. Legends show (mean RLS, number of mother cells scored)

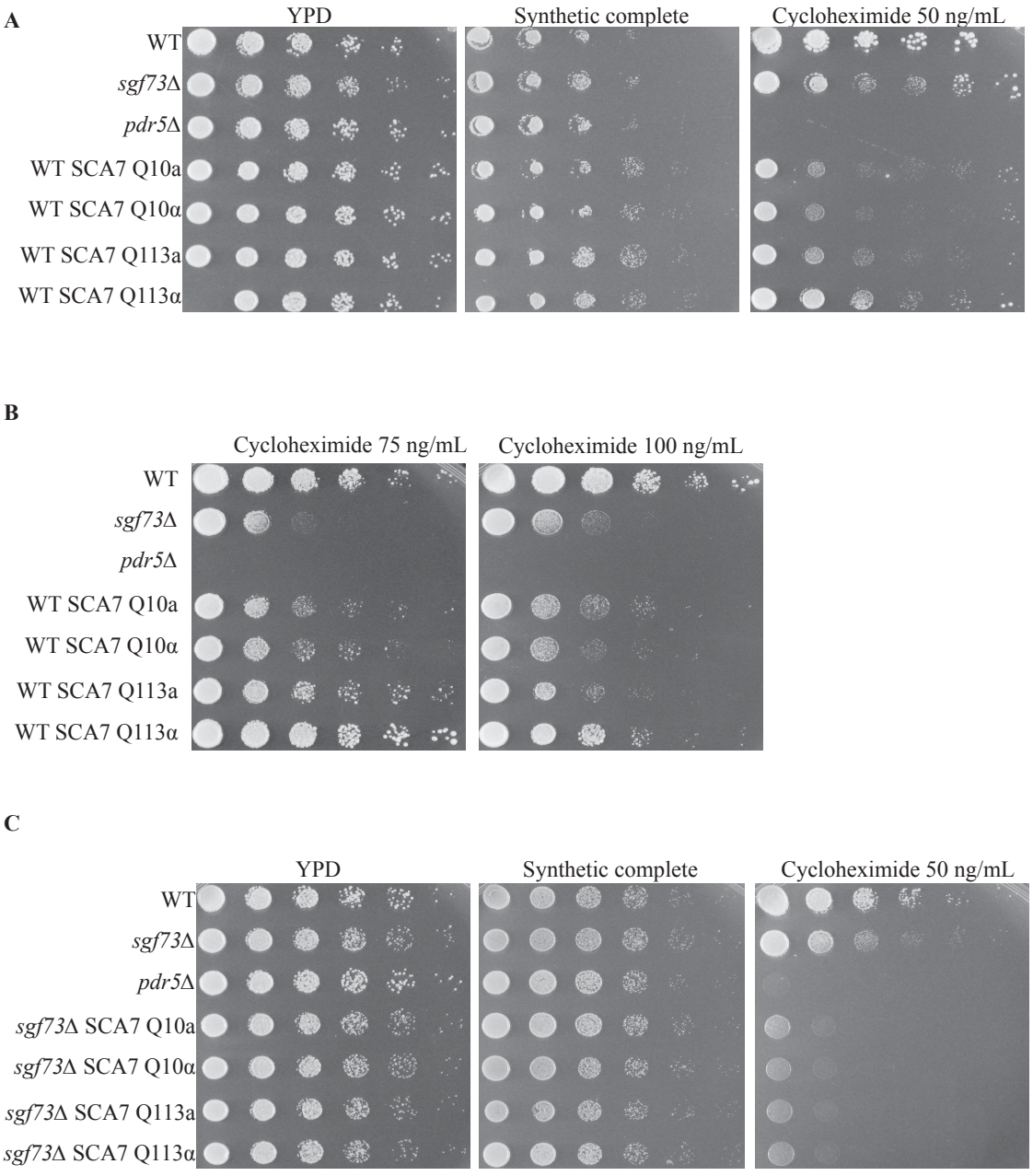


Figure 2-5: Cycloheximide sensitivity dilution assay of ataxin-7 integrated yeast strains

A) Testing cycloheximide sensitivity of WT strains integrated with SCA7Q10 and SCA7Q113, integrated strains are more sensitive to 50ng/mL cycloheximide than the WT strain. B) Exacerbating sensitivity of WT integrated SCA7Q10 and SCA7Q113 strains by increasing cycloheximide concentration to 75ng/mL and 100ng/mL. The WT integrated strain show a sensitivity between WT and *sgf73Δ*. C) Testing cycloheximide sensitivity of *sgf73Δ* strains integrated with SCA7Q10 and SCA7Q113, integrated strains are more sensitive to cycloheximide than the *sgf73Δ* strain

Discussion:

In this study, a collaborative effort between the Kennedy, Pillus, and La Spada labs, we report that the deletion of SAGA DUBm components *SGF73*, *UBP8*, and *SGF11* result in extended RLS, with the deletion of *SGF73* resulting in the longest life extension. We find that life extension in *sgf73Δ* strains requires Sir2 and/or Fob1 and is also dependent on an otherwise functioning SAGA but not SLIK complex. We find that the *sgf73Δ* strains have increased Sir2 dependent functions and that there is a physical interaction between Sir2 and Sgf73. Additionally we find that when the human ortholog of *SGF73*, ataxin-7, is expressed in the *sgf73Δ* strain it is not able to rescue the RLS phenotype and in fact causes RLS extension when expressed in WT yeast. Our work leads us to believe that the RLS extension seen in *sgf73Δ* yeast is a result of the dysregulation of otherwise coordinated chromatin-modifying activities.

Sir2 has long been a player in the yeast aging field, with seminal discoveries finding that the overexpression of Sir2 results in RLS extension [2]. One of the main activities attributed to Sir2OE life extension is its repression if rDNA recombination, limiting ERC formation [2,50]. ERC accumulation promotes aging and thus reducing ERC formation increases life span. Our experiments find that there is reduced rDNA recombination in the *sgf73Δ* yeast, consistent with promoting longevity, and implicates Sir2 function in *sgf73Δ*. Additionally it is believed that lower levels of rDNA replication leads to increased successful genome replication, increasing lifespan [51]. Our findings thereby suggest that reduced rDNA recombination in the *sgf73Δ* yeast leads to RLS extension through successful genome replication for a longer period of time.

SAGA is known as a stress-induced transcriptional co-activator complex [52], but more recently it has been shown to modulate the expression of many growth related genes in an acetyl-CoA dependent fashion [53]. Upon entry into the oxidative growth phase, a sharp increase in acetyl-CoA causes Gcn5 acetylation of a number of H3 sites, including the activating mark H3K9, at genes necessary for growth, including a majority of the ribosomal genes [53,54]. Many ribosomal genes are regulated by Ifh1 [55], a transcription factor that is recruited to ribosomal gene promoters upon nutrient stimulus [56]. Ifh1 becomes inactive upon Gcn5-mediated acetylation. Sir2 along with Hst1 then deacetylates Ifh1 [57]. This is consistent with the previous findings of SAGA occupation at ribosomal protein genes [53,58]. Additionally, the deletion of many ribosomal protein genes results in RLS extension. Our findings, along with previous studies, suggest that Sgf73 coordination of SAGA and Sir2 activities may be modulating ribosomal protein levels contributing to RLS extension. The deletion of *SGF73* could interfere with the recruitment of Sir2 to deacetylate Ifh1. Ifh1 would remain in the inactive, Gcn5-acetylated form thereby reducing ribosomal biogenesis.

The yeast *SGF73* has limited sequence homology to human ataxin-7 with two conserved domains: the SCA7 box, known to be important for Sgf73 nucleosome binding, and an amino terminal zinc-finger domain known to be important for Sgf73 binding to Ubp8 [11,18]. Since the two proteins share these critical partner interaction domains we hypothesized that the ataxin-7 protein might be able to substitute for Sgf73 in the *sgf73Δ* strain reverting the RLS phenotype. However this was not the case, with neither the SCA7Q10 or SCA7Q113 protein able to compensate for Sgf73 loss. The human protein may not be able to substitute for the yeast protein due to

protein structural differences as the proteins differ in predicted helical and turn structures (http://www.uniprot.org/)[59]. A definitive answer is complicated by the fact that the 3D structure is not known for the same conserved regions of the proteins. The majority of both proteins is currently listed as disordered. Only the first region of Sgf73 (aa 5-95, Supplementary Figure 2-1 a) is predicted, but the majority of the SCA7 domain of ataxin-7 (aa 339-397 Supplementary Figure 2-1 b) is predicted (phyre2 [60]). Even though human ataxin-7 can not be substituted for Sgf73, we find that it does interfere with normal Sgf73 function when expressed in WT yeast. WT strains expressing either SCA7Q10 or SCA7Q113 show a phenotype similar to *sgf73Δ*. Though conservation between the proteins is not enough to substitute for function, it appears to be enough to disrupt endogenous Sgf73 activity. Further studies are needed to determine if ataxin-7 substitution induces SAGA structural problems.

The studies into the mechanisms behind increased RLS in *sgf73Δ* are especially interesting since a polyglutamine expansion in the human ortholog, ataxin-7, causes the dominantly inherited neurodegenerative disease SCA7 [8]. The finding that *sgf73Δ* yeast have altered Sir2-dependent activity, as well as a defect in Ubp8 ubiquitination implicates these functions may also be altered when ataxin-7 is polyglutamine expanded. Additionally, the interaction that we observe between Sgf73 and Sir2 is supported by a recent study showing that the human ortholog of Ubp8, USP22, and the Sir2 ortholog, Sirt1, interact [40,41]. This suggests that there may be altered Sirt1 function in SCA7 patients caused by expanded ataxin-7 and this altered function may contribute to disease pathogenesis. Polyglutamine expanded ataxin-7 is integrated into the STAGA complex [61], and it has previously been reported to interfere

with GCN5-dependent histone acetyltransferase activity [62]. However changes in STAGA USP22 function in the presence of polyglutamine expanded ataxin-7 have not been documented. With our findings of altered Ubp8 function contributing to *sgf73Δ* RLS, it is likely that USP22 function is altered when expanded ataxin-7 is incorporated into STAGA. Our studies predict that expanded ataxin-7 might alter the normal function of Sirt1 and USP22, thereby contributing to SCA7 phenotypes. These downstream activities that affect disease pathogenesis as a result of expanded ataxin-7 will be important targets for further studies of neural function and aging. Additionally, continued studies of the yeast ortholog *SGF73* will likely uncover more avenues of study that are important to both neurodegeneration and the aging field. Overlap with the mammalian system can be explored for potential avenues for SCA7 therapeutic development as well as a better understanding of aging in general.

Materials and Methods

Strains and media

All yeast strains were derived from the parent strains of the haploid yeast ORF deletion collections [63], BY4742 (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) and BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*). *sgf73Δ* was reconstructed by deleting SGF73 via homologous recombination of a selectable *URA3* marker in the deletion collection wild-type BY strain using standard PCR-mediated gene disruption, and this parent was used to generate all *sgf73Δ* containing double and triple mutants. RGY43 was a gift of Richard Gardner. All strains used in this study are summarized in Supplementary Table 1.

Ataxin-7 10Q and 113Q (SCA7Q10 and SCA7Q113) integrated strains were made using the vector backbone pGADT7 AD where *LEU2* had been deleted and a 500bp fragment of the *URA3* gene inserted in place of the 2u ori along with the complete ataxin-7 cDNA 10Q (plasmid 2523-6) or 113Q (plasmid 2941-4) sequence. The plasmid was then linearized using BstBI, which cuts in the middle of the *URA3* cassette and the linearized plasmid was transfected into WT and *sgf73Δ* competent cells and integrated into the *URA3* locus via homologous recombination.

Cells were grown on standard YPD containing 1% yeast extract, 2% peptone and 2% glucose.

Replicative Lifespan

RLS assays were performed as described previously [64], and were all performed in the Kennedy lab at the Buck Institute for Regenerative Medicine. p-values for RLS

survival curve comparisons were calculated using a Wilcoxon rank-sum test [65]. Kaplan-Meier survival curves [66] were plotted using Prism (GraphPad, USA).

Yeast Interaction Studies

Yeast Lysis: Standard yeast glass bead lysis protocol was followed. Briefly; Yeast were grown to an A600 of 0.8 and lysed in 50mM HEPES pH 7.5, 0.1M NaCl, 0.5% Np40, 10% glycerol, 1mM EDTA, with protease inhibitors including PMSF. Lysates were pre-cleared with 10µl of protein A/G (dynabeads, Life technologies) magnetic beads at 4°C for one hour.

Recombinant protein purification: BL21 cells were transformed with one of three plasmids Sir2-GST, Sir2R139K-GST, or GST empty vector. A 100ml culture was grown to A600 0.6 and IPTG was added to a final concentration of 0.5mM. The culture was then incubated with shaking at room temperature for 4 hours. Cells were collected and lysed (20mM Tris pH 8, 1mM EDTA, 1mM EGTA, 1% NP-40, 350mM NaCl, 10mM dithiothreitol, protease inhibitors (Roche 04693124001), PMSF, 200mg/ml lysozyme) for 30 min on ice. Samples were then sonicated for 3 times 1 minute with 5 sec on followed by 0.5 sec off pulses. Samples were then spun at 4°C for 20 minutes. Glutathione agarose was prepared and added to the bacterial lysates and incubated at 4°C for one hour.

Strains: All tagged protein strains used were endogenously tagged with either GFP or 13Myc.

Immunoprecipitation: In Vitro: Conjugated beads were added to prepared yeast lysates and rotated at 4°C overnight. Beads were washed twice with yeast lysis buffer and twice with IP wash buffer (50mM HEPES pH 7.5, 150mM NaCl, 1mM EDTA).

Proteins were analyzed by western blot. In vivo: Anti-GFP antibody (living colors JL8 632381) was added to yeast lysates and rotated at 4°C overnight, 50ul of protein A/G beads were then added and sample rotated for 1 hour at 4°C. Beads were washed twice with lysis buffer and twice with wash buffer. Protein was analyzed via immunoblotting.

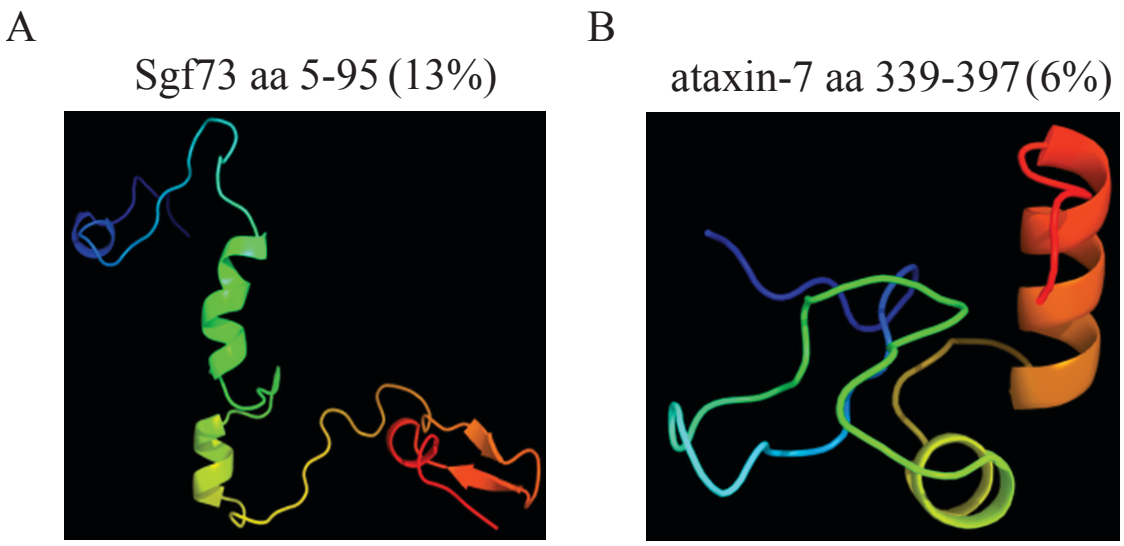
Immunoblot: Standard SDS-PAGE was used to separate proteins on 4-12% gels. After transfer to PVDF, membrane was blocked with 5% milk and incubated with either anti-GFP (living colors JL8 632381) or anti-Sir2 [67].

Dilution Assay

Strains were grown overnight at 30°C in YPD, one OD of cells was then washed and re-suspended in 1ml of water. 5 fold dilutions were then made and spotted on YPD, or cycloheximide plates and allowed to grow at 30°C.

Acknowledgments

Chapter 2, in part, is a reprint of material that is currently submitted for publication. Mark A. McCormick[*], Amanda G. Mason[*] Stephan J. Guyenet[*], Weiwei Dang, Renee M. Garza, Marc K. Ting, Rick M. Moller, Shelley L. Berger, Matt Kaeberlein, Lorraine Pillus, Albert R. La Spada, and Brian K. Kennedy, 2014. The thesis author was one of the primary investigators and authors of this paper.



Supplementary Figure 2-1: Partial predicted protein structures of Sgf73 and ataxin-7

A) Predicted structure of the first region of Sgf73 (aa 5-95) is predicted B) Predicted structure for aa 339-397 of ataxin-7 corresponding to the majority of the SCA7 box domain (aa 334-402) of ataxin-7. All structures were predicted using phyre2 [60].

Supplementary Table 2-1: Yeast strains used in studies

Strain number	Genotype	MAT
LPy6494	WT	α
LPy18835	*sgf73Δ*	α
LPy18910	Sgf73-13Myc	α
LPy18919	*sir2Δ fob1Δ*	α
LPy17446	*pdr5Δ*	α
LPy20492	Sir2-GFP Sgf73-13Myc	α
LPy18882	Sgf73-GFP	α
LPy20502	Sir2-GFP	α
LPy18884	Sgf73-GFP *sir2Δ* *	α
LPy19267	*sgf73Δ* SCA7Q10	a
LPy19268	*sgf73Δ* SCA7Q10	α
LPy19313	*sgf73Δ* SCA7Q113	a
LPy19314	*sgf73Δ* SCA7Q113	α
LPy19269	WT SCA7Q10	a
LPy19270	WT SCA7Q10	α
LPy19311	WT SCA7Q113	a
LPy19312	WT SCA7Q113	α

* Frozen with *SIR2* covering plasmid

References:

1. Mortimer, R.K. & Johnston, J.R. Life span of individual yeast cells. *Nature* **183**, 1751-1752 (1959).

2. Kaeberlein, M., McVey, M. & Guarente, L. The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms. *Genes Dev* **13**, 2570-80 (1999).

3. Blander, G. & Guarente, L. The Sir2 family of protein deacetylases. *Annu Rev Biochem* **73**, 417-35 (2004).

4. Rogina, B. & Helfand, S.L. Sir2 mediates longevity in the fly through a pathway related to calorie restriction. *Proc Natl Acad Sci U S A* **101**, 15998-6003 (2004).

5. Chua, K.F., Mostoslavsky, R., Lombard, D.B., Pang, W.W., Saito, S., Franco, S., Kaushal, D., Cheng, H.L., Fischer, M.R., Stokes, N., Murphy, M.M., Appella, E. & Alt, F.W. Mammalian SIRT1 limits replicative life span in response to chronic genotoxic stress. *Cell Metab* **2**, 67-76 (2005).

6. Langley, E., Pearson, M., Faretta, M., Bauer, U.M., Frye, R.A., Minucci, S., Pelicci, P.G. & Kouzarides, T. Human SIR2 deacetylates p53 and antagonizes PML/p53-induced cellular senescence. *EMBO J* **21**, 2383-96 (2002).

7. Chang, H.C. & Guarente, L. SIRT1 mediates central circadian control in the SCN by a mechanism that decays with aging. *Cell* **153**, 1448-60 (2013).

8. David, G., Abbas, N., Stevanin, G., Durr, A., Yvert, G., Cancel, G., Weber, C., Imbert, G., Saudou, F., Antoniou, E., Drabkin, H., Gemmill, R., Giunti, P., Benomar, A., Wood, N., Ruberg, M., Agid, Y., Mandel, J.L. & Brice, A. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat Genet* **17**, 65-70 (1997).

9. Stevanin, G., Durr, A. & Brice, A. Clinical and molecular advances in autosomal dominant cerebellar ataxias: from genotype to phenotype and physiopathology. *Eur J Hum Genet* **8**, 4-18 (2000).

10. Lebre, A.S. & Brice, A. Spinocerebellar ataxia 7 (SCA7). *Cytogenet Genome Res* **100**, 154-63 (2003).

11. Helmlinger, D., Hardy, S., Sasorith, S., Klein, F., Robert, F., Weber, C., Miguet, L., Potier, N., Van-Dorsselaer, A., Wurtz, J.M., Mandel, J.L., Tora, L. & Devys, D. Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. *Hum Mol Genet* **13**, 1257-65 (2004).

12. Martinez, E., Palhan, V.B., Tjernberg, A., Lymar, E.S., Gamper, A.M., Kundu, T.K., Chait, B.T. & Roeder, R.G. Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors in vivo. *Mol Cell Biol* **21**, 6782-95 (2001).

13. Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y. & Allis, C.D. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**, 843-51 (1996).

14. Zhang, X.Y., Pfeiffer, H.K., Thorne, A.W. & McMahon, S.B. USP22, an hSAGA subunit and potential cancer stem cell marker, reverses the polycomb-catalyzed ubiquitylation of histone H2A. *Cell Cycle* **7**, 1522-4 (2008).

15. Zhao, Y., Lang, G., Ito, S., Bonnet, J., Metzger, E., Sawatsubashi, S., Suzuki, E., Le Guezennec, X., Stunnenberg, H.G., Krasnov, A., Georgieva, S.G., Schule, R., Takeyama, K., Kato, S., Tora, L. & Devys, D. A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. *Mol Cell* **29**, 92-101 (2008).

16. Mushegian, A.R., Vishnivetskiy, S.A. & Gurevich, V.V. Conserved phosphoprotein interaction motif is functionally interchangeable between ataxin-7 and arrestins. *Biochemistry* **39**, 6809-13 (2000).

17. Scheel, H., Tomiuk, S. & Hofmann, K. Elucidation of ataxin-3 and ataxin-7 function by integrative bioinformatics. *Hum Mol Genet* **12**, 2845-52 (2003).

18. Bonnet, J., Wang, Y.H., Spedale, G., Atkinson, R.A., Romier, C., Hamiche, A., Pijnappel, W.W., Timmers, H.T., Tora, L., Devys, D. & Kieffer, B. The structural plasticity of SCA7 domains defines their differential nucleosome-binding properties. *EMBO Rep* **11**, 612-8 (2010).

19. Timmers, H.T. & Tora, L. SAGA unveiled. *Trends Biochem Sci* **30**, 7-10 (2005).

20. Henry, K.W., Wyce, A., Lo, W.S., Duggan, L.J., Emre, N.C., Kao, C.F., Pillus, L., Shilatifard, A., Osley, M.A. & Berger, S.L. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev* **17**, 2648-63 (2003).

21. Baker, S.P. & Grant, P.A. The SAGA continues: expanding the cellular role of a transcriptional co-activator complex. *Oncogene* **26**, 5329-40 (2007).

22. Kohler, A., Schneider, M., Cabal, G.G., Nehrbass, U. & Hurt, E. Yeast Ataxin-7 links histone deubiquitination with gene gating and mRNA export. *Nat Cell Biol* **10**, 707-15 (2008).

23. Lee, K.K., Swanson, S.K., Florens, L., Washburn, M.P. & Workman, J.L. Yeast Sgf73/Ataxin-7 serves to anchor the deubiquitination module into both SAGA and Slik(SALSA) HAT complexes. *Epigenetics Chromatin* **2**, 2 (2009).

24. Ruiz-Garcia, A.B., Sendra, R., Pamblanco, M. & Tordera, V. Gcn5p is involved in the acetylation of histone H3 in nucleosomes. *FEBS Lett* **403**, 186-90 (1997).

25. Grant, P.A., Duggan, L., Cote, J., Roberts, S.M., Brownell, J.E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C.D., Winston, F., Berger, S.L. & Workman, J.L. Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev* **11**, 1640-50 (1997).

26. Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y., Winston, F. & Allis, C.D. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes Dev* **15**, 3286-95 (2001).

27. Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J. & Kouzarides, T. Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407-11 (2002).

28. Sun, Z.W. & Allis, C.D. Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* **418**, 104-8 (2002).

29. Jiang, L., Smith, J.N., Anderson, S.L., Ma, P., Mizzen, C.A. & Kelleher, N.L. Global assessment of combinatorial post-translational modification of core histones in yeast using contemporary mass spectrometry. LYS4 trimethylation correlates with degree of acetylation on the same H3 tail. *J Biol Chem* **282**, 27923-34 (2007).

30. Sterner, D.E., Belotserkovskaya, R. & Berger, S.L. SALSA, a variant of yeast SAGA, contains truncated Spt7, which correlates with activated transcription. *Proc Natl Acad Sci U S A* **99**, 11622-7 (2002).

31. Pray-Grant, M.G., Schieltz, D., McMahon, S.J., Wood, J.M., Kennedy, E.L., Cook, R.G., Workman, J.L., Yates, J.R., 3rd & Grant, P.A. The novel SLIK histone acetyltransferase complex functions in the yeast retrograde response pathway. *Mol Cell Biol* **22**, 8774-86 (2002).

32. Spedale, G., Mischerikow, N., Heck, A.J., Timmers, H.T. & Pijnappel, W.W. Identification of Pep4p as the protease responsible for formation of the SAGA-related SLIK protein complex. *J Biol Chem* **285**, 22793-9 (2010).

33. Wu, P.Y. & Winston, F. Analysis of Spt7 function in the Saccharomyces cerevisiae SAGA coactivator complex. *Mol Cell Biol* **22**, 5367-79 (2002).

34. Belotserkovskaya, R., Sterner, D.E., Deng, M., Sayre, M.H., Lieberman, P.M. & Berger, S.L. Inhibition of TATA-binding protein function by SAGA subunits Spt3 and Spt8 at Gcn4-activated promoters. *Mol Cell Biol* **20**, 634-47 (2000).

35. Smolle, M., Venkatesh, S., Gogol, M.M., Li, H., Zhang, Y., Florens, L., Washburn, M.P. & Workman, J.L. Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nat Struct Mol Biol* **19**, 884-92 (2012).

36. Shilatifard, A. Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annu Rev Biochem* **75**, 243-69 (2006).

37. Schulze, J.M., Hentrich, T., Nakanishi, S., Gupta, A., Emberly, E., Shilatifard, A. & Kobor, M.S. Splitting the task: Ubp8 and Ubp10 deubiquitinate different cellular pools of H2BK123. *Genes Dev* **25**, 2242-7 (2011).

38. Kaeberlein, M., Kirkland, K.T., Fields, S. & Kennedy, B.K. Genes determining yeast replicative life span in a long-lived genetic background. *Mech. Ageing Dev.* **126**, 491-504 (2005).

39. Sowa, M.E., Bennett, E.J., Gygi, S.P. & Harper, J.W. Defining the human deubiquitinating enzyme interaction landscape. *Cell* **138**, 389-403 (2009).

40. Lin, Z., Yang, H., Kong, Q., Li, J., Lee, S.M., Gao, B., Dong, H., Wei, J., Song, J., Zhang, D.D. & Fang, D. USP22 antagonizes p53 transcriptional activation by deubiquitinating Sirt1 to suppress cell apoptosis and is required for mouse embryonic development. *Mol Cell* **46**, 484-94 (2012).

41. Armour, S.M., Bennett, E.J., Braun, C.R., Zhang, X.Y., McMahon, S.B., Gygi, S.P., Harper, J.W. & Sinclair, D.A. A high-confidence interaction map identifies SIRT1 as a mediator of acetylation of USP22 and the SAGA coactivator complex. *Mol Cell Biol* **33**, 1487-502 (2013).

42. Kaeberlein, M., Kirkland, K.T., Fields, S. & Kennedy, B.K. Sir2-independent life span extension by calorie restriction in yeast. *PLOS Biology* **2**, 1381-1387 (2004).

43. Delaney, J.R., Sutphin, G.L., Dulken, B., Sim, S., Kim, J.R., Robison, B., Schleit, J., Murakami, C.J., Carr, D., An, E.H., Choi, E., Chou, A., Fletcher, M., Jelic, M., Liu, B., Lockshon, D., Moller, R.M., Pak, D.N., Peng, Q., Peng, Z.J., Pham, K.M., Sage, M., Solanky, A., Steffen, K.K., Tsuchiya, M., Tsuchiyama, S., Johnson, S., Raabe, C., Suh, Y., Zhou, Z., Liu, X., Kennedy, B.K. & Kaeberlein, M. Sir2 deletion prevents lifespan extension in 32 long-lived mutants. *Aging Cell* (2011).

44. Defossez, P.A., Prusty, R., Kaeberlein, M., Lin, S.J., Ferrigno, P., Silver, P.A., Keil, R.L. & Guarente, L. Elimination of replication block protein Fob1 extends the life span of yeast mother cells. *Mol Cell* **3**, 447-55. (1999).

45. Dang, W., Steffen, K.K., Perry, R., Dorsey, J.A., Johnson, F.B., Shilatifard, A., Kaeberlein, M., Kennedy, B.K. & Berger, S.L. Histone H4 lysine 16 acetylation regulates cellular lifespan. *Nature* **459**, 802-7 (2009).

46. Sinclair, D.A. & Guarente, L. Extrachromosomal rDNA circles-a cause of aging in yeast. *Cell* **91**, 1033-1042 (1997).

47. Moazed, D. Enzymatic activities of Sir2 and chromatin silencing. *Curr Opin Cell Biol* **13**, 232-8 (2001).

48. Jacobson, S. & Pillus, L. The SAGA subunit Ada2 functions in transcriptional silencing. *Mol Cell Biol* **29**, 6033-45 (2009).

49. Alamgir, M., Erukova, V., Jessulat, M., Azizi, A. & Golshani, A. Chemical-genetic profile analysis of five inhibitory compounds in yeast. *BMC Chem Biol* **10**, 6 (2010).

50. Gottlieb, S. & Esposito, R.E. A new role for a yeast transcriptional silencer gene, SIR2, in regulation of recombination in ribosomal DNA. *Cell* **56**, 771-6. (1989).

51. Kwan, E.X., Foss, E.J., Tsuchiyama, S., Alvino, G.M., Kruglyak, L., Kaeberlein, M., Raghuraman, M.K., Brewer, B.J., Kennedy, B.K. & Bedalov, A. A Natural Polymorphism in rDNA Replication Origins Links Origin Activation with Calorie Restriction and Lifespan. *PLoS Genet* **9**, e1003329 (2013).

52. Huisinga, K.L. & Pugh, B.F. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. *Mol Cell* **13**, 573-85 (2004).

53. Cai, L., Sutter, B.M., Li, B. & Tu, B.P. Acetyl-CoA induces cell growth and proliferation by promoting the acetylation of histones at growth genes. *Mol Cell* **42**, 426-37 (2011).

54. Tu, B.P., Kudlicki, A., Rowicka, M. & McKnight, S.L. Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science* **310**, 1152-8 (2005).

55. Warner, J.R. The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* **24**, 437-40 (1999).

56. Schawalder, S.B., Kabani, M., Howald, I., Choudhury, U., Werner, M. & Shore, D. Growth-regulated recruitment of the essential yeast ribosomal protein gene activator Ifh1. *Nature* **432**, 1058-61 (2004).

57. Downey, M., Knight, B., Vashisht, A.A., Seller, C.A., Wohlschlegel, J.A., Shore, D. & Toczyski, D.P. Gcn5 and sirtuins regulate acetylation of the ribosomal protein transcription factor ifh1. *Curr Biol* **23**, 1638-48 (2013).

58. Robert, F., Pokholok, D.K., Hannett, N.M., Rinaldi, N.J., Chandy, M., Rolfe, A., Workman, J.L., Gifford, D.K. & Young, R.A. Global position and recruitment of HATs and HDACs in the yeast genome. *Mol Cell* **16**, 199-209 (2004).

59. UniProt, C. Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res* **41**, D43-7 (2013).

60. Kelley, L.A. & Sternberg, M.J. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363-71 (2009).

61. Helmlinger, D., Hardy, S., Eberlin, A., Devys, D. & Tora, L. Both normal and polyglutamine- expanded ataxin-7 are components of TFTC-type GCN5 histone acetyltransferase- containing complexes. *Biochem Soc Symp*, 155-63 (2006).

62. Palhan, V.B., Chen, S., Peng, G.H., Tjernberg, A., Gamper, A.M., Fan, Y., Chait, B.T., La Spada, A.R. & Roeder, R.G. Polyglutamine-expanded ataxin-7 inhibits STAGA histone acetyltransferase activity to produce retinal degeneration. *Proc Natl Acad Sci U S A* **102**, 8472-7 (2005).

63. Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., Chu, A.M., Connelly, C., Davis, K., Dietrich, F., Dow, S.W., El Bakkoury, M., Foury, F.,

Friend, S.H., Gentalen, E., Giaever, G., Hegemann, J.H., Jones, T., Laub, M., Liao, H., Liebundguth, N., Lockhart, D.J., Lucau-Danila, A., Lussier, M., M'Rabet, N., Menard, P., Mittmann, M., Pai, C., Rebischung, C., Revuelta, J.L., Riles, L., Roberts, C.J., Ross-MacDonald, P., Scherens, B., Snyder, M., Sookhai-Mahadeo, S., Storms, R.K., Veronneau, S., Voet, M., Volckaert, G., Ward, T.R., Wysocki, R., Yen, G.S., Yu, K., Zimmermann, K., Philippsen, P., Johnston, M. & Davis, R.W. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285**, 901-6 (1999).

64. Steffen, K.K., Kennedy, B.K. & Kaeberlein, M. Measuring replicative life span in the budding yeast. *J Vis Exp* (2009).

65. Wilcoxon, F. Individual comparisons of grouped data by ranking methods. *J Econ Entomol* **39**, 269 (1946).

66. Kaplan, E.L. & Meier, P. Nonparametric estimation from incomplete observations. *J Amer Statist Assn* **53**(1958).

67. Garcia, S.N. & Pillus, L. A unique class of conditional sir2 mutants displays distinct silencing defects in Saccharomyces cerevisiae. *Genetics* **162**, 721-36 (2002).

Chapter 3: Profiling genomic occupancy of Sgf73, Ubp8, and Sir2

Introduction:

The SAGA (Spt-Ada-Gcn5 acetyltransferase) complex is a major transcriptional co-activator complex responsible for the regulation of approximately 10% of yeast genes, the majority of which are stress induced [1]. SAGA is recruited to genomic locations by a variety of transcription factors where it is involved in gene activation through both its Gcn5 acetylation and Ubp8 deubiquitination chromatin modifying activities, and by facilitating binding of TATA-binding protein (TBP) at promoters to aid in transcriptional activation and elongation (reviewed in [2]). The SAGA complex has four distinct modules responsible for acetyltransferase activity, deubiquitination activity, complex recruitment, and complex architecture (Figure 3-1a). Sgf73 is one subunit of the SAGA complex [3,4]. It acts in the Ubp8 deubiquitinase module (DUBm), linking the core SAGA complex harboring Gcn5 acetylation activity to the Ubp8 DUBm [5-7]. Sgf73 contains two highly conserved zinc-finger binding domains, one in the amino-terminal domain that interacts with the Ubp8 module, and a separate zinc-finger domain (aa 227-272) within the Sca7 box that binds nucleosomes and is within the region involved in Sgf73 binding to SAGA [3,5,8] (Figure 3-1 b).

The deletion of *SGF73* in the budding yeast *Saccharomyces cerevisiae* results in a dramatic extension in replicative lifespan (RLS), i.e. the number of daughter cells that a mother cell can produce (McCormick, Mason, & Guyenet et al submitted). *SGF73* is required for Ubp8-mediated histone H2B deubiquitination, with *sgf73Δ* strains having higher levels of monoubiquitinated H2BK123, as is seen in *ubp8Δ*

strains [5]. This increase in ubiquitinated H2BK123 contributes to the RLS extension of *sgf73Δ* mutants as strains that lack *UBP8* are also long lived, and strains that cannot ubiquitinate H2BK123 or harbor the H2BK123R mutation are short lived (McCormick, Mason, & Guyenet et al submitted).

The RLS extension in *sgf73Δ* mutants is Sir2 dependent (McCormick, Mason, & Guyenet et al submitted). *SIR2* encodes a NAD+ dependent deacetylase that has been implicated in both lifespan regulation and caloric restriction in a variety of species, making it a central focus of aging research. The modest overexpression of Sir2, by integration of a second copy under the endogenous promoter, has been shown to extend RLS by ~30% [9]. Sir2 is also required for silencing at telomeres, and represses rDNA recombination; rDNA recombination causes the accumulation of extrachromosomal ribosomal DNA circles (ERCs) a cause of aging in yeast [10]. It was also found that *sgf73Δ* mutants have altered Sir2-related activities (McCormick, Mason, & Guyenet et al submitted).

The human counterpart of *SGF73* is ataxin-7, a protein that causes spinocerebellar ataxia type 7 (SCA7) upon polyglutamine expansion. Ataxin-7 is a member of the mammalian transcriptional coactivating complex STAGA (Spt3-Taf9-Gcn5 acetyltransferase), the human SAGA equivalent [3,11]. Ataxin-7 has an N-terminal polyglutamine tract that ranges from 4-35 repeats in unaffected individuals, but which has the propensity to expand to repeat lengths greater than 37, causing the neurodegenerative disorder SCA7 [12,13]. SCA7 is one of nine known inherited polyglutamine expansion disorders that result from the transcription and translation of an expanded polyglutamine tract. Accumulation of expanded ataxin-7 protein leads to

the dysfunction and death of neurons in the retina, cerebellum, and cerebellar associated structures causing blindness, progressive loss of coordination, and premature death [14].

By exploring the chromatin regions bound by Sgf73, as well as two related proteins Ubp8 and Sir2 we hope to reveal genomic regions relevant to RLS. Sgf73, Ubp8, and Sir2 function to contribute and disrupt RLS extension upon *SGF73* deletion. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is a technique used to identify genomic regions bound by chromatin binding proteins [15]. ChIP-seq is a tool that facilitates the creation of protein global occupancy maps to determine where a protein is bound. These targeted regions are potentially regulated by the bound protein. Regions of the yeast genome bound by Sgf73, Ubp8, and Sir2 have implications in the field of yeast aging to help understand their regulatory roles. Also, findings will contribute to the study of SCA7 neurodegeneration as the functions of these proteins are deeply conserved.

Using a ChIP-seq approach I identified 389 genomic regions bound by Sgf73 and found that there was an enrichment of binding in the region 5' to many ribosomal protein encoding genes. We found a significant overlap in Sgf73- and Ubp8-bound genomic regions, and enrichment in ribosomal protein encoding genes. In addition, from 389 Sgf73 genomic targets found, 31 genes had previously been linked to RLS extension. Using these data integrated with transcriptome data from SAGA-related mutants, we validated a set of target genes subject to Sgf73 regulation and propose that these target genes are integral to pathways of aging regulation in yeast, and are likely candidates to SCA7-linked neurodegeneration in humans.

Results:

Myc tagging of Sgf73 and Ubp8 does not interfere with protein function

To perform ChIP on Sgf73 and Ubp8, it was necessary to epitope tag the proteins at their genomic loci because no protein-specific antibodies are. A 13-Myc epitope tag sequence was incorporated at the genomic locus of each gene independently as previously described [16], using G418 selection to identify potential tag integrants. Successful integration was verified by molecular genotyping. After screening we obtained one Ubp8 (Ubp8-13myc #12) and two Sgf73 (Sgf73-13myc #4, Sgf73-13myc #5) possitive clones (Figure 3-2 a). To ensure that the tagged protein was being successfully made in the yeast cells we confirmed expression of the tagged protein via immunoblot analysis (Figure 3-2 b). To evaluate Ubp8-13myc protein function, we performed a temperature sensitivity growth assay. It is known that *ubp8Δ* yeast have increased temperature sensitivity, having a slower growth rate at 37°C in comparison to growth at 30°C [17]. We performed a dilution assay with WT, *ubp8Δ*, and our tagged Ubp8-13myc #12 strain spotted on YPD plates at 30°, and 39°C. The growth difference was striking at 39°C between the WT and null strain but there were no significant differences between the WT and tagged strain (Figure 3-2 c), thus confirming function of the Ubp8-13myc strain. To evaluate Sgf73-13myc protein function, a cycloheximide sensitivity assay was performed. It has previously been reported that *sgf73Δ* yeast are sensitive to cycloheximide, having decreased growth relative to WT when plated on medium containing cycloheximide[18]. A dilution assay was prepared with WT, *sgf73Δ,* Sgf73-13myc #4, and Sgf73-13myc #5 on YPD plates

and on YPD plates containing 50ng/mL cycloheximide. There was a striking difference between WT and *sgf73Δ* strains grown on cycloheximide medium but not between the WT and tagged Sgf73-13myc strains (Figure 3-2 d), thus demonstrating that the tag does not interfere with Sgf73 function. From these experiments, we conclude that the yeast strains generated express functional tagged versions of Sgf73 and Ubp8, behave like WT, and are therefore appropriate for ChIP-seq experiments.

Sgf73, Ubp8, and Sir2 are successfully immunoprecipitated

After the successful tagging of Sgf73 and Ubp8, the proteins were ChIPed using an anti-c-myc resin, and Sir2 was ChIPed using an anti-Sir2 antibody. A WT untagged strain was used as a control for non-specific c-myc antibody resin binding. For Sir2 ChIP, using the Sir2 antibody, we used a strain lacking *SIR2* and *FOB1* as negative (non-specific) control. Strains lacking *SIR2* alone are short lived due in part ERC accumulation and deletion of *FOB1* limits ERC formation, rescuing lifespan to WT levels [19,20]. To check that the protein had been pulled down, input, unbound, and bound protein samples taken during the ChIP process were processed by immunoblot analysis. The immunoblots of the protein from the Sgf73-13myc ChIP and the Ubp8-13myc ChIP show that the baits are recovered with little remaining in the unbound fractions (Figure 3-3 a, b). The immunoblots on protein from the Sir2 ChIP show that the Sir2 signal is detectable and that the protein is bound by the Sir2 antibody during the ChIP process (Figure 3-3 c). These results show that all protein baits were effectively immunoprecipitated.

Established targets are bound by Sgf73, Ubp8, and Sir2

Based on reports in the literature, several genomic targets were chosen as positive controls for the ChIP procedures. To validate the targets we performed PCR on the purified DNA originating from the immunoprecipitated protein bait-DNA complexes. For Sir2 bound regions, primers that amplify the telomere VI-R as well as 5S and 25S rDNA were used as published in Darst et al. [21]. For Sgf73 and Ubp8 previously published primers for either the upstream activator sequences (UAS) or core promoter of SAGA dependent genes: *PHO84*-UAS, *CUP1*-UAS, *GAL1*-UAS, and *GAL1*-CORE were used [22,23]. PCR was performed on input samples as well as ChIPed DNA, revealing detection of the regions of interest in bound samples of interest but not in the control samples (Figure 3-4 a, b).

Library generation and library fragment size confirmation

ChIP experiments were performed in biological duplicate starting from two independent yeast colonies, and the cells were never frozen (6 myc ChIPs, and 8 Sir2 ChIPs). After ChIP, libraries were generated. For optimal sequencing conditions, seven samples would be multiplexed in one sequencing lane. Barcodes were incorporated into the library-making process to make multiplex sequencing possible. Each library was identified by number for sequencing: Table 3-1 outlines the original sample number and which strain and ChIP replicate it represents. After following the NuGen kit protocol, DNA library fragments were separated on a 6% acrylamide gel (Figure 3-5 a), size-selected for ~150-400bp fragments (Figure 3-5 b), and purified.

Successful fragment size selection was confirmed on the Agilent 2100 Bioanalyzer, as were the concentrations (in ng/μl) of the purified libraries calculated.

Sequence alignment with Bowtie and analysis using HOMER

After the raw sequence data was received, the quality of the reads was evaluated using the program FastQC (Supplementary Figures 3-1 to 3-14). Sequence duplication levels were high due to PCR amplification. This was compensated for in later analysis steps, as explained below. Raw sequence reads were then aligned to the *S. cerevisiae* S288C genome 3 (sacCer3) using the program Bowtie[24]. After sequence alignment the program HOMER (**H**ypergeometric **O**ptimization of **M**otif **E**n**R**ichment) [25] v4.2 was used to visualize all mapped reads on the University of California Santa Cruz (UCSC) genome browser. A "peak" is defined as a collection of sequence reads, collectively mapping to a genomic location at a significantly higher density than background and surrounding genomic area. Depiction of each chromosome on the UCSC genome browser [26] shows the clear identification of peaks between runs that are specific or shared for each bait protein. Also shown are peaks that are not specific as they are present in the control track, that is, merged reads from the no myc tag samples (Supplementary Figure 3-15 a-p).

Sgf73 and Ubp8 share 42 peaks enriched for ribosomal genes

Using HOMER, significant peaks, defined by a peak signal four times greater than the surrounding 10kb region, were identified in each replicate using the control ChIP as the background filter. There were 389 significant peaks shared between the

two Sgf73-13myc ChIP runs (Supplementary Table 3-1). These resulted in 388 unique elements because two significant peaks were 5' to the same transcriptional start site (TSS). We found 44 significant peaks present in both Ubp8-13myc ChIP runs (Supplementary Table 3-2). Of the 44 Ubp8 peaks, 42 were also found in the Sgf73 ChIPs. For these 42 peaks we manually annotated genes most proximal to the peak, the peak was denoted as 5' or 3' for genes within ~400bp (Table 3-2). Our gene annotation is in contrast to the output from HOMER, which only includes the closest 5' TSS to the peak. Our manual annotation expands the number of candidate genes, as peaks can be in close proximity to 5' TSS of two opposing genes as well as potentially important 3' regulatory sequences[27]. Peaks in the table are ranked by tag count, an indication of the number of unique reads mapped to that peak. Gene ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis[28,29] was performed to reveal enriched protein families or pathways (Figure 3-6). In the biological process category of the KEGG analysis there was enrichment for genes involved in 3'-end of small subunit (SSU)-rRNA processing and maturation, translation, for genes with roles in transition metal ion transport, and for genes that function in endogenous and chemical stimuli responses. In the molecular function category there was enrichment in oxidoreductase activity, ribosome structure, and metal ion transmembrane transporter activity. The only enriched cellular category was the ribosome. Most striking from this analysis is that the only enriched KEGG pathway was the ribosome, with 16 of the 42 peaks (38%) having Sgf73 & Ubp8 occupancy proximal to a ribosomal gene. Those ribosomal protein (RP) genes are:

RPL10, RPL20A, RPL23B, RPL27B, RPL28, RPL2A, RPL33A, RPL35A, RPL42A, RPL6A, RPL8A, RPS0A, RPS0B, RPS23A, RPS23B, RPS5.

Occupancy peaks from Sgf73 and Ubp8 ChIP-seq can be validated by ChIP-PCR

To validate Sgf73 and Ubp8 binding/occupancy sites we chose primers (Methods Table 3-1) spanning the significant peaks. DNA used for validation experiments was ChIPed in the same manner as that sequenced. We chose to validate peaks with the highest number of uniquely mapped reads (hereafter referred to as tag counts), which are: 5' *YPR036W-A*; 5' *BTN2* and 3'*VPS62*; and 5' *FET3* and *AAC1*, as well as a few peaks near RP genes (5'*RPL6A*, 5' *RPL10*, 5' *RPL20A*). The sequencing peaks for these six locations are clearly visible and significant from traces on the UCSC genome browser [26] (Figure 3-7 a-f). We were able to validate our sequenced peaks by seeing that there was at least a 3-fold increase in the ChIPed samples of interest over the negative ChIP control (Figure 3-8 a-f). Additionally, some regions lacking mapped reads in the ChIP-seq data served as additional negative controls (Figure 3-9). Based on the ChIP-PCR validation we are confident that our ChIP-seq experiment was successful and specific in identifying regions of DNA that our proteins of interest bound.

Translational differences of genes with Sgf73 and Ubp8 occupancy upon their deletion

Our expectation was that Sgf73 and Ubp8 occupancy indicates that the region was not only bound by these proteins, but by the SAGA transcriptional co-activator

complex. Thus we hypothesized that in *sgf73Δ* or *ubp8Δ* cells, transcriptional activity of the genes to which the proteins were proximally bound might be altered. We chose to look at the mRNA expression level of a subset of the genes that are occupied by both Sgf73 and Ubp8 in the null backgrounds in comparison to WT.

The genes *VPS62*, *FET3*, *AAC1*, *POM33*, *YPR036W-A*, and *HSP78* were chosen for analysis since they have the highest tag counts of the genes with duel occupancy. RT-qPCR analysis revealed that *VPS62*, *AAC1*, and *HSP78* had significantly increased expression in *sgf73Δ* cells in comparison to WT (Figure 25 a) ($n=3$, $p<0.05$). *FET3* and *YPR036W-A* had significantly increased expression levels in the *ubp8Δ* mutant in comparison to WT (Figure 3-10 a) ($n=3$, $p<0.05$).

Additionally we chose to look at the expression of a number of the ribosomal/translation related genes with duel occupancy: *EFB1*, *RPL33A, RPL6A, RPL28, RPL27B, RPL35A, RPL20A, RPS23B, RPS0B, RPL23B, RPS0A, RPS23A, RPL2A,* and *RPL42A*. We found that expression of *RPL27B, RPS23A, RPS23B, RPL23B, RPL2A, RPL42A,* and *RPL35A* was significantly reduced in *sgf73Δ* in comparison to WT, (Figure 3-10 b) ($n=6$, $p<0.05$). In the *ubp8Δ* mutant we found *RPL6A, RPL35A, RPS23B,* and *RPS23A* had significantly reduced expression while *RPL28,* and *RPL27* had significantly increased expression in comparison to WT (Figure 3-10 b) ($n=3$, $p<0.05$), (note that most changes represent reduced transcription, except for *RPL28, RPL27* which had extremely elevated expression levels in *ubp8Δ*). Additionally both null mutant strains trended toward having reduced expression in *RPL20A* (*sgf73Δ* $p=0.19$ and *ubp8Δ* $p=0.06$) in comparison to WT (Figure 3-10 b).

388 genes have Sgf73 occupancy with enrichment of the ribosomal pathway

To expand the scope of our data analysis we chose to study the occupancy peaks present in both Sgf73-13myc ChIP-seq replicates, as presence in both independent ChIP-seq experiments adds confidence to the binding site. Sgf73 binds DNA though its zinc finger domain [8] whereas there is no evidence for Ubp8 directly binding DNA. Closer contacts with the DNA may explain the difference between the number of targets for each protein of interest: Sgf73 yielded 389 direct targets and Ubp8 only 44. Significant Sgf73 peaks are listed in Supplementary Table 3-1. We looked for motif enrichment among the 389 peaks using the HOMER software findMotifsGenome.pl. The program generated a list of 17 motifs present among the 389 peaks, but listed 10 of them as possible false positives since their p value was <1e-12 (Figure 3-11). Of the 7 motifs predicted to be significant based on p-value, 3 are the known binding sequence of transcription factors which are involved in translation and ribosomal biogenesis. These factors are Sfp1, Hsf1, and Rlr1.

Using the Database for Annotation, Visualization and Integrated Discovery (DAVID)[30], we searched for pathway enrichment among the genes occupied by Sgf73 and found that three KEGG pathways were enriched: the ribosome (57 genes p=2.63E-33), and with lower confidence glycolysis/gluconeogenesis (9 genes p=5.9E-2), and the pathway for fructose and mannose metabolism (6 genes p= 8.2E-2). Most striking was the ribosomal pathway, (Table 3-3), with 57 genes in the ribosomal pathway representing ~14.5% of all genes occupied by Sgf73 in our study. Of these genes, 36 encode proteins of the large ribosomal subunit, 20 of the small ribosomal subunit, and 1 of the ribosomal stalk.

Many previous studies have linked ribosomal biogenesis and aging. In particular the reduction of large ribosomal subunit levels leads to increased RLS [31,32]. Our finding thus suggests that the increased RLS seen by the deletion of *SGF73* may occur through altered ribosomal biogenesis regulation upon alterations in Sgf73/SAGA occupancy.

We performed an independent genome ontology enrichment analysis on the 388 genes with 5' Sgf73 occupancy (Figure 3-12) to classify which gene ontology (GO) categories were overrepresented in our gene list. Again we see a strong role of ribosomal genes in all three main categories: biological process, molecular function, and cellular component. Additionally in the biological process category we see an enrichment of organic substance biosynthetic processes, including hexose biosynthesis, glycolysis, and gluconeogenesis and in the ion transport category there is an enrichment of copper ion transmembrane transport. In the molecular function category there is enrichment in genes encoding proteins involved in iron ion binding, genes with glycosyl and o-glycosyl hydrolase activity, and oxidoreductase activity.

31 genes with Sgf73 occupancy have extended RLS upon deletion

As previously discussed, the deletion of *SGF73* results in a drastic extension in RLS. We have made observations on what we believe is contributing to this increased RLS (chapter 2, and McCormick, Mason, & Guyenet et al submitted) but there are likely further connections to be found through exploring Sgf73 occupancy and all single gene deletions known to result in lifespan extension. Our hypothesis is that Sgf73 may be involved in the regulation of other genes known to cause increased RLS

upon deletion, thus contributing to the life extension seen when *SGF73* is deleted. The Kennedy lab has completed RLS analysis on all viable single deletion yeast strains, generating a list of 236 deletion mutants that have extended RLS (personal communication Kennedy lab). When we compare the genes with Sgf73 occupancy (based on the closest 5' TSS) and all genes that when deleted extend RLS, we find that there is an overlap of 31 genes (Figure 3-13 a, b), four of which also have Ubp8 occupancy (red boxes Figure 3-13 b). Of the 31 genes, 19 of them are RP genes, with 17 of those 19 encoding proteins of the large ribosomal subunit. Of all single deletion RLS-extending genes, 25 encode large ribosomal proteins, meaning that 68% of those genes have Sgf73 occupancy. This high level of overlap suggests that Sgf73 is involved in coordinating RP gene regulation and that this altered regulation contributes to lifespan extension in *sgf73Δ* cells.

RP genes with Sgf73 occupancy have reduced expression in *sgf73Δ* mutants

To explore if there was an altered transcriptional profile of RP genes that have Sgf73 occupancy in *sgf73Δ* mutants we selected 31 RP genes for expression analysis by RT-qPCR. This analysis included all 19 RLS-linked Sgf73-occupied RP genes. The expression of all RP genes explored trended towards down-regulation in the *sgf73Δ* mutant in comparison to WT (Figure 3-14 a, b). Of the 19 RLS-related RP genes, nine have significantly down-regulated expression: *RPL6B, RPL19B, RPL20B, RPL23A, RPL31A, RPL34B, RPL35A, RPL37B, RPL43B* (Figure 3-14a) ($p < 0.05$, t-test). An additional seven non-RLS RP genes including: *RPL27B*, *RPS23A, RPL23B, RPS23B, RPL2A, RPL42A, RPL10,* also have significantly down-regulated expression

(Figure 3-14 b) ($p < 0.05$, t-test). This transcriptional alteration further supports the evidence that RP gene transcription levels play a role in *sgf73Δ* mediated lifespan extension. The reduced level of multiple RP gene transcripts likely will have widespread effects on cell growth and metabolism, and is a documented response to various stimuli. For example, RP transcription is altered upon cellular stress (reviewed in [33]) thus in *sgf73Δ* mutants, their already reduced levels of RP gene transcription and the loss of regulation for many genes by Sgf73 may result in a defect in stress-induced RP transcriptional reduction.

There is a muted RP transcriptional response upon rapamycin treatment in *sgf73Δ* mutants

To test if a defect exists in RP gene translation down-regulation upon stress in *sgf73Δ* mutants, cells were treated with rapamycin. Rapamycin causes the inhibition of target of rapamycin (Tor) signaling, resulting in the repression of RP genes (reviewed in [33]). WT and *sgf73Δ* cells were treated with 200ng/ml rapamycin or ethanol as the vehicle control. The expression fold reduction in a number of RP genes was then evaluated via RT-qPCR.

The expression of all RP genes investigated did decrease upon rapamycin treatment, but there was a higher level of transcript, in a number of RP genes in the *sgf73Δ* mutant compared to WT (Figure 3-15 a, b). Strikingly, nine of the RLS-linked RP genes (47%) had significantly higher transcript levels in the *sgf73Δ* mutant (Figure 3-15 a). This is in contrast to the only two (16%) non RLS-linked RP genes (Figure 3-15 b). Thus despite lower RP levels in the *sgf73Δ* mutant in unchallenged conditions,

the expression of a number of RP genes is higher than WT when treated with rapamycin. We also looked at a number of non-RP genes that have Sgf73 occupancy and found that their expression did not follow the same general pattern observed with RP genes (Figure 3-15 c). This suggests that there is altered RP gene regulation under rapamycin stress conditions that is specific to this set of transcripts, in particular highlighting the RLS–linked RP genes.

There is no alteration in RP reduction upon MG132 treatment of *sgf73Δ*

MG132 is a proteasome inhibitor known to cause a reduction in the transcription of RP genes. To test whether there is an altered response to proteasome inhibition stress in *sgf73Δ* mutants, cells were treated for 20 and 45 minutes with 25μM MG132 or DMSO vehicle. All strains were deleted for *PDR5*, which encodes an ATP-binding cassette (ABC) transporter, to prevent drug efflux from the cells.

There does not appear to be a significant alteration in the response of RP genes or non-RP genes to MG132 in *sgf73Δ* mutants in comparison to WT cells (Figure 3-16 a, b). There is a trend towards *sgf73Δ* cells having a lower level of RP gene transcripts after 45 minutes of MG132 treatment than in WT cells. This suggests that *sgf73Δ* mutants may be able to reduce transcript levels to a greater extent upon proteasome inhibition, thus helping to cope with the accumulated protein in the cell.

Correlating Sgf73 occupancy with *sgf73Δ* and *ubp8Δ* microarray data

A series of microarray experiments were also performed to define gene expression patterns in the *sgf73Δ* and *ubp8Δ* strains in comparison to WT yeast.

Microarray data were filtered to produce lists of genes that were either up- or down-regulated by 1.5-fold or more. Based on these lists we were able to compare Sgf73 occupancy with genes that had differential expression upon *sgf73Δ* or *ubp8Δ* deletion. There was a significant overlap of 50 genes that had differential expression in the *sgf73Δ* strain that were also occupied by Sgf73 (hypogeometric test p (>=50)= 0.00013), with 13 having increased expression and 37 having decreased expression upon *SGF73* deletion (Figure 3-17 a, b). Surprisingly none of the Sgf73-occupied ribosomal protein genes had altered expression in the *sgf73Δ* microarray. Despite the lack of expression differences in RP genes found in the microarrays, we have seen with the previous RT-qPCR studies there does appear to be expression alterations in some of the Sgf73-occupied ribosomal protein genes. There are 15 genes with differential expression in the *ubp8Δ* microarray that had Sgf73 occupancy, 13 of them have increased expression and two have decreased expression (Figure 3-18 a, b). Ubp8 did not bind 5' to any of the genes that had altered expression in the *ubp8Δ* microarray. Comparing both microarray data sets and Sgf73 occupancy we see there is an overlap of 11 genes (Figure 3-19 a), with eight genes showing the same direction of expression change in both the *sgf73Δ* and *ubp8Δ* strains, and three genes having opposite expression changes in the two strains. In each of these three instances there was decreased expression in *sgf73Δ* and increased expression in *ubp8Δ* (Figure 3-19 b). Of the 11 overlapping genes, four code for proteins with roles in the mitochondria.

Connecting Sgf73 occupancy, expression, and RLS data

From our connections of Sgf73 occupancy to deletions resulting in increased RLS and Sgf73 occupancy to *sgf73Δ* and *ubp8Δ* microarray data, we decided to connect all four data sets to find targets that likely had a role in Sgf73-mediated life span extension (Figure 3-20). Two genes were shared among all four data sets. The first *FMP43* encodes a mitochondrial pyruvate carrier whose increases upon DNA replication stress[34], had decreased expression in both microarrays. The second *IDH1* encodes a subunit of a mitochondrial NAD(+)-dependent isocitrate dehydrogenase [35], its expression was increased in both microarrays. Two additional genes were found to be shared between the Sgf73 occupancy, *sgf73Δ* microarray, and increased RLS data sets: *MNN1*, encoding an alpha-1,3-mannosyltransferase [36-38], and *YBR238C*, encoding a mitochondrial membrane protein [39]. Both had decreased expression in the *sgf73Δ* microarray.

We chose to evaluate the Sgf73 occupancy/binding to the four overlapping genes via ChIP-PCR (Figure 3-21). We observed that all were enriched in both Sgf73-13myc and Ubp8-13myc ChIP. We also looked at the sequencing peaks for these four targets using the UCSC genome browser to visualize sequence runs for the Sgf73-13myc ChIPs (Figure 34 a-d). *PHO89*, encoding a sodium (Na+) and phosphate (Pi) cotransporter [40], also became of interest since it is the only additional gene with decreased expression in both microarrays having Sgf73 occupancy. The subset of five genes were analyzed using RT-qPCR (Figure 3-22). Significant reductions in *FMP43*, *MNN1*, *YBR238C*, and *PHO89* expression were seen in *sgf73Δ* ($p<0.05$), similar to the microarray. *FMP43* also has a paralog *MPC2* that has Sgf73 occupancy and has

decreased expression in *sgf73Δ* mutants (Figure 3-22). Significant changes in *IDH1* could not be confirmed in a statistically significant manner, although they trended towards increased expression. In *ubp8Δ* a significant expression decrease was seen for *FMP43,* and a significant increase for *IDH1*. The protein levels of epitope tagged *FMP43*, *IDH1*, *MNN1*, and *MPC2* were also analyzed by immunoblot in *sgf73Δ* mutants (Figure 3-23 a, b, c, d). All proteins showed the same trends as seen at the transcript level by RT-qPCR.

Limited Sir2 occupancy in *sgf73Δ* resembles Sir2 occupancy in *SIR2* overexpressing strain

Sequencing data from our Sir2 ChIP-seq experiments was not optimal. There was very low sequence alignment from the raw sequence files resulting in a low predicted IP efficiency. Additionally previous studies have indicated that silenced chromatin is less amenable to sheering in ChIP, resulting in an underrepresentation of these silenced regions[41]. As Sir2 is known to be involved in chromatin silencing this phenomenon my also account for our low IP efficiency. Furthermore a high level of background alignment from the *sir2Δfob1Δ* ChIP-seq runs filtered out most of the sequencing peaks. Alternatively, Sir2 may not have many DNA-binding activities or DNA targets, and therefore we would not expect to have a high number of sequencing peaks. After filtering with the control strain as well as using standard parameters we were left with 10 significant peaks in the WT strain, 10 significant peaks in the *SIR2* overexpressing (Sir2OE) strain, and 9 significant peaks in the *sgf73Δ* strain (Supplementary Table 3-3). These peaks were calculated based on two combined

replicates for each strain and do not represent peaks present in both sequencing runs for each strain as this would further limit our results. An encouraging finding from these limited data is that Sir2 occupancy in the Sir2OE strains overlaps significantly with the Sir2 occupancy in *sgf73Δ,* sharing six out of 10 binding targets/peaks, whereas in WT only one target/peak is shared. This suggests Sir2 occupancy in *sgf73Δ* resembles that in the Sir2OE strain. With further examination of the peaks we do find two peaks that have Sir2 occupancy in WT yeast and also have Sgf73 occupancy. One is *MUP1*, encoding a high affinity methionine permease membrane protein involved in cysteine uptake [42,43], and the other is *MNN1* already a key target due to its deletion resulting in RLS extension and confirmed decreased expression in *sgf73Δ* yeast (Figure 3-22). Another interesting peak in the WT Sir2 ChIP data is 5' of the gene *FTR1*, encoding a high affinity iron permease which forms a complex with Fet3, a ferro-O_2-oxidoreductase, that acts as a cell-surface high affinity iron uptake system for the cell [44-48]. *FET3* has 5' occupancy by both Sgf73 and Ubp8 in our studies, and has increased RLS upon deletion. Both *FTR1* and *FET3* have increased expression upon DNA replication stress [34], and are regulated by Aft2, an enriched binding motif from our Sgf73 ChIP-seq (Figure 3-11). Thus, Sgf73 might be involved in regulating iron transport during DNA replication stress.

Sgf73-occupied genes with Sir2 dependent and independent expression changes

We were interested in defining the relationship between both Sir2 dependent and Sir2 independent gene expression changes upon *SGF73* deletion and Sgf73 occupancy (Figure 3-24). If *SGF73* deletion still causes expression changes in the

sir2Δ fob1Δ background, the genes have expression changes that are likely independent of Sir2. Conversely, if gene expression differences are only seen with *SGF73* deletion in the presence or absence of Sir2, then their expression changes may be dependent on Sir2 (of note is that we cannot distinguish between *SIR2* and *FOB1* dependent/independent expression changes but will refer to them as Sir2 dependent/independent changes.) Based on microarray data that has been filtered to have a fold change of 1.5 or greater from the strains *sgf73Δ* vs. WT and *sgf73Δ sir2Δ fob1Δ* vs. *sir2Δ fob1Δ* we find that 22 genes have Sgf73 occupancy and Sir2 independent expression changes upon *SGF73* deletion. 28 genes have Sgf73 occupancy and only change upon *SGF73* deletion if Sir2 is present. Seven genes have Sgf73 occupancy and only change upon *SGF73* deletion if Sir2 is not present. A total of 35 genes have possible Sir2-dependent expression changes upon *SGF73* deletion (Figure 3-24, Table 3-4).

We examined the *ubp8*Δ microarray data. If *UBP8* deletion also causes expression changes in the *sir2Δ fob1Δ* background, the genes have expression changes that are likely independent of Sir2. Conversely, if gene expression differences are only seen with *UBP8* deletion in the presence or absence of Sir2 then their expression changes may be dependent on Sir2. To explore Sir2-connected gene expression changes that have Sgf73 occupancy we evaluated microarray data from *ubp8Δ* vs. WT and *ubp8Δ sir2Δ fob1Δ* vs. *sir2Δ fob1Δ* along with Sgf73 occupancy data (Figure 3-25). Seven genes have Sgf73 occupancy and Sir2 independent expression changes in *ubp8Δ*; eight genes have Sgf73 occupancy and only change in *ubp8Δ* if Sir2 is present; and seven genes have Sgf73 occupancy and only change upon in *ubp8Δ* if Sir2 is not

present. Thus, there is a total of 15 genes that have potential Sir2 dependent expression changes in *ubp8Δ* (Table 3-5).

By combining 5 data sets (Sgf73 occupancy, *sgf73Δ* vs. WT microarray, *sgf73Δ sir2Δ fob1Δ* vs. *sir2Δ fob1Δ* microarray, *ubp8Δ* vs. WT microarray, *ubp8 Δsir2Δ fob1Δ* vs. *sir2Δ fob1Δ* microarray) we sought to identify genes whose expression was dependent on Sir2 or independent of Sir2 presence in both the *sgf73Δ* and *ubp8Δ* microarrays, and had 5' Sgf73 occupancy (Figure 3-26). Six genes have Sgf73 occupancy and conferred an expression change in the deletion strains regardless of the strains' Sir2 status, rendering them Sir2-independent (Table 3-6). Three genes have Sgf73 occupancy and change upon *SGF73* or *UBP8* deletion only if Sir2 is present. We did not identify any genes that have Sgf73 occupancy and only change upon *SGF73* or *UBP8* deletion if Sir2 is not present, therefore a total of three genes have possible Sir2-dependent expression changes in *sgf73Δ* and *ubp8Δ* (Table 3-6).

Shared expression changes in *SIR2*OE, *sgf73Δ*, and *ubp8Δ* lack Sgf73 occupancy

Based on strain phenotypes it is hypothesized that the deletion on *SGF73* and the over-expression of *SIR2* result in similar dysregulation of gene expression. This was investigated by exploring the overlap of microarray data from *sgf73Δ* vs. WT and *SIR2*OE vs. WT (only 20 differentially expressed genes). We find that the two data sets share 13 genes (Figure 3-27 a, b). In addition we evaluated gene expression alterations that were shared between the *ubp8Δ* vs. WT and *SIR2*OE vs. WT microarrays, and found 10 genes shared between the two data sets (Figure 3-27 c, d). Combining all three data sets we find that 9 genes have expression changes in all three

conditions (Figure 3-27 e, f). However, when we incorporate the Sgf73 occupancy data we do not find any genes that have Sgf73 occupancy in addition to differential expression in the *SIR2*OE vs. WT microarray.

Extrapolating findings to SCA7

A goal of studying *SGF73*, *UBP8*, and *SIR2* in yeast is to learn more about the function of the SAGA complex and importantly, help identify disease mechanisms related to SCA7 and the human STAGA complex. From the yeast occupancy and expression studies we hypothesized that dysregulation of candidate ribosomal protein genes and mitochondrial related genes might be conserved. A number of the top candidates of interest in yeast have human orthologs (Table 3-7), thus we examined the RNA expression levels of these genes from SCA7 patient and control fibroblast cells (gift from J. Ward) via RT-qPCR. We found transcriptional dysregulation of some of the candidate genes in patient fibroblast cells vs. unaffected control fibroblast cells (Figure 3-28). These results are encouraging and suggest that similar SAGA and STAGA targets are disrupted upon loss of of Sgf73 or when ataxin-7 has an expanded poly-Q track. More experiments are needed to fully assess the extent of dysregulation for each chosen target and its effects on SCA7 disease pathology or progression.

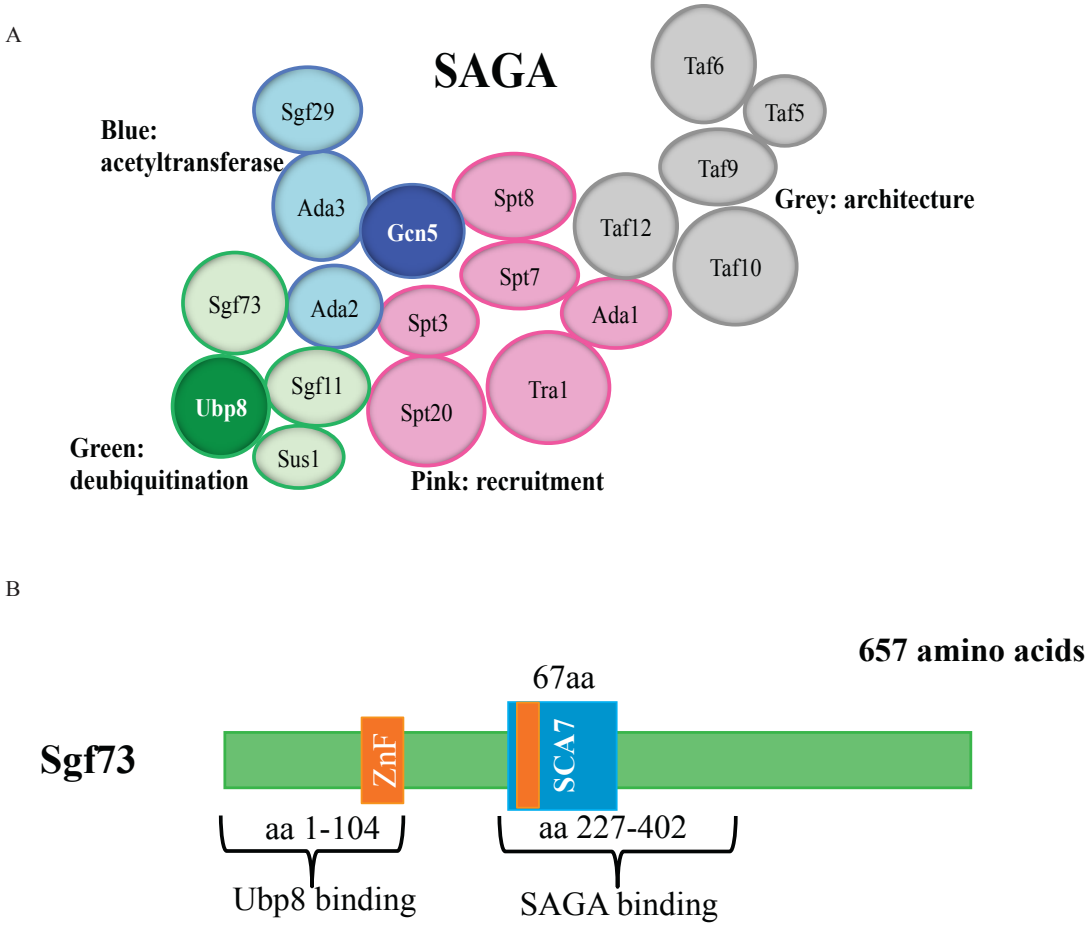


Figure 3-1: SAGA and Sgf73 schematic
A) Depiction of the SAGA complex subunits in their respective modules, indicated by color-coding. B) Schematic of Sgf73 depicting the N-terminal zinc-finger binding domain, the conserved SCA7 box domain containing the second zinc-finger binding domain. Regions known to be essential for Sgf73 binding to Ubp8 and SAGA are indicated.

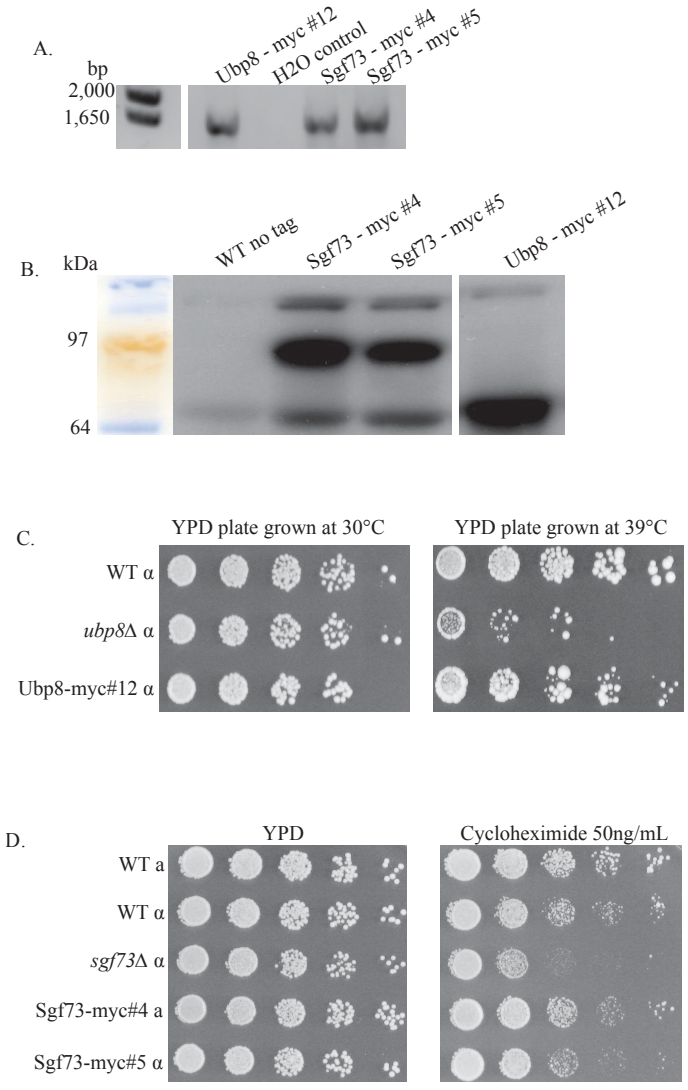


Figure 3-2: Tagging and expression of Sgf73-myc and Ubp8-myc
A) PCR of the genomic DNA from Sgf73-myc and Ubp8-myc, forward primer recognizes upstream of integrated host genomic DNA sequence, reverse primer recognizes the *kanMX* selective sequence present after the myc tag. B) Immunoblot of whole cell protein lysate from Sgf73-myc and Ubp8-myc strains, blot is probed with anti-myc. C) Dilution heat sensitivity assay on Ubp8-myc strain, control plate grown at 30°C and experimental plate at 39°C. Ubp8-myc strain does not show heat sensitivity. D) Dilution cycloheximide sensitivity assay on Sgf73-myc strains. Sgf73-myc tagged strains are not cycloheximide sensitive.

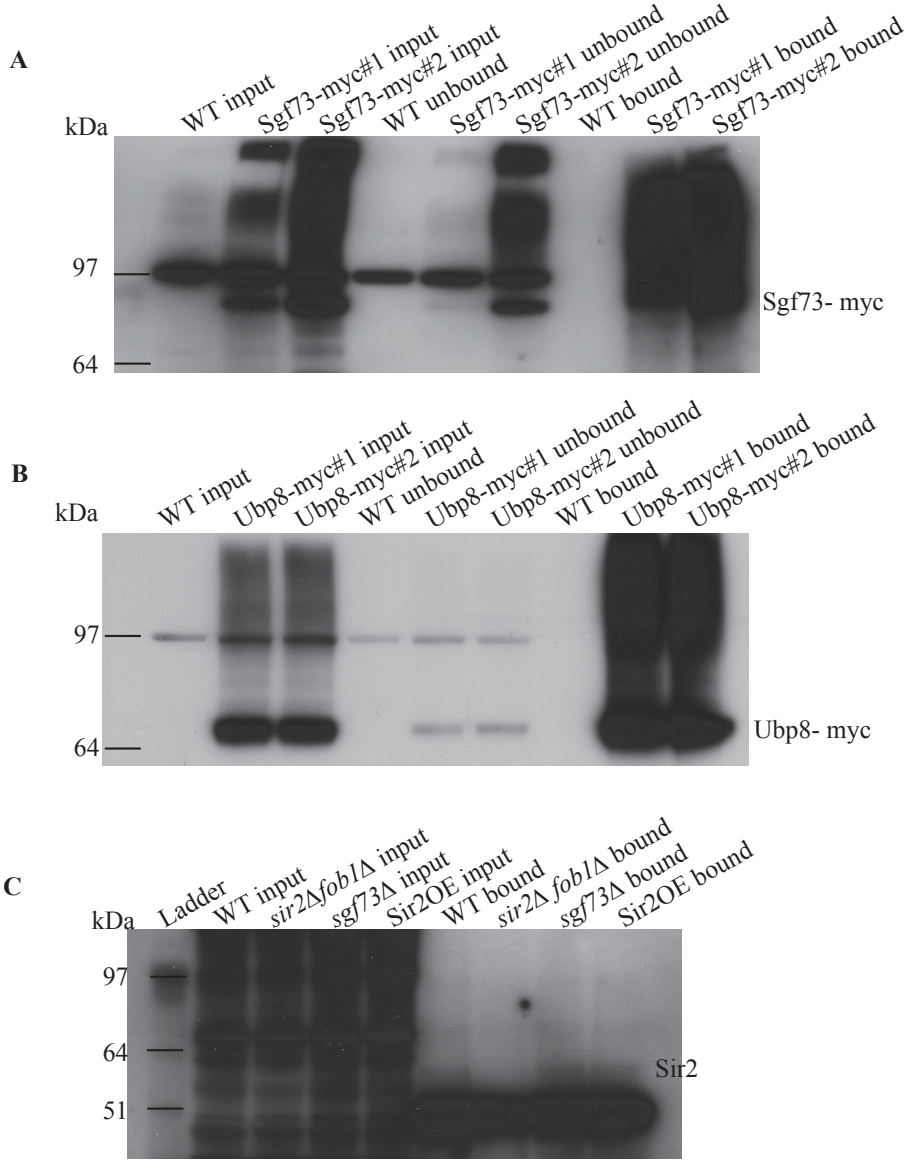


Figure 3-3: Immunoblots of ChIP bait

One percent of the input, unbound, and bound ChIP fractions was used in each immunoblot for: A) Both Sgf73-myc ChIP replicates. Gel was probed with anti-myc. B) Both Ubp8-myc ChIP replicates. Gel was probed with anti-myc. C) One Sir2 ChIP. Note due to the increased number of strains being ChIPed the unbound protein was not included. Only input and bound samples were loaded, gel was probed with anti-Sir2 antibody.

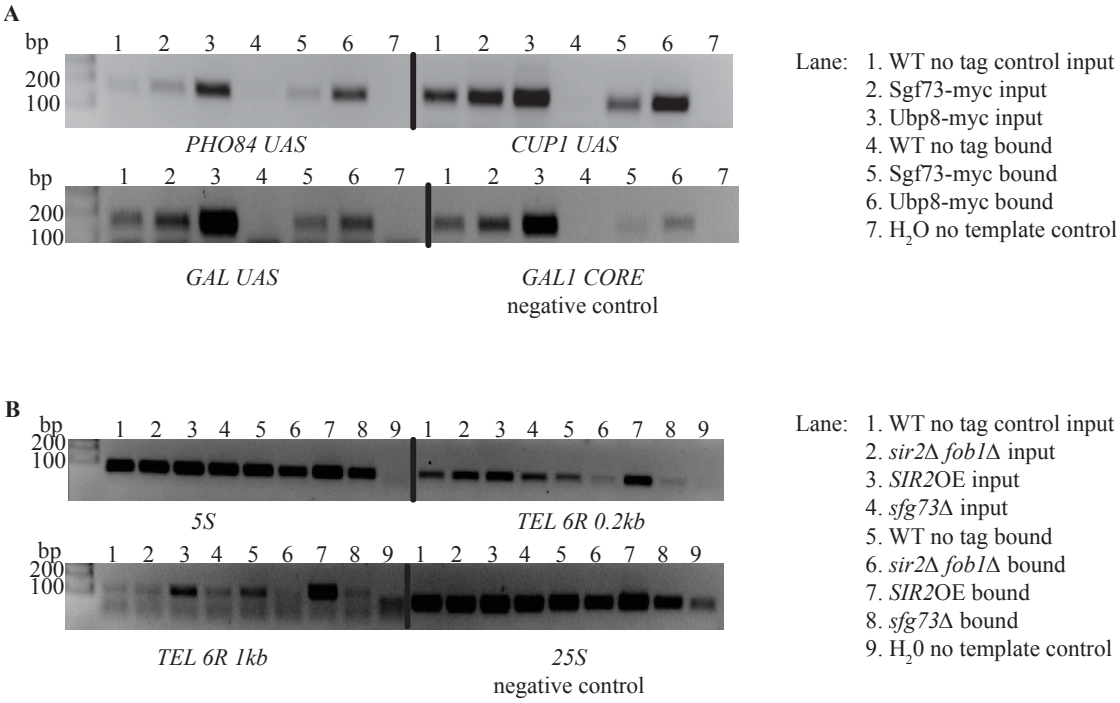


Figure 3-4: ChIP-PCR of established SAGA and Sir2 targets
A) PCR reactions on myc-ChIPed samples: input reaction is on DNA post sonication but pre-ChIP, bound PCR samples were preformed on DNA post-ChIP B) PCR reactions on Sir2-ChIPed samples: input reaction is on DNA post sonication but pre-ChIP, bound PCR samples were preformed on DNA post-ChIP.

Table 3-1: Sequencing library information. List of strains and antibodies used in ChIPs for preparing sequencing libraries and the resulting file names.

Library #	Strain	IP	Replicate	Barcode	Sequencing file names, top: BGI file name bottom: Converted file name
1	WT	αMyc	1	L1 - AAGGGA	1_FCD1UP0ACXX_L8_CHKSE13040343_1.fq.gz 1_LaSpada AGM-052513-WTScer-MycChIP 1.fq.gz
2	WT	αMyc	2	L2- CCTTCA	2_FCD1UP0ACXX_L8_CHKSE13040344_1.fq.gz 2_LaSpada AGM-052513-WTScer-MycChIP 2.fq.gz
3	Sgf73-13myc	αMyc	1	L3- GGACCC	3_FCD1UP0ACXX_L8_CHKSE13040345_1.fq.gz 3_LaSpada AGM-052513-Sgf73MycScer-MycChIP1.fq.gz
4	Sgf73-13myc	αMyc	2	L4- TTCAGC	4_FCD1UP0ACXX_L8_CHKSE13040346_1.fq.gz 4_LaSpada AGM-052513-Sgf73MycScer-MycChIP 2.fq.gz
5	Ubp8-13myc	αMyc	1	L5- AAGACG	5_FCD1UP0ACXX_L8_CHKSE13040347_1.fq.gz 5_LaSpada AGM-052513-Ubp8MycScer-MycChIP 1.fq
6	Ubp8-13myc	αMyc	2	L6- CCTCGG	6_FCD1UP0ACXX_L8_CHKSE13040348_1.fq.gz 6_LaSpada AGM-052513-Ubp8MycScer-MycChIP 2.fq
7	WT	αSir2	1	L7- GGATGT	7_FCD1UP0ACXX_L8_CHKSE13040349_1.fq.gz 7_LaSpada AGM-052513-WTScer-Sir2ChIP 1.fq
8	WT	αSir2	2	L8- TTCGCT	8_FCC1U29ACXX_L8_CHKSE13040350_1.fq.gz 8_LaSpada AGM-052513-WTScer-Sir2ChIP 2.fq
9	*sir2Δ fob1Δ*	αSir2	1	L1 - AAGGGA	9_FCC1U29ACXX_L8_CHKSE13040351_1.fq.gz 9_LaSpada AGM-052513-sir2nullfob1nullScer-Sir2ChIP_1.fq
10	*sir2Δ fob1Δ*	αSir2	2	L2- CCTTCA	10_FCC1U29ACXX_L8_CHKSE13040352_1.fq.gz 10_LaSpada AGM-052513-sir2nullfob1nullScer-Sir2ChIP 2.fq
11	*sgf73Δ*	αSir2	1	L3- GGACCC	11_FCC1U29ACXX_L8_CHKSE13040353_1.fq.gz 11_LaSpada AGM-052513-sgf73nullScer-Sir2ChIP 1.fq
12	*sgf73Δ*	αSir2	2	L4- TTCAGC	12_FCC1U29ACXX_L8_CHKSE13040354_1.fq.gz 12_LaSpada AGM-052513-sgf73nullScer-Sir2ChIP 2.fq
13	Sir2OE	αSir2	1	L5- AAGACG	13_FCC1U29ACXX_L8_CHKSE13040355_1.fq.gz 13_LaSpada AGM-052513-Sir2OEScer-Sir2ChIP 1.fq
14	Sir2OE	αSir2	2	L6- CCTCGG	14_FCC1U29ACXX_L8_CHKSE13040356_1.fq.gz 14_LaSpada AGM-052513-Sir2OEScer-Sir2ChIP 2.fq

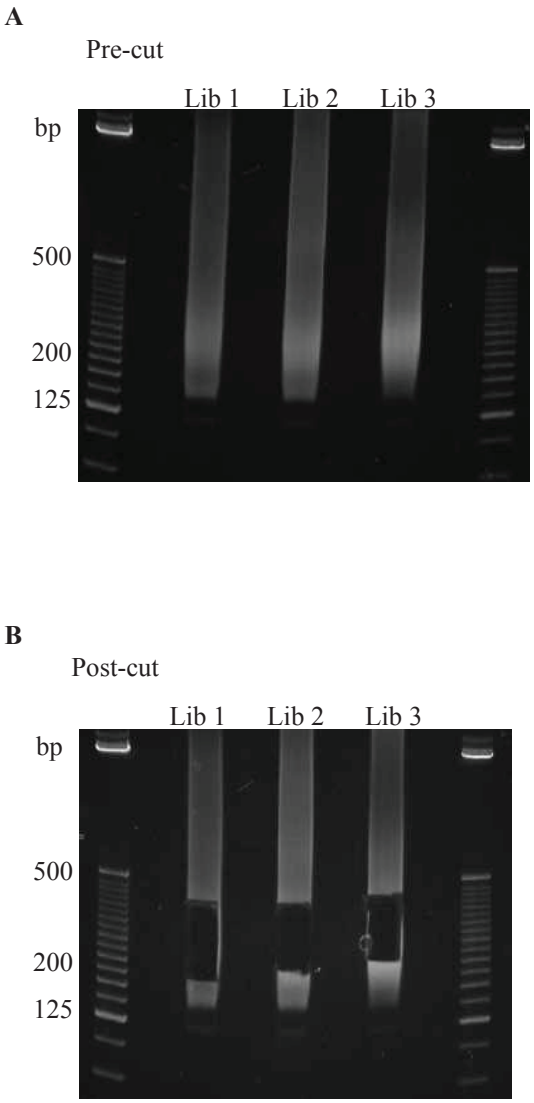


Figure 3-5: Sequencing library gel size selection
A) Libraries were loaded on 6% acrylamide gels following adaptor ligation and PCR amplification. Note enrichment of fragments between 250 and 300bp B) Image depicts gel post-cut for library size selection of fragments between ~150 and 400bp

Table 3-2: Significant peaks with both Sgf73 and Ubp8 occupancy. Table lists the chromosome number and start - stop point of each significant peak. The peak score refers to the number of uniquely mapped reads to that peak location. The table also lists the approximate distance to the nearest 5' and/or 3' gene.

Chr #	Start	End	Peak Score	Position #1	Gene #1	Position #2	Gene #2
chr16	645466	645572	7609.93	300bp 5'	YPR036W-A		
chr07	772110	772215	7363.08	250bp 5'	BTN2	150bp 3'	VPS62
chr13	388445	388570	7275.38	200bp 5'	FET3	200bp 5'	AAC1
chr01	141717	141822	7154.03	400bp 5'	EFB1	250bp 5'	SSA1
chr12	97654	97759	6956.17	200bp 5'	SSA2	250bp 3'	POM33
chr14	619830	619947	6429.64	100bp 5'	LST8	250bp 5'	SIS1
chr07	371319	371436	6344.09	in/5'	YGL072C	75bp 3'	HSF1
chr15	83242	83353	6246.09	300bp 5'	MDH2	450bp5'	TRM13
chr07	483476	483617	6205.88	in	BRP1	400bp 5'	YGL007C-A
chr07	914973	915080	6172.74	150bp 5'	ZPR1	250bp 5'	YGR210C
chr04	1490154	1490260	6063.54	150bp5'	YDR524C-A	in	API2
chr07	310724	310827	5664.14	150bp 5'	RPL28		
chr06	221770	221908	5579.99	350bp 5'	RPL2A	200bp 3'	RRT5
chr07	609944	610075	5574.8	450bp 5'	ERG25	250bp3'	snR48
chr11	518636	518759	5488.31	in	YKR014W	100bp5'	YKR040C
chr12	368378	368559	5372.52	225bp 5'	AHP1	250bp 5'	YLR108C
chr16	785825	785925	5307.68	275bp 5'	CTR1		
chr02	444794	444896	5191.48	100bp 5'	FES1	150bp 3'	EXO84
chr15	619333	619436	5141.47	400bp 5'	PDR5		
chr14	739439	739543	5015.02	400bp 5'	FRE4		
chr11	258357	258497	4693.92	3'	YKL097C		
chr04	974378	974511	4533.99	100bp 5'	HSP78	125bp 3'	YAP6
chr07	726624	726768	4435.94	225bp 5'	RPS23A	150bp5'	YGR117C
chr13	124351	124475	4433.71	150bp 5'	RPL6A	200bp 3'	TCB3
chr11	99842	99989	4408.93	200bp3'	LOT5		
chr12	941024	941130	4299.36	100bp 5'	YLRWTy2-1	400bp 5'	YLR410W-A and -B
chr16	794681	794784	4290.31	150bp5'	RPS23B	150bp 5'	NAT3
chr12	282717	282849	4274.88	75bp 5'	RPL10	250bp 5'	BUD20
chr10	651432	651537	4153.98	350bp 5'	RPS5	150bp3'	IBA57
chr15	109732	109929	4146.57	in	Shr5	400bp 5'	ZEO1
chr12	241888	241995	4144.37	200bp 5'	RPS0B		
chr16	281811	281913	4136.48	200bp 5'	RPL33A	250bp5'	snR17b
chr13	754426	754537	3973.01	200bp 5'	RPL20A	50bp3'	ZRC1, YMR242WA
chr07	920094	920238	3918.71	300bp 5'	RPS0A		
chr05	396424	396541	3787.9	200bp5'	SLX8	200bp 5'	RPL23B
chr08	35953	36069	3757.66	in/5'	RPL8A	325bp 3'	Gut1
chr04	117378	117532	3754.12	300bp5'	RPL35A		
chr14	331027	331136	3747.99	200bp5'	RPL42A		
chr04	1401462	1401583	3704.85	150bp5'	RPL27B	200bp 5'	UGO1
chr04	976569	976699	3619.74	25bp3'	SWM1		
chr07	398028	398167	3026.78	400bp 5'	OLE1	350bp 5'	SDS23
chr11	533361	533495	2781.82	in/5'	MRS4		

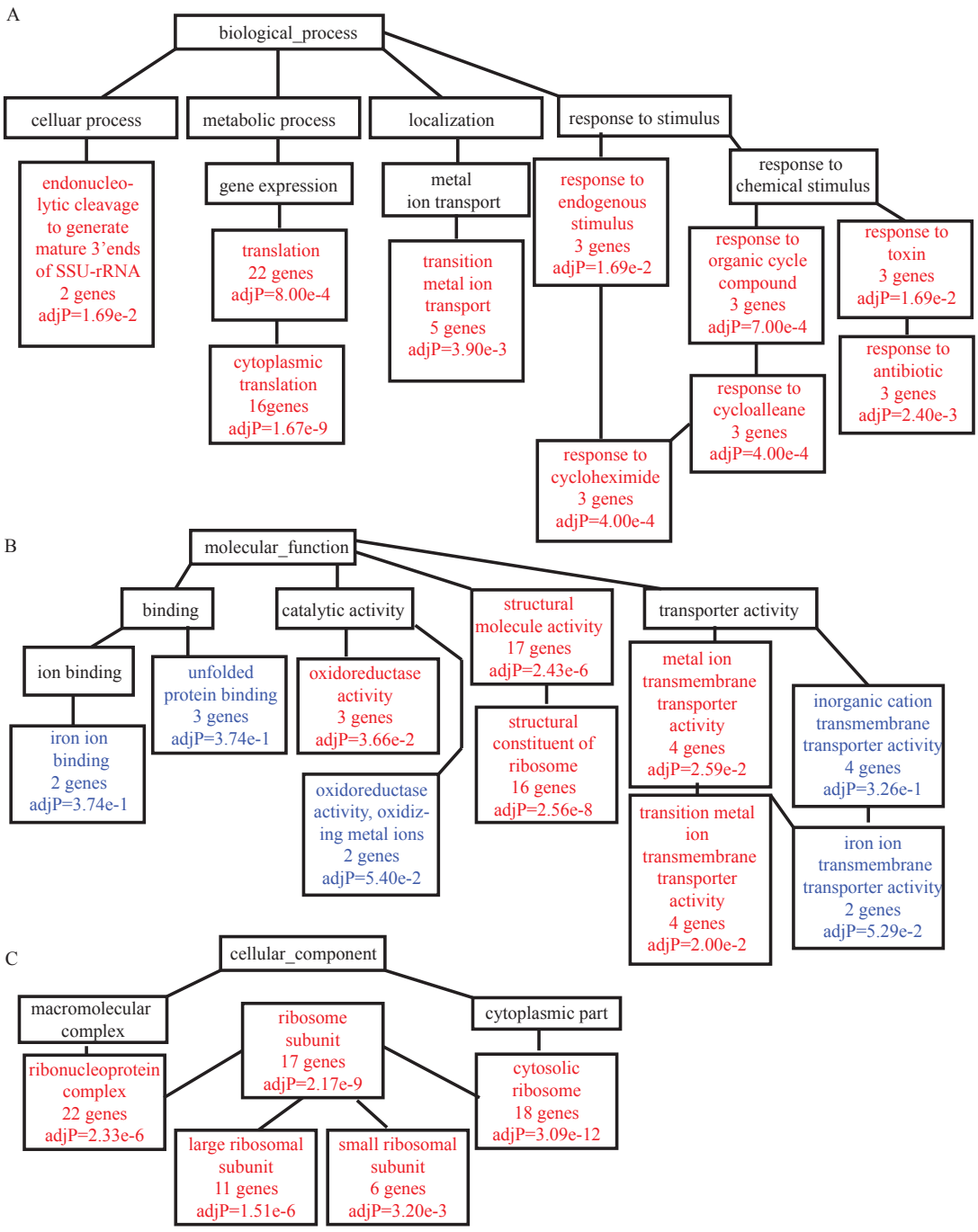


Figure 3-6: Sgf73 and Ubp8 occupied genes GO term analysis
Each box represents a category in the gene ontology cataloging within the broader category of A) biological process, B) molecular function, and C) cellular component. If a box has red text that category was significantly enriched, in the Sgf73 and Ubp8 occupied gene set, meaning more genes from the data set fell into that category than would be expected by random chance. Blue text categories are less significant. In significant categories the number of genes from the data set in that category is listed along with the p-value.

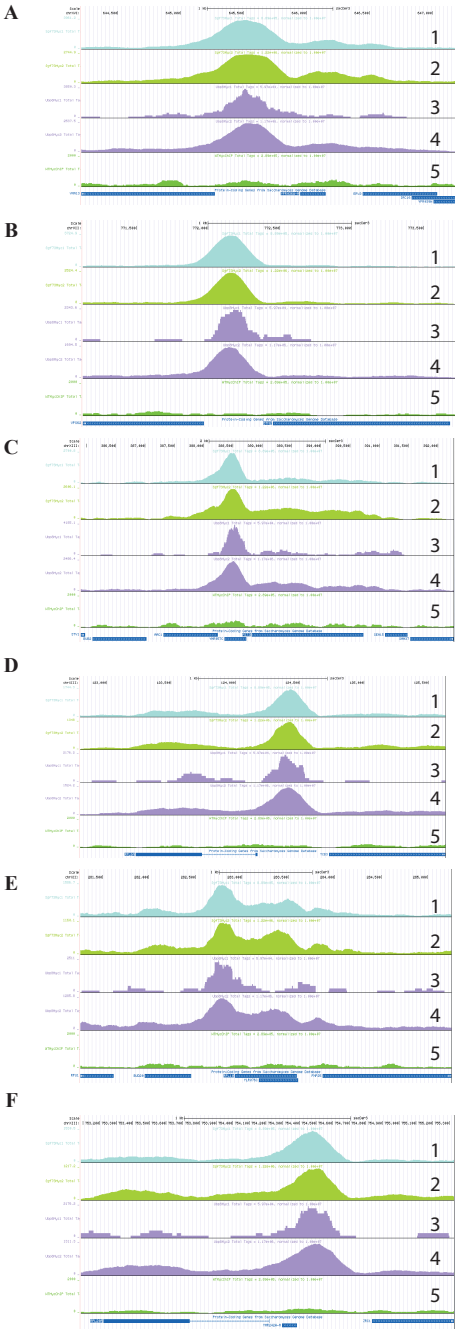


Figure 3-7: Targets-of-interest peak visualization on the UCSC genome browser
Each track depicts one sequencing run, track 1 and 2 are replicates 1 and 2 of Sgf73-myc ChIP, tracks 3 and 4 are replicates 1 and 2 of Ubp8-myc ChIP, the bottom track 5 is the WT no tag control representing background reads. A) 5' *YPR036W-A* B) 5' *BTN2* and 3'*VPS62* C) 5' *FET3* and *AAC1* D) 5' *RPL6A* E) 5' *RPL10* F) 5' *RPL20A*

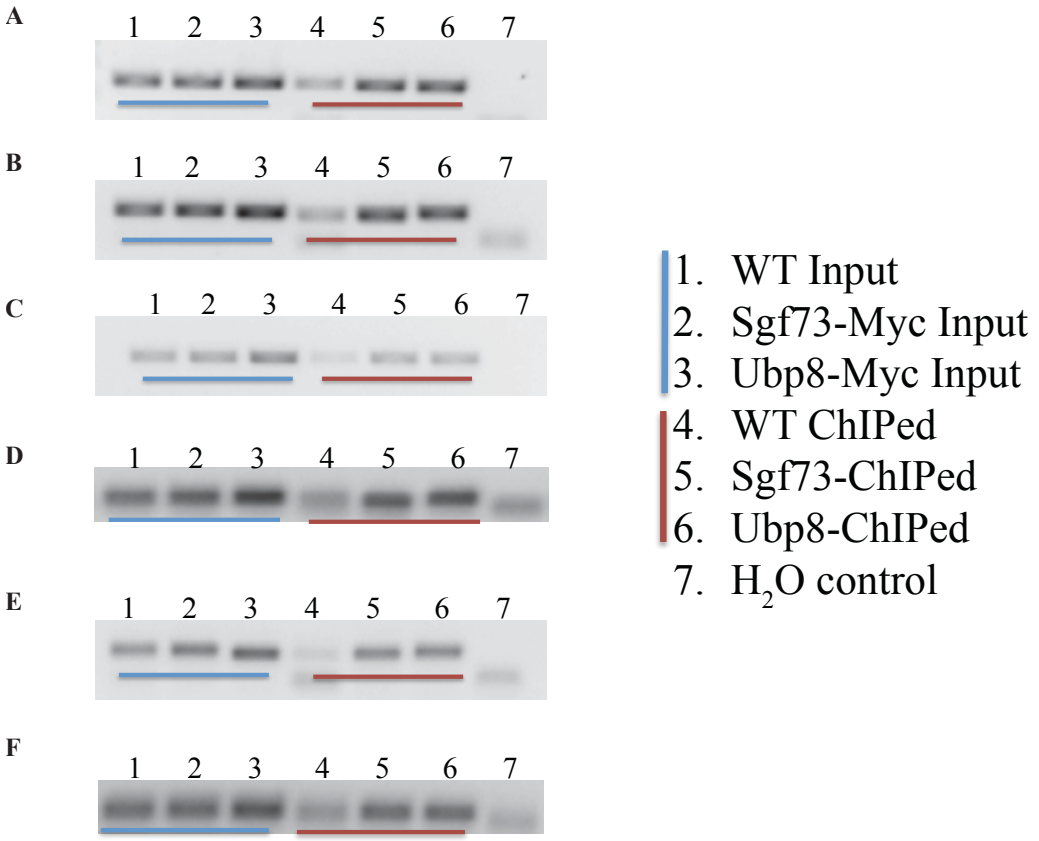


Figure 3-8: Target peak validation determined by ChIP-PCR
PCR amplifies the peak region identified by ChIP-seq A) 5' *YPR036W-A* B) 5' *BTN2* and 3'*VPS62* C) 5' *FET3* and *AAC1* D) 5' *RPL6A* E) 5' *RPL10* F) 5' *RPL20A*

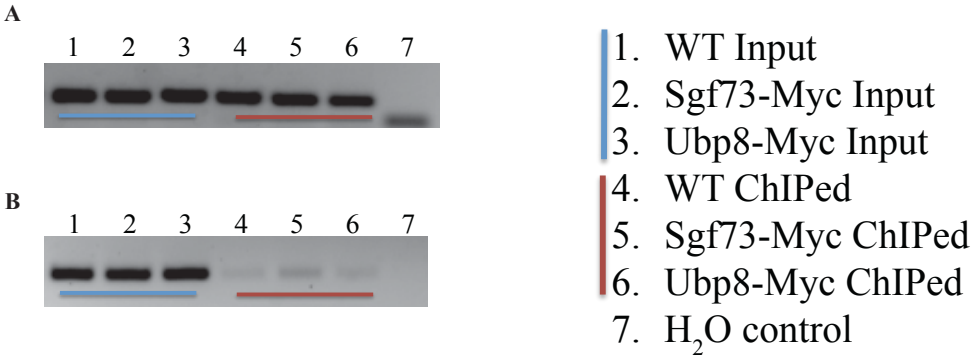


Figure 3-9: ChIP PCR of non-ChIPed genomic regions
Regions that did not have significant peaks by ChIP-seq were chosen for off target ChIP-PCR evaluation A) chrVII region from bases 334116 - 334213 B) chrXII region from bases 324679 - 324875

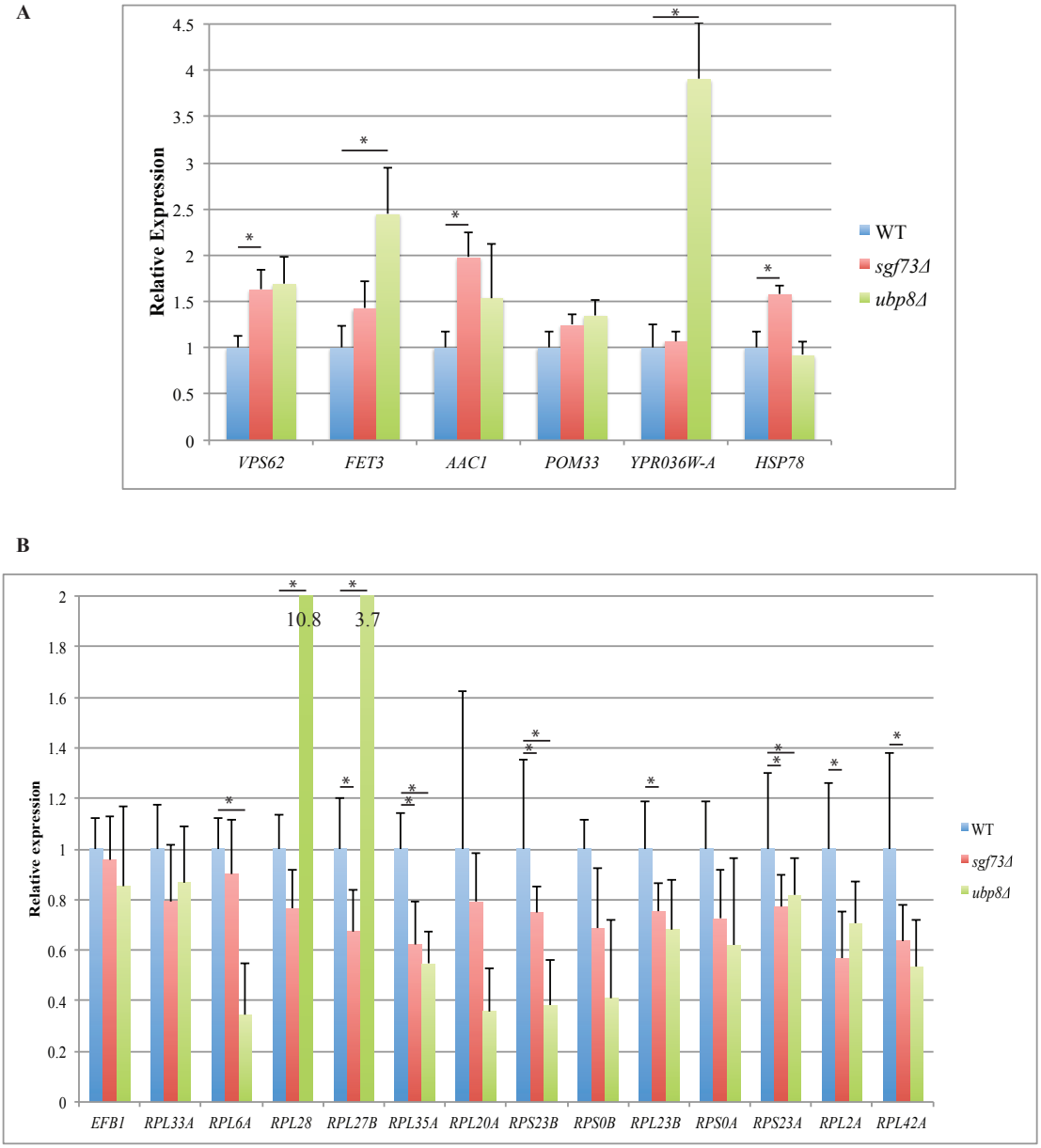


Figure 3-10: Profiling expression changes of Sgf73 and Ubp8 occupied genes Using qRT-PCR we looked at a number of Sgf73 and Ubp8 occupied target genes in *sgf73Δ* and *ubp8Δ* mutants. All values were first normalized to the internal control *SCR1* and then to WT, * represents a significant p-value <0.05. A) Expression level of genes with the highest peak score. B) Expression level of ribosomal/ transcription related genes.

* - possible false positive

Rank	Motif	P-value	log P-pvalue	% of Targets	% of Background	STD(Bg STD)	Best Match/Details
1		1e-19	-4.567e+01	53.98%	31.25%	24.8bp (45.7bp)	RGT1(MacIsaac)/Yeast More Information | Similar Motifs Found
2		1e-18	-4.235e+01	40.10%	20.30%	25.2bp (56.8bp)	MA0374.1_RSC3/Jaspar More Information | Similar Motifs Found
3		1e-17	-3.966e+01	6.94%	0.74%	26.8bp (53.9bp)	SFP1(MacIsaac)/Yeast More Information | Similar Motifs Found
4		1e-16	-3.825e+01	4.63%	0.25%	22.2bp (38.8bp)	HSF1(MacIsaac)/Yeast More Information | Similar Motifs Found
5		1e-13	-3.210e+01	9.77%	2.11%	23.8bp (47.9bp)	SWI4(MacIsaac)/Yeast More Information | Similar Motifs Found
6		1e-13	-3.155e+01	21.59%	8.86%	27.9bp (46.7bp)	RLR1?/SacCer-Promoters/Homer More Information | Similar Motifs Found
7		1e-13	-3.017e+01	8.74%	1.80%	21.1bp (53.7bp)	SWI6(MacIsaac)/Yeast More Information | Similar Motifs Found
8 *		1e-11	-2.669e+01	5.91%	0.90%	22.9bp (38.2bp)	NDD1(MacIsaac)/Yeast More Information | Similar Motifs Found
9 *		1e-11	-2.555e+01	19.54%	8.50%	27.5bp (50.3bp)	MA0270.1_AFT2/Jaspar More Information | Similar Motifs Found
10 *		1e-9	-2.089e+01	6.17%	1.32%	26.7bp (36.6bp)	MA0266.1_ABF2/Jaspar More Information | Similar Motifs Found
11 *		1e-8	-2.024e+01	1.80%	0.05%	26.6bp (30.1bp)	MA0393.1_STE12/Jaspar More Information | Similar Motifs Found
12 *		1e-8	-2.023e+01	9.51%	3.04%	23.9bp (45.2bp)	MA0325.1_LYS14/Jaspar More Information | Similar Motifs Found
13 *		1e-8	-1.918e+01	1.80%	0.06%	34.9bp (19.6bp)	MA0403.1_TBF1/Jaspar More Information | Similar Motifs Found
14 *		1e-8	-1.878e+01	2.31%	0.14%	21.4bp (49.0bp)	MA0273.1_ARO80/Jaspar More Information | Similar Motifs Found
15 *		1e-6	-1.553e+01	2.57%	0.28%	27.7bp (32.3bp)	RLM1(MacIsaac)/Yeast More Information | Similar Motifs Found
16 *		1e-5	-1.333e+01	0.77%	0.01%	32.1bp (8.3bp)	SFL1(MacIsaac)/Yeast More Information | Similar Motifs Found
17 *		1e-4	-9.234e+00	1.80%	0.27%	20.4bp (43.8bp)	REB1/SacCer-Promoters/Homer More Information | Similar Motifs Found

Figure 3-11: Homer output of de novo motifs found from analysis of Sgf73 occupancy peaks. Each sequence corresponds to the known binding motif of the protein listed in the far right column. These motifs are enriched in the Sgf73-myc ChIPed sequences to varying degrees of significance. Motifs with a * may be false positives due to a strict significance p-value cutoff of <1e-12 to be considered a positively enriched motif[25].

Table 3-3: Genes in the KEGG ribosome pathway with Sgf73 occupancy

Ensembl Gene Stable ID	Entrez ID	Gene Symbol	Protein Name
YBL027W	852254	*RPL19B*	60S ribosomal protein L19
YBL087C	852191	*RPL23A*	60S ribosomal protein L23
YBR048W	852337	*RPS11B*	40S ribosomal protein S11
YBR084C-A	852379	*RPL19A*	60S ribosomal protein L19
YBR189W	852487	*RPS9B*	40S ribosomal protein S9-B
YCR031C	850397	*RPS14A*	40S ribosomal protein S14-A
YDL075W	851484	*RPL31A*	60S ribosomal protein L31-A
YDL083C	851476	*RPS16B*	40S ribosomal protein S16
YDL133C-A	851422	*RPL41B*	60S ribosomal protein L41
YDL184C	851344	*RPL41A*	60S ribosomal protein L41
YDL191W	851336	*RPL35A*	60S ribosomal protein L35
YDR447C	852058	*RPS17B*	40S ribosomal protein S17-B
YDR471W	852082	*RPL27B*	60S ribosomal protein L27-B
YDR500C	852111	*RPL37B*	60S ribosomal protein L37-B
YER074W	856805	*RPS24A*	40S ribosomal protein S24
YER102W	856839	*RPS8B*	40S ribosomal protein S8
YER117W	856853	*RPL23B*	60S ribosomal protein L23
YFR031C-A	850590	*RPL2A*	60S ribosomal protein L2
YGL030W	852853	*RPL30*	60S ribosomal protein L30
YGL076C	852804	*RPL7A*	60S ribosomal protein L7-A
YGL103W	852775	*RPL28*	60S ribosomal protein L28
YGL135W	852742	*RPL1B*	60S ribosomal protein L1
YGR214W	853128	*RPS0A*	40S ribosomal protein S0-A
YHL033C	856352	*RPL8A*	60S ribosomal protein L8-A
YIL052C	854759	*RPL34B*	60S ribosomal protein L34-B
YIL148W	854658	*RPL40A*	60S ribosomal protein L40; Ubiquitin
YJL136C	853305	*RPS21B*	40S ribosomal protein S21-B
YJL189W	853250	*RPL39*	60S ribosomal protein L39
YJR094W-A	853557	*RPL43B*	60S ribosomal protein L43
YJR123W	853587	*RPS5*	40S ribosomal protein S5
YKL006W	853864	*RPL14A*	60S ribosomal protein L14-A
YKL180W	853674	*RPL17A*	60S ribosomal protein L17-A
YLR048W	850737	*RPS0B*	40S ribosomal protein S0-B
YLR061W	850750	*RPL22A*	60S ribosomal protein L22-A
YLR075W	850764	*RPL10*	60S ribosomal protein L10
YLR167W	850864	*RPS31*	40S ribosomal protein S31; Ubiquitin
YLR344W	851058	*RPL26A*	60S ribosomal protein L26-A
YLR441C	851162	*RPS1A*	YLR441C
YLR448W	851169	*RPL6B*	60S ribosomal protein L6-B
YML026C	854982	*RPS18B*	40S ribosomal protein S18
YML073C	854902	*RPL6A*	60S ribosomal protein L6-A
YMR142C	855173	*RPL13B*	60S ribosomal protein L13-B
YMR242C	855283	*RPL20A*	60S ribosomal protein L20A

Table 3-3: Genes in the KEGG ribosome pathway with Sgf73 occupancy continued

Ensembl Gene Stable ID	Entrez ID	Gene Symbol	Protein Name
YNL069C	855655	*RPL16B*	60S ribosomal protein L16-B
YNL096C	855628	*RPS7B*	40S ribosomal protein S7-B
YNL162W	855560	*RPL42A*	60S ribosomal protein L42
YNL178W	855543	*RPS3*	40S ribosomal protein S3
YNL302C	855414	*RPS19B*	40S ribosomal protein S19-B
YOL039W	854118	*RPP2A*	60S acidic ribosomal protein P2-alpha
YOR096W	854263	*RPS7A*	40S ribosomal protein S7-A
YOR312C	854489	*RPL20B*	60S ribosomal protein L20
YOR369C	854551	*RPS12*	40S ribosomal protein S12
YPL081W	856024	*RPS9A*	40S ribosomal protein S9-A
YPL131W	855972	*RPL5*	60S ribosomal protein L5
YPL143W	855960	*RPL33A*	60S ribosomal protein L33-A
YPL249C-A	855826	*RPL36B*	60S ribosomal protein L36-B
YPR043W	856156	*RPL43A*	60S ribosomal protein L43

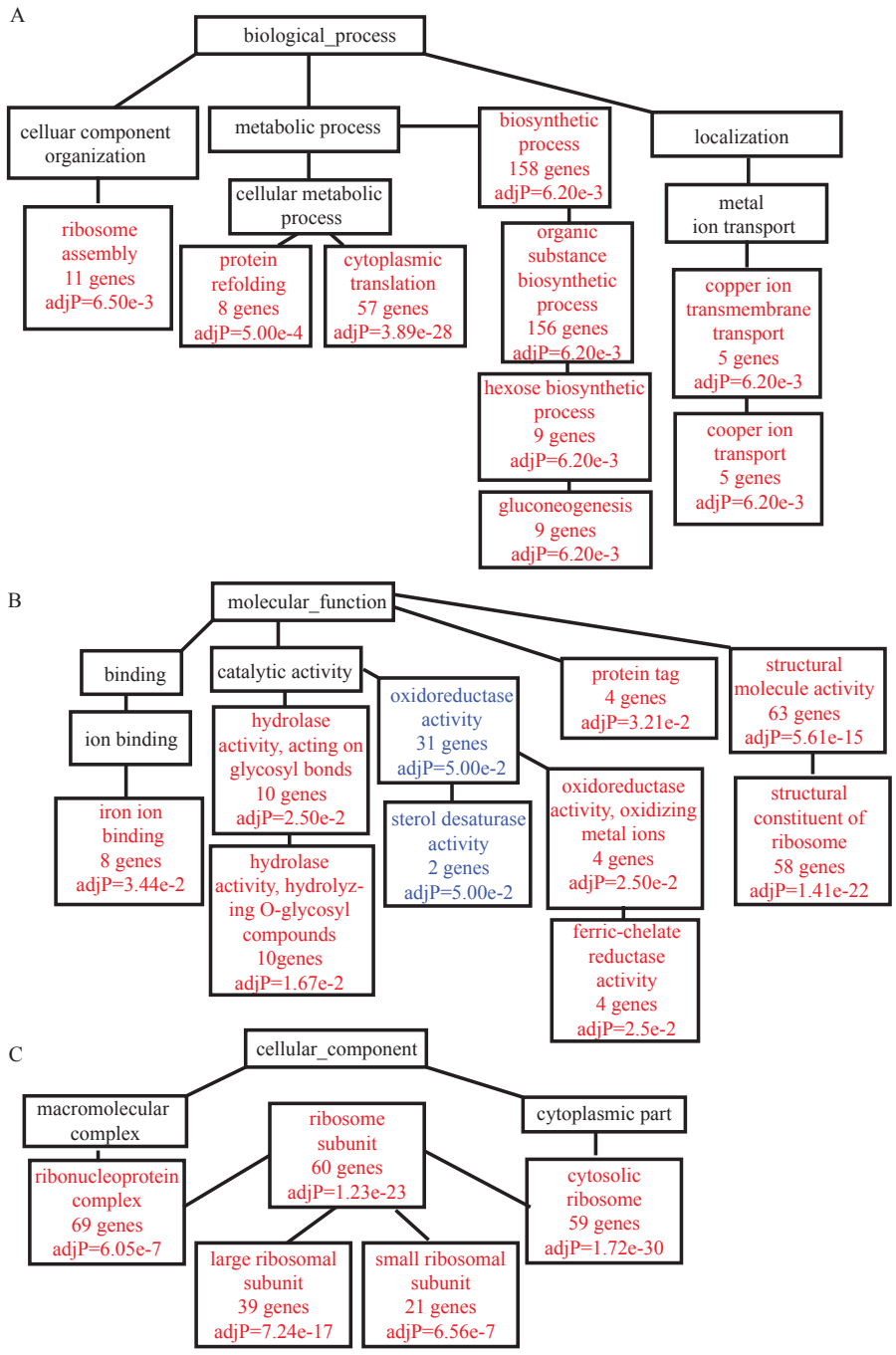
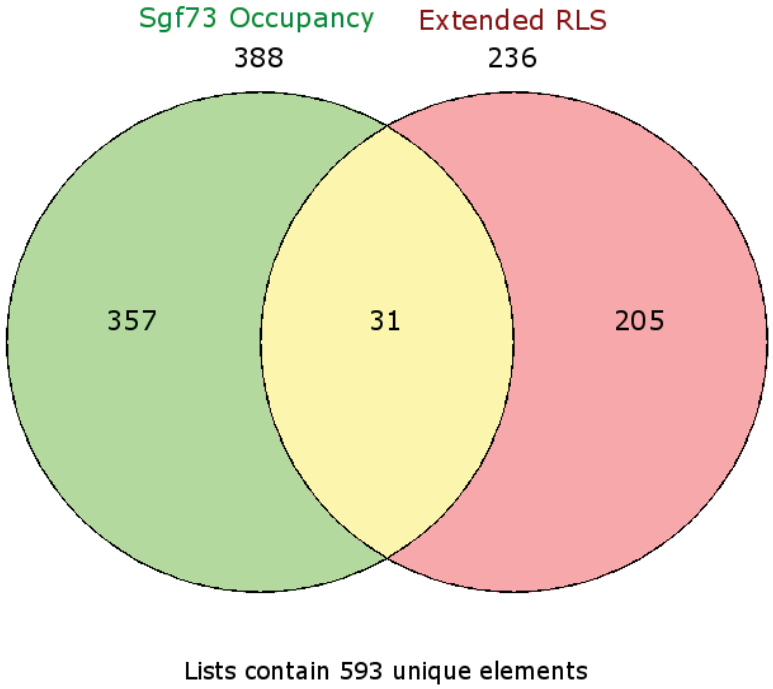


Figure 3-12: Sgf73 occupied genes GO term analysis

Each box represents a category in the gene ontology cataloging within the broader category of A) biological process, B) molecular function, and C) cellular component. If a box has red text that category was significantly enriched in the Sgf73 occupied gene set, meaning more genes from the data set fell into that category than would be expected by random chance. Blue text categories are less significant. If the category is significant, the number of genes from the data set in that category is listed along with the p-value.

A



B

CTH1	*MNN1*	*RPL20A*	*RPL34B*	*RPP2A*
CYC1	*RPL13B*	*RPL20B*	*RPL35A*	*RPS12*
FET3	*RPL16B*	*RPL22A*	*RPL37B*	*SAM1*
FMP43	*RPL19A*	*RPL23A*	*RPL43B*	*SCW4*
HXK2	*RPL19B*	*RPL26A*	*RPL6A*	*YBR238C*
IDH1	*RPL1B*	*RPL31A*	*RPL6B*	*YEH1*
				YPR064W

Figure 3-13: Overlapping Sgf73 occupied genes and RLS-liked genes
A) Venn diagram overlapping all 388 Sgf73 occupied genes and 236 RLS-linked genes known to cause RLS extension upon deletion identifies 31 shared genes B) Table of all shared Sgf73 occupied and RLS-linked genes, blue boxed genes have Sgf73 occupancy whereas red boxed genes have Sgf73 and Ubp8 occupancy.

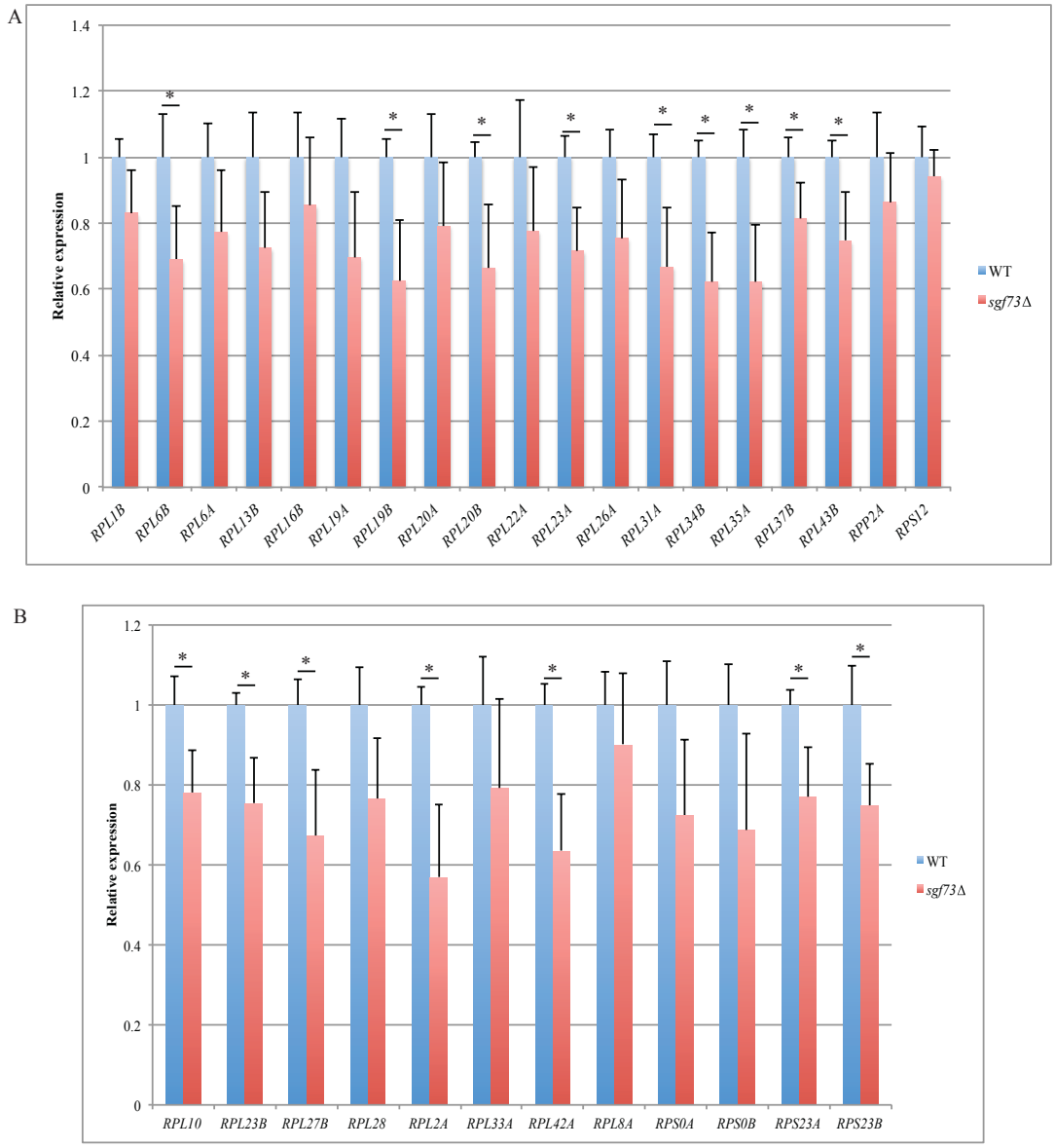


Figure 3-14: Sgf73 occupied ribosomal protein gene expression in *sgf73Δ*
A) Expression of all Sgf73 occupied RP genes that are linked to RLS extension upon their deletion. B) Expression of select additional Sgf73 occupied RP genes that are not linked to RLS extension. Expression analysis was determined by qRT-PCR, values were first normalized to the internal control *SCR1*. Expression values displayed are normalized to that in the WT strain. Error bars represent SEM, n=6, * represents $p<0.05$.

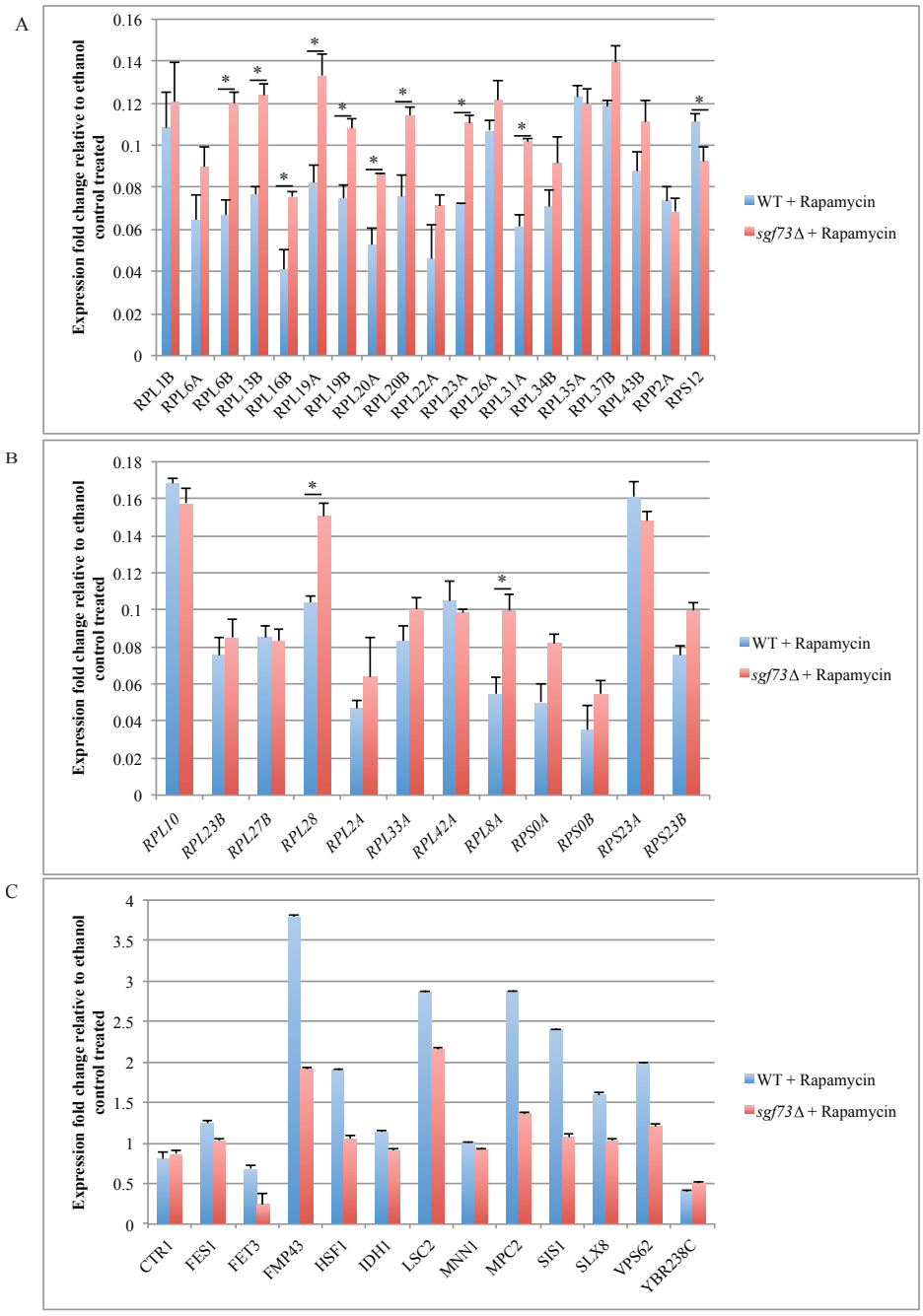


Figure 3-15: Expression of Sgf73 occupied genes upon rapamycin treatment
. A) Expression of all Sgf73 occupied RP genes that are linked to RLS extension upon their deletion. B) Expression of select additional Sgf73 occupied RP genes that are not linked to RLS extension. C) Expression of select Sgf73 occupied non-RP gene targets. Expression analysis by qRT-PCR, values were normalized to the internal control *SCR1*. Fold change was calculated based on comparison of 200ng/mL rapamycin treatment to ethanol vehicle, note scale difference in charts. Error bars represent SEM, n=6, * represents $p<0.05$

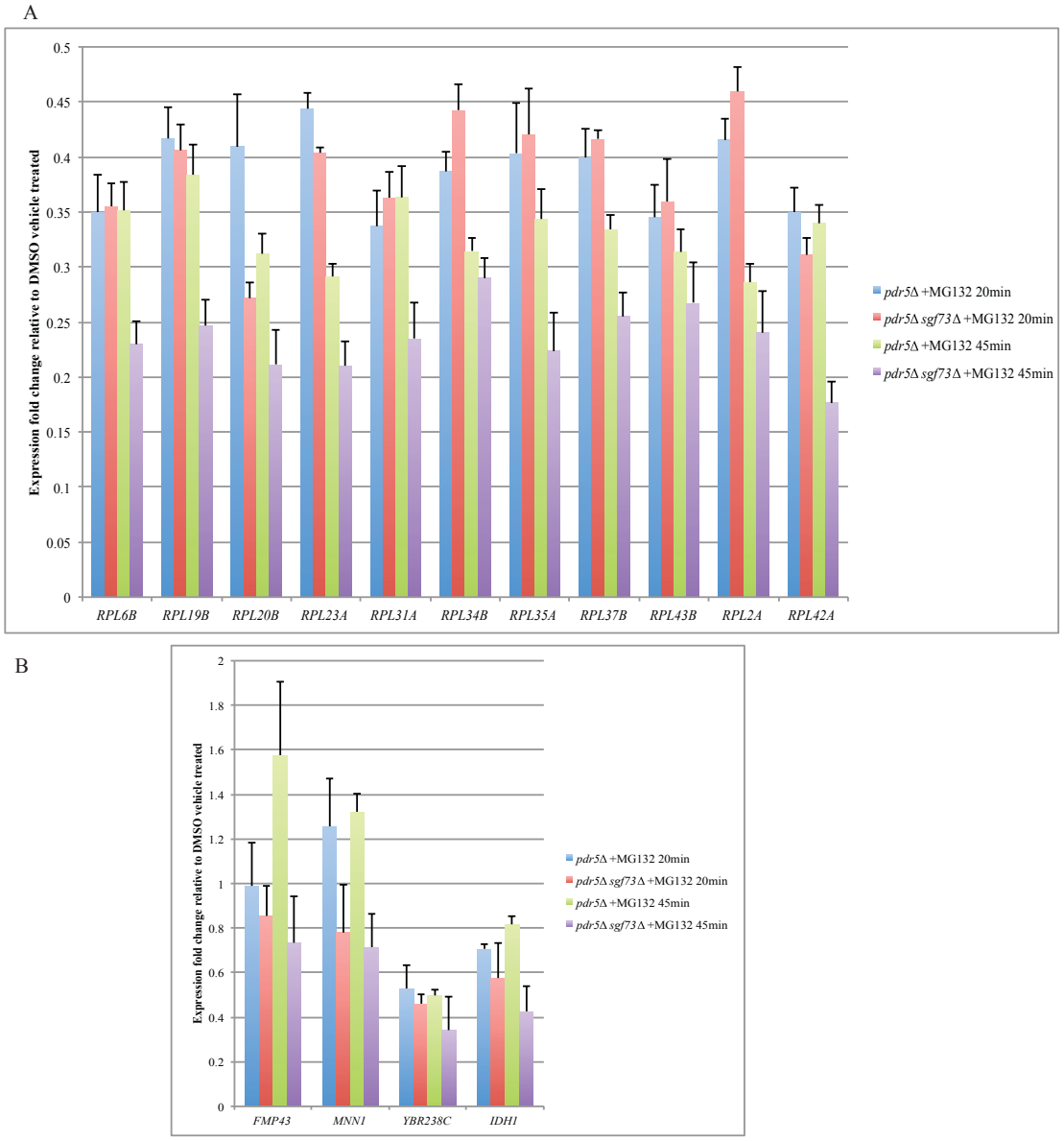
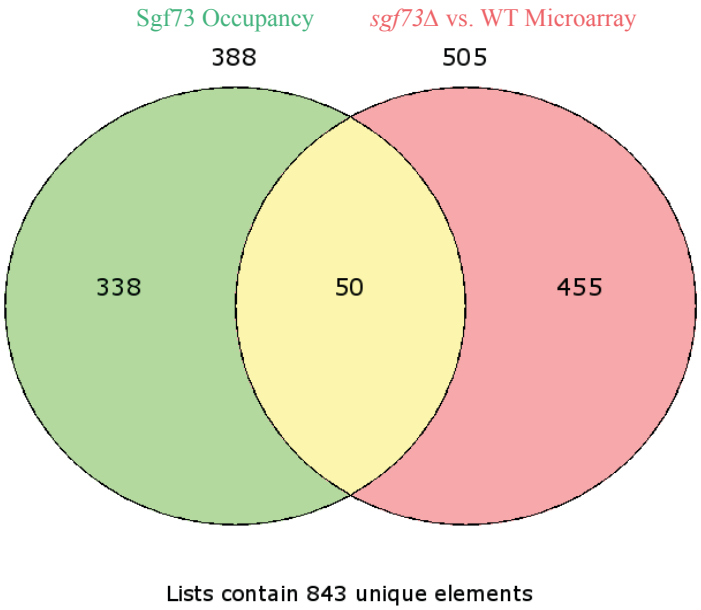


Figure 3-16: Expression of Sgf73 occupied genes upon MG132 treatment.
A) Expression of select Sgf73 occupied RP genes that are linked to RLS extension upon their deletion and two RP genes, *RPL2A* and *RPL42A* that are not RLS-linked. B) Expression of select Sgf73 occupied non-RP gene targets. Expression analysis by qRT-PCR, values were normalized to the internal control *SCR1*. Fold change was then calculated based on comparison of 25μM MG132 treatment to DMSO vehicle, note scale difference in charts. Error bars represent SEM, n=6, * represents $p<0.05$.

A

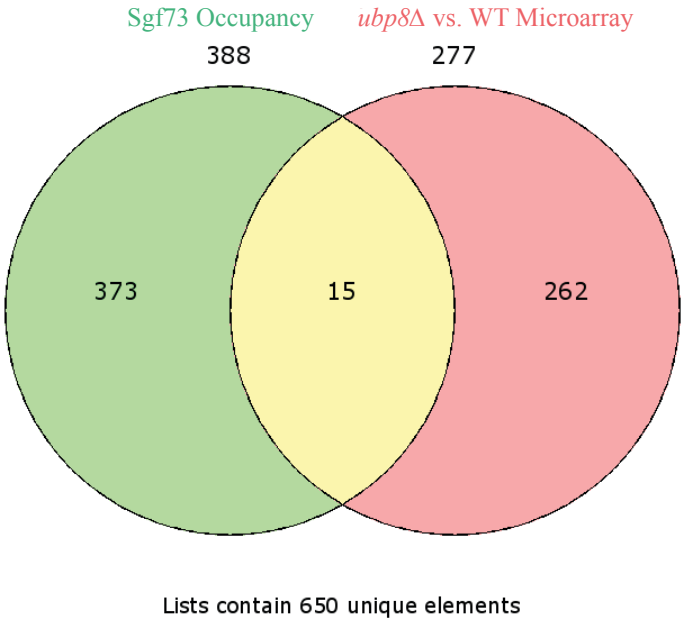


B

ACO1	*AGP1*	*HAP4*	*SAG1*
ARN1	*ARG3*	*ICY2*	*SAM3*
FRE7	*BAP3*	*IPT1*	*STD1*
GDS1	*BSC1*	*MCM1*	*TEC1*
IDH1	*CLB2*	*MDH2*	*TIR1*
RCK1	*CLN1*	*MNN1*	*YBR196C-A*
SPO20	*CLN2*	*NCA3*	*YBR238C*
TGL2	*DDR2*	*PFK27*	*YDR222W*
YCR102C	*ERG5*	*PHO89*	*YGR035W-A*
YIL046W-A	*FMP43*	*RGS2*	*YLR412C-A*
YIR007W	*FRE4*	*RIB4*	*YNR014W*
YOL155W-A	*GIC2*	*RPI1*	
YPR196W	*GIS3*	*RTS3*	

Figure 3-17: Overlapping Sgf73 occupied genes and *sgf73Δ* microarray data
A) Venn diagram overlapping Sgf73 occupied genes and all genes that had an expression difference of 1.5 fold or more in *sgf73Δ* mutants compared to WT by microarray expression analysis. B) List of shared genes between occupancy and microarray data; blue boxes indicate genes that have higher expression levels upon *SGF73* deletion, red boxes indicate genes that have lower expression levels upon *SGF73* deletion.

A

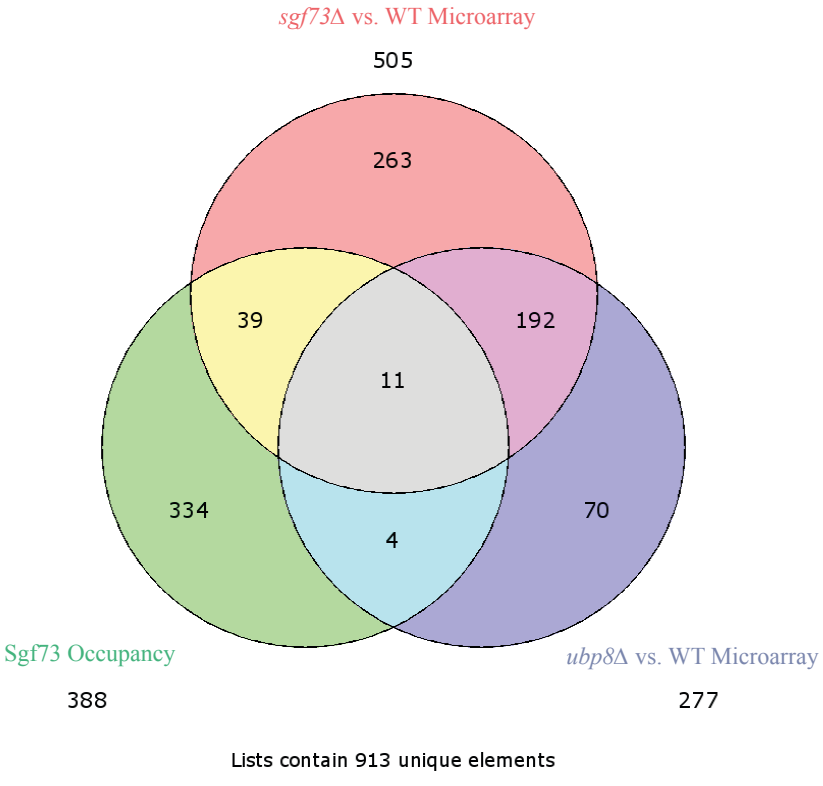


B

ARN1	*FMP43*
DDR2	*PHO89*
IDH1	
NCA3	
PDR15	
PHM8	
RIM4	
SPO20	
TGL2	
TIR1	
YCR102C	
YJL133C-A	
YOL155W-A	

Figure 3-18: Overlapping Sgf73 occupied genes and *ubp8Δ* microarray data
A) Venn diagram overlapping Sgf73 occupied genes and genes that had an expression difference of 1.5 fold or more in *ubp8Δ* mutants compared to WT by microarray expression analysis. B) List of shared genes between occupancy and microarray data; blue boxes indicate genes that have higher expression levels upon *UBP8* deletion, red boxes indicate genes that have lower expression levels upon *UBP8* deletion.

A



B

ARN1	*DDR2*	*FMP43*
IDH1	*NCA3*	*PHO89*
SPO20	*TIR1*	
TGL2		
YCR102C		
YOL155W-A		

Figure 3-19: Overlapping Sgf73 occupied genes, *sgf73Δ* and *ubp8Δ* microarray data

A) Venn diagram overlapping Sgf73 occupied genes and genes that had an expression difference of 1.5 fold or more in *sgf73Δ* or *ubp8Δ* mutants compared to WT by microarray expression analysis. B) List of shared genes between occupancy and microarray data; blue boxes are genes that have higher expression levels upon *SGF73* or *UBP8* deletion, yellow boxes are genes that have opposite expression change upon *SGF73* or *UBP8* deletion, red boxes are genes that have lower expression levels upon *SGF73* or *UBP8* deletion.

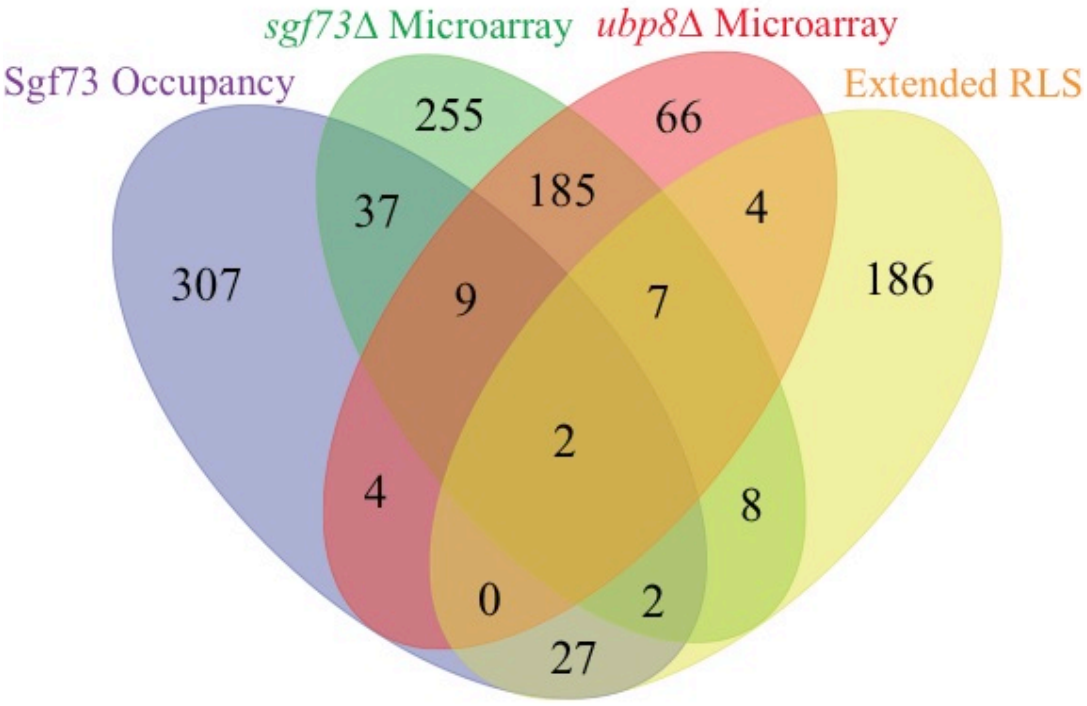


Figure 3-20: Overlapping Sgf73 Occupancy, *sgf73Δ* and *ubp8Δ* microarray, and RLS – linked gene lists

The four-way Venn diagram overlaps the gene lists of Sgf73 occupied genes, genes that had an expression difference of 1.5 fold or more in *sgf73Δ* or *ubp8Δ* mutants compared to WT by microarray expression analysis, and the gene list of single deletion mutants which are known to result in RLS extension upon deletion.

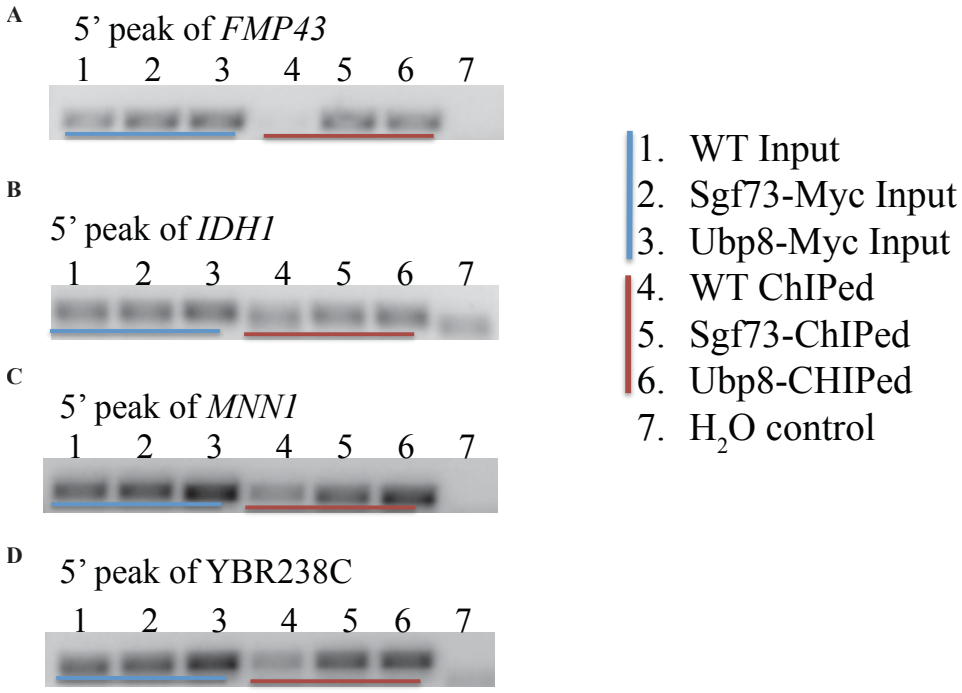


Figure 3-21: ChIP-PCR of key targets shared among Sgf73 occupancy, *sgf73Δ* and/or *ubp8Δ* microarray, and RLS – linked gene lists
PCR amplifies the peak region identified by ChIP-seq A) 5' *FMP43* B) 5' *IDH1* C) 5' *MNN1* D) 5' *YBR238C*

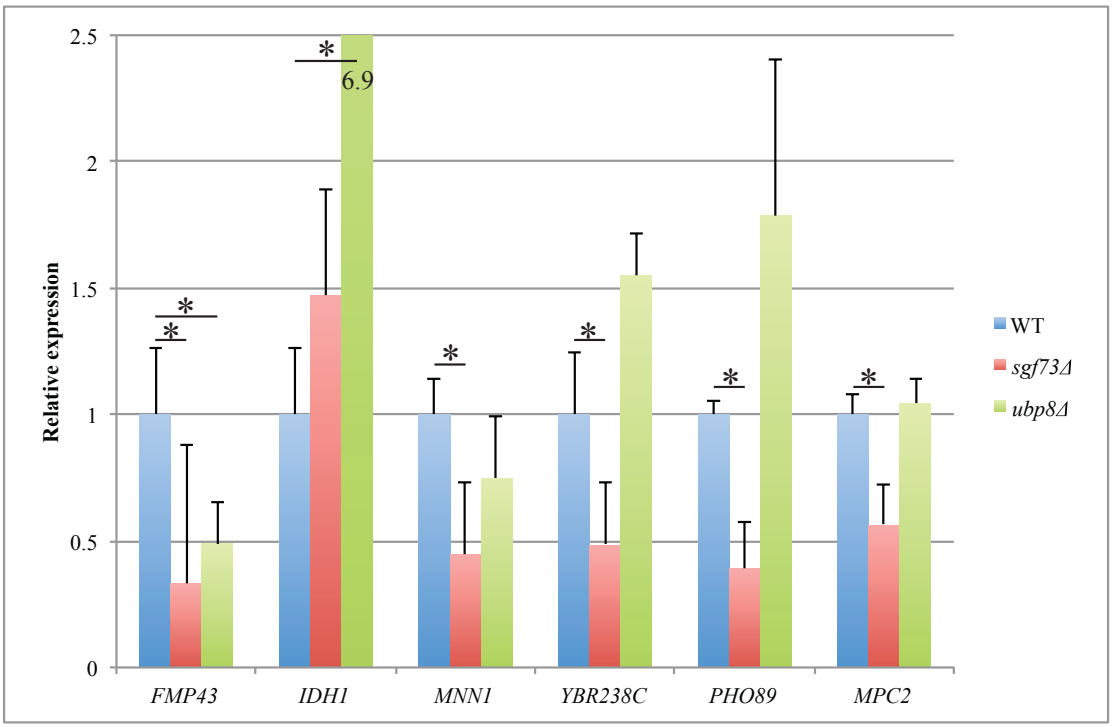


Figure 3-22: Altered expression of key targets shared among Sgf73 occupancy, *sgf73Δ* and/or *ubp8Δ* microarray, and RLS – linked gene lists
Expression analysis by qRT-PCR, all values were first normalized to the internal control *SCR1*, and then to WT expression levels; n=6, * represents a p-value <0.05.

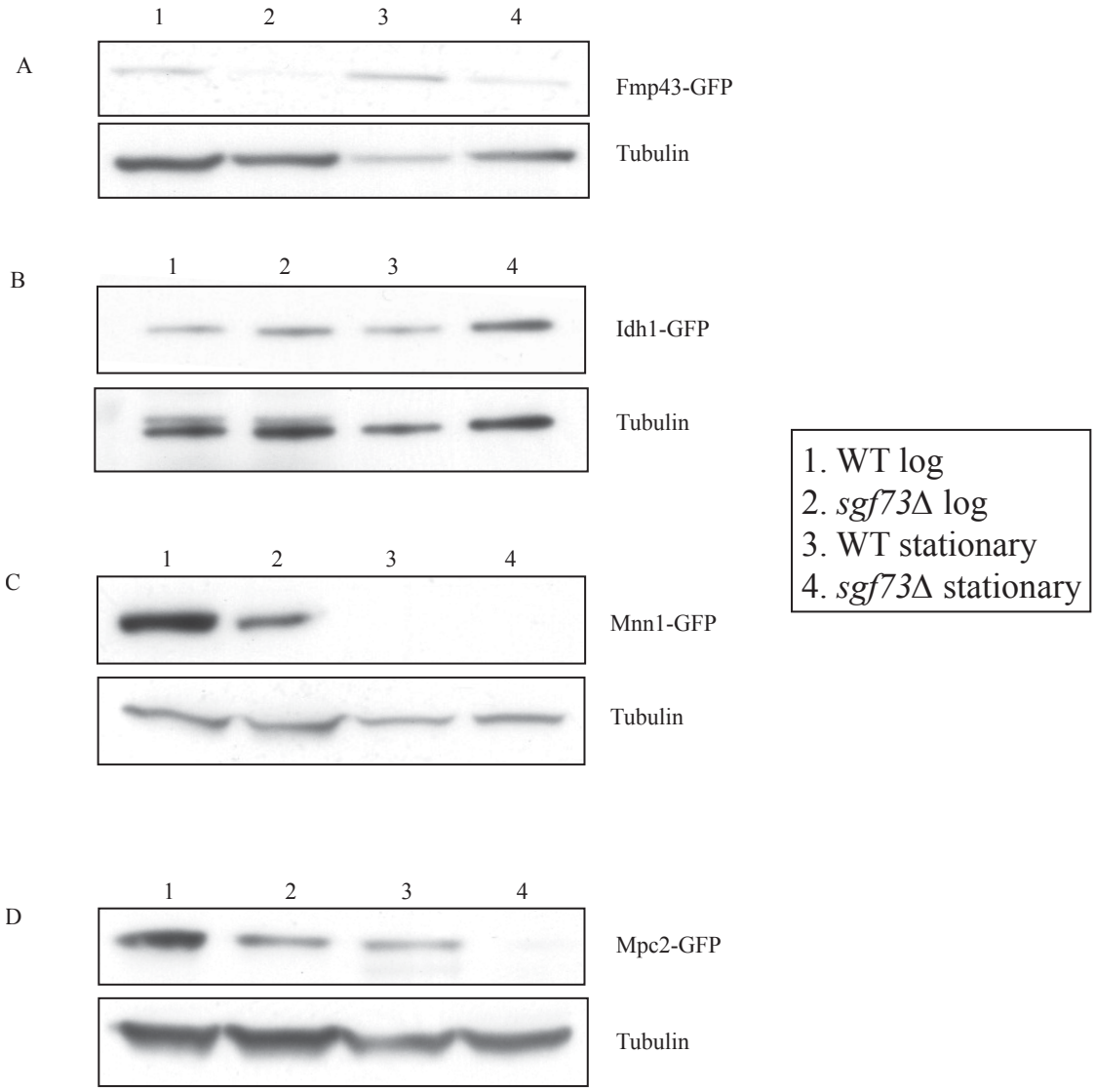


Figure 3-23: Protein levels of key non-RP gene targets
Whole cell protein lysates were prepared from WT and *sgf73Δ* cells after they had been grown to logarithmic phase (0.9OD/mL) or to a point where they were in stationary phase (24 hours). Blots were probed with anti-GFP antibody to identify the protein of interest. Anti-tubulin as a loading control.

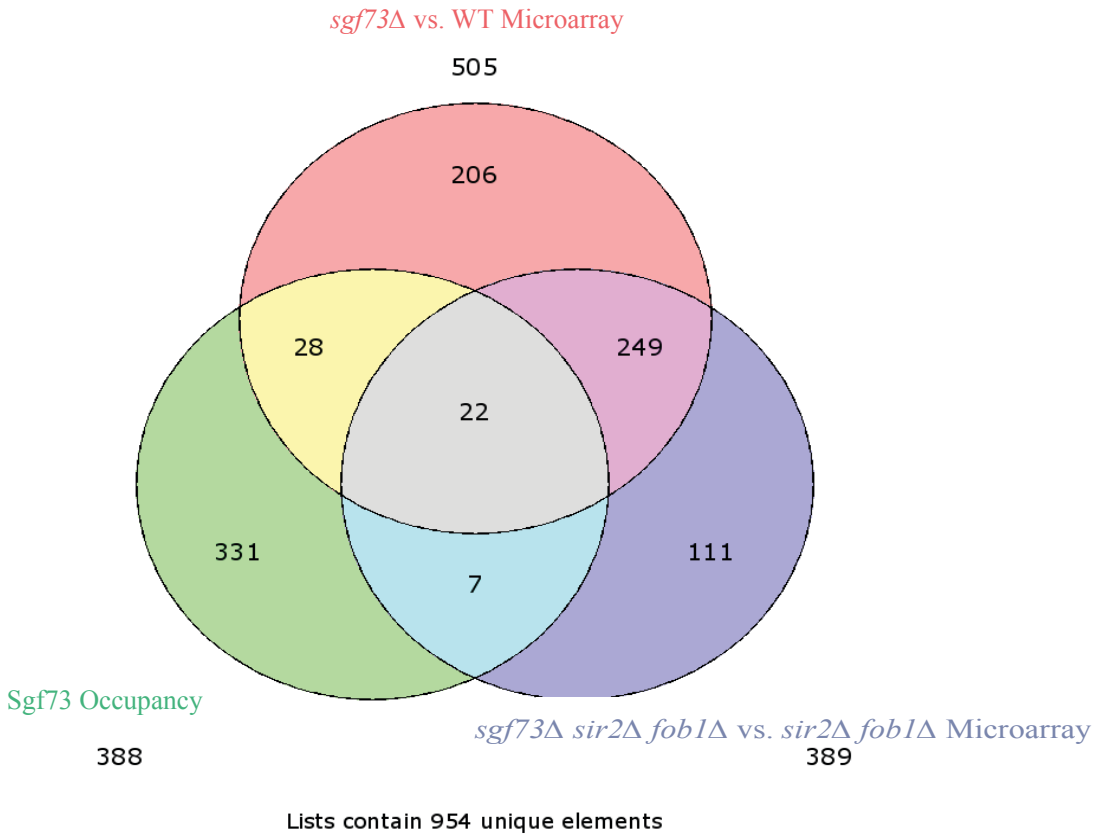


Figure 3-24: Determining Sir2-dependent changes in *sgf73Δ* mutants of Sgf73 occupied genes

The 3-way Venn diagram compares Sgf73 occupied genes and gene lists from two microarray studies that had an expression difference of 1.5 fold or more in either *sgf73Δ* mutants vs. WT or *sgf73Δ sir2Δ fob1Δ* mutants vs. *sir2Δ fob1Δ* mutants**.**

Table 3-4: Gene list of Sgf73 occupied genes with expression changes in *sgf73Δ* mutants and their possible Sir2 dependence

Sgf73 occupied genes were compared with data sets for microarray expression changes of 1.5 fold or more on the *sgf73Δ* background in either the presence or absence of Sir2 and Fob1. This resulted in groups of genes classified as having an expression change in *sgf73Δ* mutants that was Sir2 dependent or Sir2 independent.

Sir2 Dependent Genes	Sir2 Independent Genes
ACO1	*ARG3*
AGP1	*BAP3*
ARN1	*DDR2*
BSC1	*FRE7*
CLB2	*GIS3*
CLN1	*HAP4*
CLN2	*ICY2*
DIP5	*MDH2*
ERG5	*MNN1*
FMP43	*NCA3*
FRE4	*PFK27*
GDS1	*PHO89*
GIC2	*RCK1*
GSH1	*RGS2*
IDH1	*RIB4*
IPT1	*RPI1*
MCM1	*SPO20*
PHM8	*TIR1*
RTS3	*YCR102C*
SAG1	*YDR222W*
SAM3	*YIL046W-A*
SPO13	*YOL155W-A*
STD1	
STR3	
TEC1	
TGL2	
YBR196C-A	
YBR238C	
YGR035W-A	
YIR007W	
YJL133C-A	
YKR011C	
YLR412C-A	
YNR014W	
YPR196W	

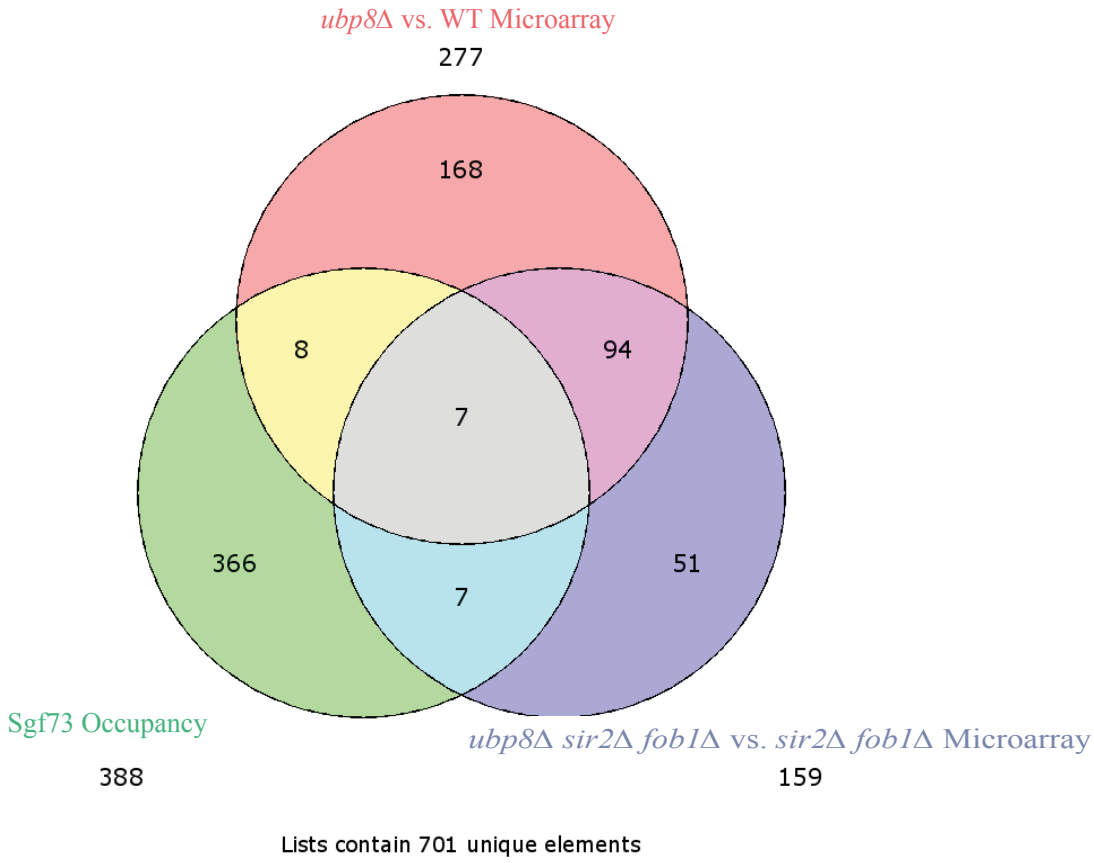


Figure 3-25: Determining Sir2-dependent changes in *ubp8Δ* mutants of Sgf73 occupied genes

The 3-way Venn diagram compares Sgf73 occupied genes and gene lists from two microarray studies that had an expression difference of 1.5 fold or more in either *ubp8Δ* mutants vs. WT or *ubp8Δ sir2Δ fob1Δ* mutants vs. *sir2Δ fob1Δ* mutants**.**

Table 3-5: Gene list of Sgf73 occupied genes with expression changes in *ubp8Δ* mutants and their possible Sir2 dependence

Sgf73 occupied genes were compared with data sets for microarray expression changes of 1.5 fold or more in the *ubp8Δ* background in either the presence or absence of Sir2 and Fob1. This resulted in groups of genes classified as having an expression change in *ubp8Δ* mutants that was Sir2 dependent or Sir2 independent.

Sir2 Dependent Genes	Sir2 Independent Genes
ARG3	*ARN1*
DDR2	*NCA3*
FMP43	*PHO89*
IDH1	*SPO20*
MCH5	*TIR1*
MDH2	*YCR102C*
PDR15	*YOL155W-A*
PHM8	
RCK1	
RIB4	
RIM4	
SUT2	
TGL2	
URA4	
YJL133C-A	

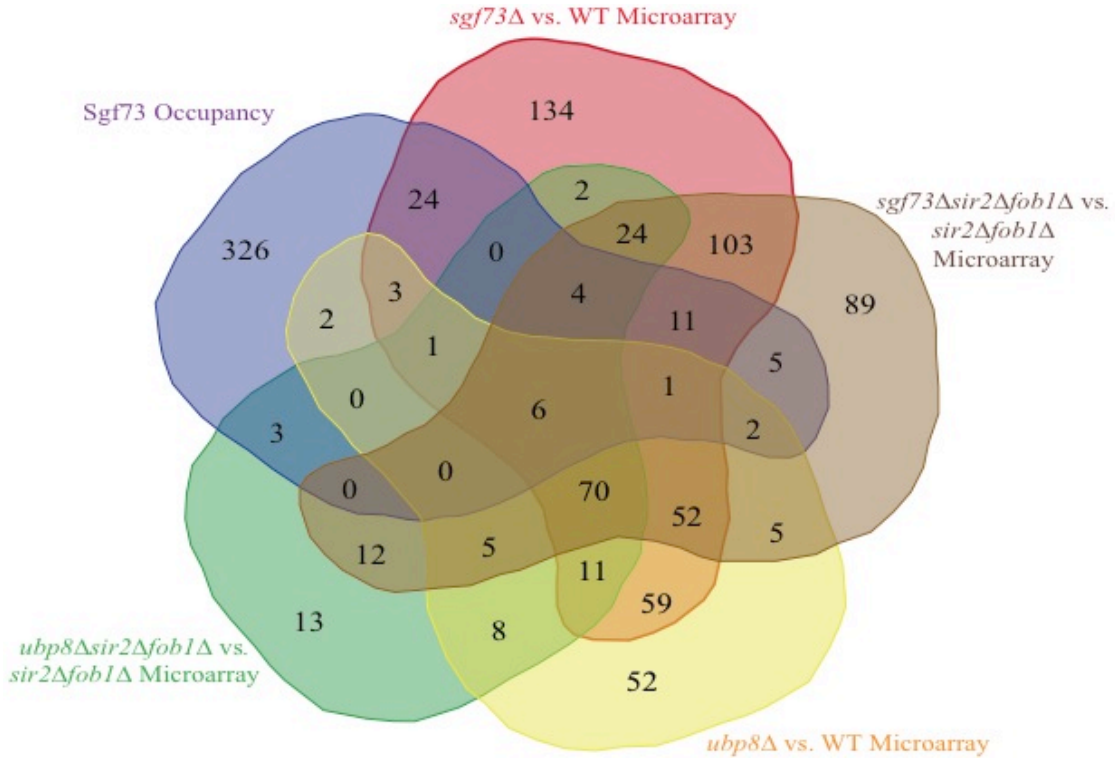


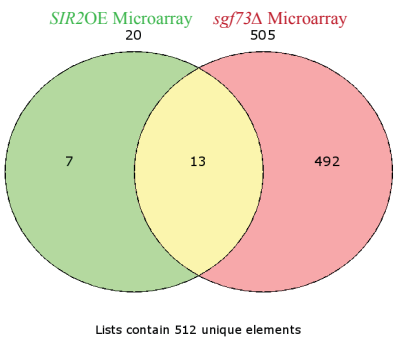
Figure 3-26: Overlapping Sgf73 occupied genes with microarray experiments from *sgf73Δ* and *ubp8Δ* mutants on both the WT and *sir2Δ fob1Δ* background.
The 5-way Venn diagram overlaps all Sgf73 occupied genes and gene lists from four microarray studies that had an expression difference of 1.5 fold or more in either *sgf73Δ* or *ubp8Δ* mutants on the WT and *sir2Δ fob1Δ* background**.** Figure generated by http://bioinformatics.psb.ugent.be.

Table 3-6: Gene list of Sgf73 occupied genes with expression changes in *sgf73Δ* and *ubp8Δ* mutants and their possible Sir2 dependence

Sgf73 occupied genes were compared with data sets for microarray expression changes of 1.5 fold or more in the *sgf73Δ* or *ubp8Δ* background in either the presence or absence of Sir2 and Fob1. This resulted in groups of genes classified as having an expression change in both *sgf73Δ* and *ubp8Δ* mutants that was Sir2 dependent or Sir2 independent.

Sir2 Dependent Genes	Sir2 Independent Genes
FMP43	*NCA3*
IDH1	*PHO89*
TGL2	*SPO20*
	TIR1
	YCR102C
	YOL155W-A

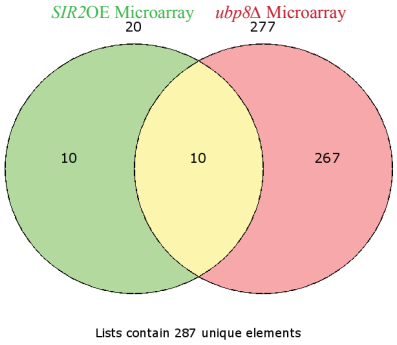
A



B

AAD15
AAD3
ADH7
ALD5
COS12
COS8
DIC1
QDR2
RDS1
SNO1
YBL101W-A
YIR042C
YPS6

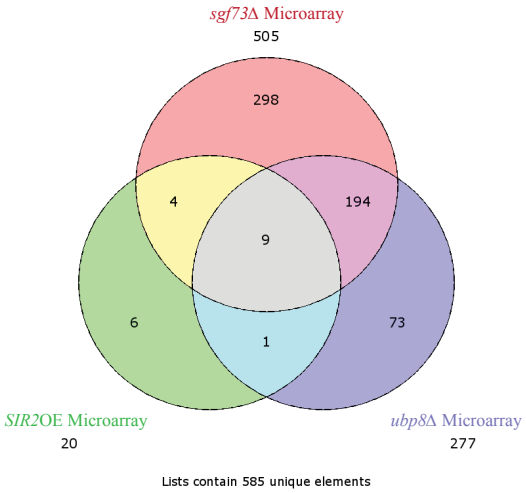
C



D

AAD15
AAD3
ADH7
COS12
RDS1
SNO1
YAR075W
YBL101W-A
YIR042C
YPS6

E



F

AAD15
AAD3
ADH7
COS12
RDS1
SNO1
YBL101W-A
YIR042C
YPS6

Figure 3-27: Shared microarray expression changes in *SIR2*OE, *sgf73Δ*, and *ubp8Δ* strains

We explored a set of microarray studies examining genes that had an expression difference of 1.5 fold or more in *SIR2*OE, *sgf73Δ*, and *ubp8Δ* mutants compared to WT. A) *SIR2*OE and *sgf73Δ* microarray expression overlap, overlapping genes are listed in B. C) *SIR2*OE and *ubp8Δ* microarray expression overlap, overlapping genes are listed in D. E) *SIR2*OE, *sgf73Δ,* and *ubp8Δ* microarray expression overlap, overlapping genes are listed in F.

Table 3-7: Select yeast targets and their human orthologs

Yeast Gene	Human Ortholog
FET3	*HEPH*
FMP43	*MPC2/BRP44*
IDH1	*IDH3B*
PHO89	*SLC20A2*
POM33	*TMEM33*
RPL10	*RPL10L*
RPL20A	*RPL18A*
RPL23B	*RPL23*
RPL27B	*RPL27*
RPL35A	*RPL35*
RPL2A	*RPL8*
RPL6A	*RPL6*
RPS23A	*RPS23*
RPS23B	*RPS23*

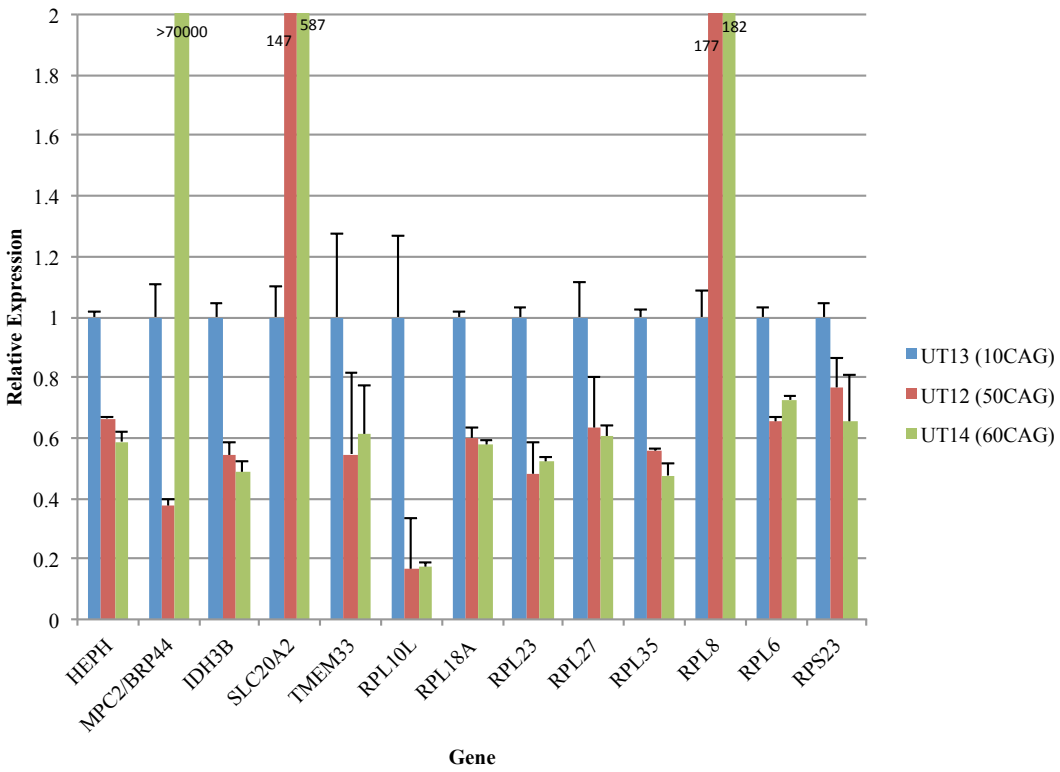


Figure 3-28: Expression of human orthologs to select yeast targets
Expression via qRT-PCR of human orthologs to select Sgf73 occupied genes in fibroblast cells from two SCA7 patients and one control individual. All values were first normalized to the internal control *GAPDH* and then to the expression of the gene in the unaffected sample.

Discussion:

Sgf73 is the transcriptional adaptor linking the core of the SAGA transcriptional co-activator complex to the Ubp8 DUB module of the complex[5-7]. Previous studies have highlighted the importance of Sgf73 for Ubp8 DUB activity, showing that deletion of *SGF73* is sufficient to abolish Ubp8 histone H2BK123 deubiquitination activity [5]. Adding to these findings previous research in our group has shown an interaction between Sgf73 and Sir2, an HDAC (see Chapter 2). Our work demonstrated that *sgf73Δ* cells have altered Sir2-dependent activity, thus implicating Sgf73 as a coordinator of gene regulation between the Sir2 and Ubp8 chromatin modifying complexes. Our research tackled the previously unanswered question as to which regions in the *S. cerevisiae* genome Sgf73 was binding, and thus which may likely coordinate epigenetic modifications in regions influencing gene transcription. Sgf73 target information is useful for further studies in both the yeast aging field as well as the study of the human neurodegenerative disease SCA7.

In our study we took a ChIP-seq approach to define the chromatin regions bound by Sgf73, Ubp8, and Sir2 in the interest of exploring the direct targets of these proteins. We found 389 distinct chromatin regions that had Sgf73 occupancy, 44 distinct chromatin regions that had Ubp8 occupancy, and limited varying occupancy by Sir2 that was dependent on the background strain. We found that of the 44 Ubp8 occupancy sites, 42 had Sgf73 occupancy. This high level of overlap is not surprising, as Sgf73 is known to bind nucleosomes through a zinc finger domain present in its SCA7 domain, facilitating the actions of Ubp8 on its targets [8]. Ubp8 does not have a known DNA binding function, thus regions ChIPed by Ubp8 may have been indirect

and may indicate regions of transient binding by Sgf73 as the samples were crosslinked. Furthermore, there is significant overlap of Gcn5 occupancy with the 42 ChIP peaks in both runs for both Sgf73 and Ubp8 [49]. Gcn5 is found at 36 of the same peak locations and an additional three genes at slightly different locations.

Most striking from our study was the highly enriched occupancy of Sgf73 5' to ribosomal genes (57 genes p = 2.63E-33). Among these ribosomal genes are a number known to extend RLS upon deletion. In total we found there was Sgf73 occupancy 5' of 31 genes known to extend RLS upon deletion. Nineteen of those are ribosomal genes. We additionally compared Sgf73 occupancy data with microarray data from *sgf73Δ*, and *ubp8Δ* to examine if transcriptional dysregulation coincided with Sgf73-occupied genes in the null strains. Of the genes with Sgf73 occupancy we found there were 50 genes in *sgf73Δ*, and 15 genes in *ubp8Δ* with altered transcription.

As previously discussed in Chapter two, in collaboration with the Kennedy group and others, we discovered that the deletion of *SGF73* results in the longest RLS extension to date, extending median RLS by 65% and maximum RLS by 53% (McCormick, Mason, & Guyenet et al Submitted). We are interested in the molecular mechanisms which are altered resulting from *SGF73* deletion, and that contribute to the extended RLS we observe. Our previous studies (McCormick, Mason, & Guyenet et al) indicated that major factors to the extended *sgf73Δ* RLS are increased H2BK123 ubiquitination as well as altered Sir2-dependent functions. However we also hypothesized that transcriptional dysregulation in the *sgf73Δ* strain contributed to extended RLS. Occupancy studies now shed light on a number of genes that may contribute to *sgf73Δ* RLS extension through transcriptional dysregulation. Promising

candidates are *FMP43*, *MNN1,* and *YBR238C*. These genes have Sgf73 occupancy, increased RLS upon deletion, and have significantly reduced levels of transcripts in *sgf73Δ* yeast in comparison to WT.

FMP43 is a mitochondrial pyruvate carrier [50], whose expression is increased upon DNA replication stress [34,51]. In a recent study, *FMP43* was found to be essential for response to the anti-oxidant ferulic acid (FA), and upon its deletion, cell growth was accelerated[51]. Fmp43 forms a multimeric complex with Mpc1 and has a paralog MPC2 with redundant function[52]. The deletion of *FMP43* alone results in impaired pyruvate uptake to the mitochondria[50] implicating its importance to cell mitochondrial metabolism. *FMP43* an interesting target in aging studies since it has a known human ortholog BRP44 (brain related protein 44, also known as MPC2, mitochondrial pyruvate carrier 2). BRP44 is functionally conserved as expressing BRP44 in *fmp43Δ* cells restores *fmp43Δ* phenotypes to WT [51]. Its conservation has also been thoroughly investigated showing that in yeast, *Drosophila*, and humans MPC1 and MPC2 are essential for mitochondrial pyruvate transport[52]. Further linking MPC2 to mitochondrial metabolism, it has been identified as a target of thiazolidinedione (TZD) insulin sensitizers in regulating pyruvate entry into mitochondria [53].

MNN1 encodes one of five alpha-1,3-mannosyltransferases located in the golgi, responsible for α1,3-mannosyl side chain post translational modifications to N- and O-linked oligosaccharides [36 38 37]. *MNN1* does not have a known human ortholog; hence, its role in other organisms is uncertain. *YBR238C* is a mitochondrial membrane protein of unknown function. It is known to be transcriptionally up-regulated by TOR.

Despite our data having a high enrichment for occupancy at ribosomal genes and there being a strong connection with ribosomal gene deletion and lifespan extension, the *sgf73Δ* and *ubp8Δ* microarray studies did not indicate any transcriptional alterations in ribosomal genes with 5' Sgf73 occupancy. In fact, in the *sgf73Δ* and *ubp8Δ* microarray data sets there was not a single ribosomal gene that appeared to have altered transcription. Our RT-qPCR studies did find that *RPL6B, RPL19B, RPL20B, RPL23A, RPL31A, RPL34B, RPL35A, RPL37B, RPL43B, RPL27B*, *RPS23A, RPL23B, RPS23B, RPL2A, RPL42A,* and *RPL10* have significantly altered expression upon *SGF73* deletion (Figure 3-14a, b); and *RPL6A, RPL28, RPL27, RPL35A, RPS23B,* and *RPS23A* have significantly altered expression in *ubp8Δ* yeast in comparison to WT (Figure 3-10 a, b). These findings suggest that there is Sgf73 occupancy on a number of ribosomal genes known to be important to aging and known to confer increased RLS upon their deletion (personal communication with Brian Kennedy). Thus, during the normal yeast growth cycle Sgf73 is causing many small changes on ribosomal transcription.

Given the fact that SAGA is a stress-induced transcriptional co-activator complex [1] we hypothesize that Sgf73 may play a role in ribosomal biogenesis under stress conditions. Previous studies have identified SAGA binding at RP genes [54] as well as direct binding of Gcn5 at RP genes [49]. Additionally, it is known that Gcn5 acetylates the RP gene transcriptional co-activator Ifh1 to stabilize and reduce RP gene transcription, and that it is subsequently de-acetylated by Sir2 [55,56]. Furthermore it has been shown that there is a reduction of SAGA occupancy at a few RP genes upon ChIP of the SAGA structural domain protein Spt7 in *sgf73Δ* mutants[54]. Indeed

Sgf73 appears to play a role in ribosomal biogenesis upon TOR inhibition by rapamycin. RP gene transcription is reduced in WT cells, and we also see a reduction in *sgf73Δ* cells although the transcript level of many RP genes is not as significantly reduced relative to WT. This elevated level of RP gene transcription with rapamycin treatment may allow *sgf73Δ* cells to grow longer under stress conditions. Additionally, upon treatment with MG132, a proteasome inhibitor, there appears to be a trend that RP gene transcript levels are lower in *sgf73Δ* mutants than WT. Thus there is a defect in RP gene transcription stress response in *sgf73Δ* mutants. This further supports the hypothesis that Sgf73 is a crucial component of SAGA regulation of RP genes.

The lack of Sgf73 occupancy at RP genes during progressive yeast aging may be an important contributing factor to *sgf73Δ* lifespan extension. It has been shown that the reduction of a number of ribosomal 60S subunits extends RLS and that the extension is partially dependent on the induced expression of *GCN4*, a nutritionally regulated transcription factor [32] (reviewed in [57]). Additionally, ER stress causes the reduction of RP gene transcription levels [58] [59], and it is thought that as cells age they experience ER stress reducing RP gene transcription [32]. Therefore altered RP gene transcription upon stress conditions in the *sgf73Δ* yeast may contribute to increased RLS by limiting ER stress signals that could further reduce RP levels.

We validated other targets that had altered transcriptional changes in either or both *sgf73Δ* and *ubp8Δ* RT-qPCR yeast studies and that had both Ubp8 and Sgf73 occupancy, but did not cause RLS extension upon deletion, and did not appear in our microarray analysis (Figure 3-10 a, b). We also validated the target *PHO89*, the only target reduced in both *sgf73Δ* and *ubp8Δ* microarray studies with Sgf73 occupancy

that did not have increased RLS upon deletion. These transcriptional changes, which do not have a direct link to increased RLS, may still be important contributors to RLS extension in the *sgf73Δ* and/or *ubp8Δ* background. Clearly not every gene under the influence of Sgf73 coordination of chromatin modification will confer RLS extension upon its individual deletion. This however does not mean that the transcriptional changes brought about in the *sgf73Δ* strain, or in the *ubp8Δ* strain where Sgf73 occupancy does not have the same power of chromatin modifying coordination due to a dysfunctional Ubp8 DUBm, are not important to overall RLS extension. The aging process is complex as is evident from the wide variety of genes whose deletion results in RLS extension. Another indication is the many published studies which implicate dietary restriction, Sir2 activity, GCN4 induction, and TOR regulation in the aging process [19] [32] [31]. Our studies of Sgf73 occupancy and transcript levels upon *SGF73* and/or *UBP8* deletion indicate that there is a varied affect on transcription, and while highlighting ribosomal involvement in Sgf73-mediated lifespan extension, these studies also implicate a number of other pathways.

Our previous studies have strongly implicated Sir2 as a key player in *sgf73Δ* RLS extension. A primary goal of our ChIP-seq experiments was to understand the overlap of genes co-occupied by Sgf73 and Sir2. Unfortunately only a limited amount of usable data was retrieved from our Sir2 experiments with the IP efficiency being predicted much below acceptable standards. Thus data garnered must be interpreted with care. Our findings suggest that the Sir2 occupancy profile does differ between WT yeast and yeast overexpressing Sir2. Additionally the Sir2 occupancy profile in *sgf73Δ* cells appears to resemble the Sir2 occupancy profile in *SIR2*OE as opposed to

that in the WT yeast. These findings are encouraging since we have previously shown an interaction between Sir2 and Sgf73 (Chapter 2), and we hypothesize that upon *SGF73* deletion Sir2 is localized to a different set of targets similar to when there is an excess amount of Sir2 in the yeast *SIR2*OE strain. These data also support our findings of altered Sir2-dependent functions in the *sgf73Δ* yeast, as the mislocalization of Sir2 will inevitably alter its normal function.

To expand our understanding of the interplay between Sgf73, Ubp8, and Sir2, we compared Sgf73 occupancy with additional microarray data, looking at expression differences in *sgf73Δ sir2Δ fob1Δ,* and *ubp8Δ sir2Δ fob1Δ,* in comparison to the *sir2Δ fob1Δ* strain and *SIR2*OE in comparison to the WT. During these analyses we also included the *sgf73Δ* and *ubp8Δ* vs. WT microarrays. With these microarrays we could explore possible Sir2 dependency of expression changes in Sgf73 bound genes. It should be taken into consideration that we are also deleting *FOB1* as *sir2Δ* strains are short lived due to ERC accumulation[9], thus while I refer to changes as Sir2 dependent/independent we can not rule out that they are Fob1- or Sir2 and Fob1-dependent/independent. Our analysis produced a list of genes with possible Sir2-dependent and Sir2-independent expression changes upon *SGF73* and/or *UBP8* deletion. In following up the top candidates from occupancy, and microarray studies we find that *PHO89* expression appears to be completely Sir2 independent in *sgf73Δ* and *ubp8Δ* cells. If we look at the top targets also considering increased RLS we find that *FMP43* and *IDH1* expression change in *sgf73Δ* appears to be Sir2 dependent whereas *MNN1* expression appears to be Sir2 independent. The finding that there appear to be Sir2 dependent and Sir2 independent populations of genes within our data

sets, again supports the idea that there are multiple pathways that are important for *sgf73Δ* RLS extension and that it is through these multiple pathway disruptions that the *sgf73Δ* strain is longer lived than either *ubp8Δ* or *SIR2*OE yeast.

Sgf73 is the yeast ortholog of the mammalian gene ataxin-7[60 61]. When the N-terminal glutamine track in ataxin-7 expands to 37 or more residues it causes the autosomal dominant neurodegenerative disease spinocerebellar ataxia type 7 (SCA7) [12]. Ataxin-7 plays a very similar role to Sgf73 in the mammalian SAGA complex STAGA, where it links core STAGA components to the Ubp8 ortholog USP22 DUBm [3]. Thus we reasoned that genes with Sgf73 occupancy and transcript alterations upon *SGF73* deletion discovered in our ChIP-seq and microarray validation might translate to transcriptional alterations of the mammalian orthologs in SCA7 patients. The expression levels of a number of human orthologs to Sgf73 bound yeast genes were explored in one control and two patient SCA7 cell lines. Using a RT-qPCR approach it appears that there is transcriptional dysregulation in a number of the genes we examined (Figure 3-28). The direction of the transcriptional change is not necessarily the same as that observed for the yeast ortholog. It is very encouraging that genes with transcriptional alterations in yeast also exhibit transcriptional differences in a mammalian model of SCA7. However, there is still more to be done to validate human targets, such as examining occupancy of ataxin-7 at sites predicted from our yeast studies. Further validation experiments pose a number of challenges but with our initial findings suggesting the yeast data do translate to mammalian cells, the yeast data can now be used as a guide for optimizing mammalian experiments. They could be used for taking a targeted ChIP approach in the mammalian system to fast track

potential target validation. Another encouraging recent finding that suggest the yeast data may translate to the mammalian system is that there are a number of ribosomal genes that are targets of the human *SIR2* ortholog, the deacetylase SIRT1 [62]. This suggests that our ChIP studies may have identified ribosomal genes that are bound by Sgf73 to coordinate Sir2 activity.

Sgf73 was successfully ChIPed and sequencing results have yielded new revelations implicating a role for regulation of ribosomal protein genes, which is disrupted in *sgf73Δ* mutants and likely a contributes to the extended RLS phenotype in *sgf73Δ*. We additionally identified targets with Sgf73 occupancy that had expression changes in *sgf73Δ* that were Sir2 dependent: *FMP43, IDH1,* and *YBR238C* and Sir2 independent targets: *MNN1* and *PHO89*. With the exception of *PHO89* the remaining genes conferred increased RLS when deleted. Our study yielded many interesting genetic interactions and mechanisms of SAGA complex members and target binding. Our findings will likely be instrumental in further yeast aging studies in relation to *SGF73* and may play an important role in understanding transcriptional dysregulation in SCA7 neurodegeneration.

Materials and Methods:

Myc tagging yeast strains

Tagging was done as described in Longtine et al[16] using their plasmid pFA6a-13Myc-kanMX6 and c terminal tagging primers. In brief, the 13Myc-KanMX6 sequence was amplified using primers which have primer overhangs corresponding to 50bp upstream of the SGF73 or UBP8 stop codon, as well as 50bp down stream of the stop codon. Competent WT Resgen cells were then prepared and transformed with the amplified product, tag was integrated by homologous recombination. Positive clones were selected for by plating on G418 plates. Molecular genotyping was then used to confirm the tags presence; positive strains were backcrossed to WT cells.

Primers for C-terminal tagging Sgf73 with 13myc:

Sgf73 Myc tagging Forward	TTGAAGTAGGTATTGGAAATTCTGTGAACCCCTACA ATGGCAGAATAAATCGGATCCCCGGGTTAATTAA
Sgf73 Myc tagging Reverse	ATTTTTTATTACTCACTTCGTGAACATGCTGGATAAC GTGCATGATTCAAGAATTCGAGCTCGTTTAAAC

Primers for C-terminal tagging Ubp8 with 13myc:

Ubp8 Myc tagging Forward	TTTTAAAGGAACAGGCATATTTATTATTCTACACCAT TCGTCAAGTAAATCGGATCCCCGGGTTAATTAA
Ubp8 Myc tagging Reverse	CTTTTTCTTCTTTTTTGTTTTATTATTATTGTTGAATGC TATTTGCTGAA GAATTCGAGCTCGTTTAAAC

Dilution assay

Cells were grown overnight in 3mLs of YPD. 1OD of the dense culture was spun down, spent media removed, and cells re-suspended in 1mL of sterile water. 1:5

serial dilutions were pinned onto control, selective, and/or drug plates, depending on the assay.

Myc ChIP and Sir2 ChIP:

A 5ml YPD culture from a single yeast colony was incubated at 30°C started the night before and diluted in 100ml of YPD in the morning to an OD of 0.25 OD/ml to be ready in ~ 4.5hours at .95-1 OD/ml. 37% formaldehyde was then added to the culture to a final concentration of 0.86% (1.15ml/50ml) and incubated with shaking at room temperature (RT) for 25 minutes. Freshly prepared 2M glycine was then added to a final concentration of 125mM (3.2ml/50ml) along with 30% NH_4OH to a final concentration of 0.2% (.365ul/50ml), and the culture was gently shaken at RT for 5min to quench the crosslinking reaction. Cells were then collected by spinning the cells down at 2500rpm for 15min. Supernatant was removed and collected as formaldehyde waste. The cell pellet was then washed three times with 20mL of cold phospate buffered saline (PBS), each time pelleting the cells at max rpm for 5min and removing supernatant after each spin. Cells were then resuspended in the remaining wash buffer and moved to a 2ml microfuge tube. The microfuge tube was then spun briefly and any additional liquid removed. The cell pellet was then resuspended in 1.5ml of FA lysis buffer (50mM HEPES-KOH pH 7.5, 140mM NaCl, 1mM EDTA pH 8.0 0.1% sodium deoxycholate, 1% Triton X-100, and 1x Roche cOmplete proteinase inhibitors) with additional detergents (0.5% NP-40, and 0.1% SDS), and then split evenly into three 1.5ml microfuge tubes. Acid washed glass beads were then added up to the meniscus and cells were lysed by bead beating in the cold room by

vortexing for 1 hour. Lysate was collected by centrifugation into a 15mL conical tube after piercing the bottom of the microfuge tubes with a needle. The recovered material was then resuspended and moved to new 1.5ml microfuge. Chromatin was then sheared in a pre-cooled horn sonicator (Sonic Dismembrator FB-505), using pulse cycles of 20sec on and 30sec off at 80% amp for a total time of 30min (majority of fragments should be around 250-400bp). After sonication, tubes were spun at 4°C at max speed for 5min and the supernatant was transferred to new low-bind microfuge tubes (Lifetech AM12450). An equal volume of FA-lysis buffer was then added to the sheared chromatin. Lysate was then pre-cleared by adding 10ul of protein A:G magnetic beads and rotating for 30 minutes at 4°C. After pre-clearing tubes were placed on a magnet and the supernatant moved to a new tube. A sample from each preparation was taken at this point to represent the input chromatin and protein.

For the Myc ChIP experiments 100µl of anti-c-myc EZview resin (Sigma E6654-1ML) was added to the tubes and samples were rotated overnight at 4°C.

For Sir2 ChIP experiments anti-Sir2 [58] coupled magnetic dynabeads (lifetech 10002D and 10004D) were prepared ahead of time. 30µl of beads per sample were placed into a 15ml conical tube and washed with 10ml of phospate buffered saline (PBS) containing 5mg/ml bovine serum albumin (BSA) to pre-block the beads. The tube was rotated for 5 minutes and then the supernatant removed, and this was repeated once more. 3µl of anti-Sir2 antibody per sample was then added along with 250µl BSA/PBS per sample and the allowed to rotate overnight at 4°C. After conjugation beads were washed twice with BSA/PBS and then resuspended in FA

lysis buffer (30µl per sample). 30µl of conjugated beads were then added to each diluted chromatin sample and allowed to rotate overnight at 4°C.

After overnight binding beads were washed by rotating samples at 4°C for 10minutes and then spinning gently (6000rpm 4 sec) after each wash (or in the case of the magnetic beads the tubes were placed on a magnet) and removing supernatant. Beads were first washed twice with 1mL of FA lysis buffer, then twice with 1ml of wash buffer (10mM Tris pH 8, 0.25M LiCl, 0.5% Np40, 0.5% sodium deoxycholate, 1mM EDTA), and finally with 1ml of Tris-EDTA (TE). After the final wash all beads underwent an additional spin and remaining liquid was removed by using a needle. To reverse the crosslinking between the protein and the DNA 100µl of TE + 1% SDS was added to the beads (10µl was taken at this point for protein analysis) and then samples were incubated at 65°C for 6 hours. Following incubation 8µl of TE and 2µl of RNAse was added to the sample and placed at 37°C for 1 hour to degrade any RNA. Protein in the sample was degraded by adding 1.5µl of proteinase K (20mg/ml) and incubating at 37°C overnight. The next day DNA was column purified from the sample using the Qiagen MinElute PCR purification kit (Qiagen 28006) and eluted with 13µl of nuclease-free water.

Fragment size was analyzed on 6% acrylamide gels.

Sequencing library preparation:

Libraries for sequencing on the Illumina HiSeq platform were generated using the ChIPed fragments acquired as previously described and the Ovation ultralow DR multiplex 1-8 library system from NuGEN (NuGEN part# 0330). The kit has linkers to

barcode up to 8 samples that can be pooled to run on one sequencing lane. The protocol supplied digitally from NuGEN with kit purchase was used to make and amplify the libraries. Briefly: End repair was performed on 10µl of the eluted DNA. After the end repair process adaptors were ligated to the ends of the fragments. This adaptor ligation process includes barcoding each library, which is needed to multiplex the samples for sequencing. Two sequencing lanes of seven pooled samples were used so it was essential that seven different ligators were used in the samples being pooled together. After adaptor ligation was complete Agencourt RNAclean XP magnetic beads were used to purify the ligator adapted fragments and subsequently the fragments were eluted off the beads. Library amplification was then performed on the fragments using 18 PCR cycles (72°C – 2min, 18 cycles (94°C – 30sec, 60°C – 30sec, 72°C – 1min), 72°C – 5min). The amplified library was then purified using the Agencourt RNAclean XP magnetic beads and libraries were eluted in 30µl of TE.

Library size selection, fragment size validation, and quantity analysis:

After library construction it was necessary to select for fragments in the 150 – 400 bp range. The libraries were loaded onto a 6% TBE acrylamide gel using Invitrogen's 25bp ladder as the size marker (Invitrogen 10597-011), only three libraries per 10-well gel were loaded, skipping lanes to avoid cross contamination between libraries upon excision from the gel. After running the gels at 150V for about 40 minutes, the DNA in the range from 150-400 bps was excised using a razor blade under minimal UV exposure. The gel block was then placed in a 0.65ml microfuge tube with 5-7 small holes punched in it using a needle that was then placed inside a

1.5ml microfuge tube. These tubes were spun to fragment the gel into small pieces. 350µl of elution buffer (300mM NaCl, 10mM Tris pH 7.6) was then added to the tube and allowed to rotate at 4°C overnight. The gel liquid mixture was then transferred to a SpinX tube containing a cellulose acetate membrane with a pore size of 0.45µm (Sigma CLS1862) to remove gel from the DNA in suspension. DNA was then precipitated by adding; 1ml cold 100% ethanol, 35µl of sodium acetate (3M pH 5.2), and 2µl glycogen (5mg/ml). The mixture was then placed at -80°C for 2 hours, thawed on ice, and spun at 4°C for 30 minutes. Pellet was washed with 75% ethanol, and spun again at 4°C for 15 minutes. The pellet was air dried for 5 minutes, and the library was then resuspended in 15µl of Qiagen EB.

After size selection by agarose gel 1µl of the library was run on a 2100 Bioanalyzer DNA 100 chip (Agilent 5067-1504). The Bioanalyzer allows for an estimation of the amount of DNA in the library as well as verifying the size of the fragments.

Sequencing on the Illumina Hiseq 2000

All libraries were shipped to the Beijing Genomics Institute (BGI) for sequencing on the Illumina Hiseq 2000 platform. Single end 50 base pair read sequencing was done, from the seven pooled ChIP-seq libraries per lane. Post sequencing, BGI sorted all sequence reads by their barcode and removed low-quality sequence reads, returning a raw reads file for each of the fourteen libraries sequenced. Upon return sequencing files were renamed to reflect; lab in which the project was

done, who did the project, the date the sequencing was returned, the yeast strain, the type of ChIP, and the replicate number (Table 1).

Sequencing analysis on Illumina HiSeq reads:

To analyze the sequencing reads obtained form the ChIP sequencing experiments the program HOMER (Hypergeometric Optimization of Motif EnRichment) v4.2, which is available for download at http://biowhat.ucsd.edu/homer/, was used. HOMER is a collection of command line tools that are useful for analyzing ChIP-seq as well as many other sequencing data sets. I ran HOMER using the Terminal application on a Mac computer: it should be noted that a number of other freely available applications need to be installed to run HOMER as outlined on their website, as is Xcode (available though the Apple App Store) if using a Mac to successfully run the program. The *Saccharomyces cerevisiae* genome SacCer3, *Saccharomyces cerevisiae* S288c assembly from Saccharomyces Genome Database (GCA_000146055.2), was used for all sequence alignments and annotations.

These instructions / methods are a variation on what is posted at http://biowhat.ucsd.edu/homer/

General sequencing data analysis instructions and example of terminal code (terminal must be in BASH mode) used to analyze files. For the code given the file used is 4_LaSpada AGM-052513-Sgf73MycScer-MycChIP2.fq (you must always direct the program to your file, you can "drag and drop" a file into terminal) after alignment for the purposes of the methods section the file will be known as ex.sam, and example ChIP-seq.

The first step carried out after retrieving and re-naming (file name indicates original sample number, lab work was done in, my initials, date of sequence generation, cell type experiment was done in, the experiment done, and the replicate number) the raw sequence reads files was to run a general quality control check on the sequence files using the program FASTQC which can be downloaded from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Once downloaded this program runs independently with the file to which you wish to run the quality control on is selected via File > Open. The program then produces a number of analyses allowing for a general idea of the sequence quality. In the case of my sequencing files most parameters suggested reads of high quality, however there was often a flag indicating a high level of sequence duplication likely due to the PCR amplification before sequencing. The duplicated reads were filtered out in the analysis process thus eliminating this problem.

Using Bowtie downloaded from http://bowtie-bio.sourceforge.net/index.shtml and run through terminal, alignment of raw sequencing reads was done to the SacCer3 genome: index for s_cerevisiae SacCer3 genome is also available for download from the same site and was used for alignment. Alignment was run separately for each experiment.

General commands:

1. Direct terminal to your downloaded Bowtie program by changing directories (this depends on where you placed these files):

```
GENAW80171CCAGV:~ agmason$ cd ~/homer/bowtie
```

2. Run alignment on your sequencing file by running bowtie: ./bowtie <genome> <fastq file> <output file name>

```
GENAW80171CCAGV:bowtie agmason$ ./bowtie s_cerevisiae
/Users/agmason/Homer/bowtie/reads/4_LaSpadaAGM-052513-Sgf73MycScer-
MycChIP2.fq ex.sam
# reads processed: 12687346
# reads with at least one reported alignment: 7664624 (60.41%)
# reads that failed to align: 5022722 (39.59%)
Reported 7664624 alignments to 1 output stream(s)
```

At this point the program has created a SAM file of aligned reads. The next step is to make a tag directory using HOMER's command makeTagDirectory. This program goes through all of the aligned reads from Bowtie and splits the tags (aligned read) into separate files based on the chromosome that they aligned to. I used the additional command –tbp 1 which limits the number of tags per base pair to 1, effectively removing all duplicate reads that started at the same position. This was done since our libraries were PCR amplified before sequencing, eliminating artificially high tag counts at certain regions due to PCR amplification and effectively getting rid of sequence duplication. Also let it be noted that more than one alignment file can be merged into one tag directory file, which was done for the control sequencing runs.

1. Run the pearl script makeTagDirectory found in the bin folder of the downloaded HOMER program: makeTagDirectory <desired output directory name>/ <alignment file.sam> [additional alignment files if want merged into one directory] -tbp1

```
GENAW80171CCAGV:~ agmason$ makeTagDirectory exampleChIpseq
/Users/agmason/Homer/bowtie/ex.sam  -tbp 1
Will parse file: /Users/agmason/Homer/bowtie/ex.sam
Creating directory: exampleChIpseq and removing existing *.tags.tsv
Reading alignment file /Users/agmason/Homer/bowtie/ex.sam
Guessing that your alignment file is bowtie format
Optimizing tag files...
Estimated genome size = 12156581
Estimated average read density = 0.100692 per bp
Total Tags = 1224068.0
Total Positions = 1224068
Average tag length = 49.0
Median tags per position = 1 (ideal: 1)
Average tags per position = 1.000
Fragment Length Estimate: 177
Peak Width Estimate: 200
Autocorrelation quality control metrics:
        Same strand fold enrichment: 1.6
        Diff strand fold enrichment: 1.5
        Same / Diff fold enrichment: 1.2
        Guessing sample is ChIP-Seq - autocorrelation looks good.
```

After the tag directory is created a number of files exist, one for each chromosome. To visualize the aligned sequencing reads on the UCSC genome browser the files must be converted into a bedGraph formatted file using the HOMER pearl script makeUCSCfile.

1. Run the makeUCSCfile program found in the bin folder of the downloaded HOMER program, on your tag directory: makeUCSCfile <tag directory> -o auto , auto can be replaced with your desired file name.bed

```
GENAW80171CCAGV:~ agmason$ makeUCSCfile
/Users/agmason/exampleChIpseq / > example.bed

Visualization fragment length = 177
No need to remove tags to get desired file size
Generating bedGraph for Scchr01
```

Generating bedGraph for Scchr02
Generating bedGraph for Scchr03
Generating bedGraph for Scchr04
Generating bedGraph for Scchr05
Generating bedGraph for Scchr06
Generating bedGraph for Scchr07
Generating bedGraph for Scchr08
Generating bedGraph for Scchr09
Generating bedGraph for Scchr10
Generating bedGraph for Scchr11
Generating bedGraph for Scchr12
Generating bedGraph for Scchr13
Generating bedGraph for Scchr14
Generating bedGraph for Scchr15
Generating bedGraph for Scchr16
Generating bedGraph for Scmito

2. Unfortunately there is a problem with HOMER's makeUCSCfile output for *Saccharomyces cerevisiae* SacCer3 with the chromosomes being annotated as Scchr followed by an Arabic numeral whereas the UCSC genome browser notates the chromosomes in the SacCer3 genome as chr followed by a roman numeral. To solve this problem a script was run as follows to convert the chromosome numbers so the file can be visualized. Script to run: cat <**bed file needing conversion**>| awk '{gsub("Scchr01","chrI"); gsub("Scchr02","chrII"); gsub("Scchr03","chrIII");gsub("Scchr04","chrIV"); gsub("Scchr05","chrV"); gsub("Scchr06","chrVI"); gsub("Scchr07","chrVII"); gsub("Scchr08","chrVIII"); gsub("Scchr09","chrIX"); gsub("Scchr10","chrX"); gsub("Scchr11","chrXI"); gsub("Scchr12","chrXII"); gsub("Scchr13","chrXIII");gsub("Scchr14","chrXIV"); gsub("Scchr15","chrXV"); gsub("Scchr16","chrXVI"); gsub("Scmito","chrM"); print}' | awk '{if((($1=="chrV")&&($3>576875))||(($1=="chrVI")&&($3>270162)) ||(($1=="chrVII")&&($3>1090940)) ||(($1=="chrIX")&&($3>439888))

||(($1=="chrXI")&&($3>666816))||(($1=="chrXIV")&&($3>784333))
||(($1=="chrXVI")&&($3>948066)))}{}else print}'><**newfilename.bed**> ; gzip <**newfilename.bed**>

```
GENAW80171CCAGV:~ agmason$ cat  example.bed | awk
'{gsub("Scchr01","chrI"); gsub("Scchr02","chrII");
gsub("Scchr03","chrIII");gsub("Scchr04","chrIV"); gsub("Scchr05","chrV");
gsub("Scchr06","chrVI"); gsub("Scchr07","chrVII"); gsub("Scchr08","chrVIII");
gsub("Scchr09","chrIX"); gsub("Scchr10","chrX"); gsub("Scchr11","chrXI");
gsub("Scchr12","chrXII"); gsub("Scchr13","chrXIII");gsub("Scchr14","chrXIV");
gsub("Scchr15","chrXV"); gsub("Scchr16","chrXVI"); gsub("Scmito","chrM");
print}' | awk '{if(($1=="chrV")&&($3>576875))||(($1=="chrVI")&&($3>270162))
||(($1=="chrVII")&&($3>1090940)) ||(($1=="chrIX")&&($3>439888))
||(($1=="chrXI")&&($3>666816))||(($1=="chrXIV")&&($3>784333))
||(($1=="chrXVI")&&($3>948066)))}{}else print}'> example.renamed.bed ;  gzip
example.renamed.bed
```

After running this script a .bed.gz file is produced (gziped bed file). This file can then be uploaded to genome.ucsc.edu by first selecting the correct genome and then selecting add custom tracks.

The next step is to find significant peaks in the sequencing data set by comparing the tag directory of interest to a background tag directory, which in my case was the combined tag directory from two control (WT no tag Myc-ChIP or sir2Δ fob1Δ Sir2 ChIP) sequencing runs/alignments, as well as filtering out peaks using standard thresholds as specified by HOMER default settings (details can be found at http://biowhat.ucsd.edu/homer/ngs/peaks.html). To find peaks the HOMER pearl script findPeaks was run.

1. Run the findPeaks program found in the bin file of the downloaded HOMER program on the created tag directory: findPeaks <tag directory of interest> -o auto -i <control tag directory>

```
GENAW80171CCAGV:bin agmason$ findpeaks
/Users/agmason/Homer/bin/exampleChIpseq / -o auto -i /Users/agmason/Tag\
Directories\ replicates\ merged\ and\ Peaks/WTMycChIP
Fragment Length = 177
!!! Estimated genome size (from tag directory) is smaller than default
   genome size.  Using estimate (12156581) [to change specify -gsize]
Total Tags = 1224068.0
Tags per bp = 0.100692
Max tags per bp set automatically to 10.0
Finding peaks of size 200, no closer than 400
        Finding peaks on Scchr01 (minCount=19.1), total tags positions = 24134
        Finding peaks on Scchr02 (minCount=19.1), total tags positions = 76278
        Finding peaks on Scchr03 (minCount=19.1), total tags positions = 30052
        Finding peaks on Scchr04 (minCount=19.1), total tags positions = 140955
        Finding peaks on Scchr05 (minCount=19.1), total tags positions = 66054
        Finding peaks on Scchr06 (minCount=19.1), total tags positions = 26351
        Finding peaks on Scchr07 (minCount=19.1), total tags positions = 110198
        Finding peaks on Scchr08 (minCount=19.1), total tags positions = 57457
        Finding peaks on Scchr09 (minCount=19.1), total tags positions = 41099
        Finding peaks on Scchr10 (minCount=19.1), total tags positions = 71313
        Finding peaks on Scchr11 (minCount=19.1), total tags positions = 57236
        Finding peaks on Scchr12 (minCount=19.1), total tags positions = 144085
        Finding peaks on Scchr13 (minCount=19.1), total tags positions = 91696
        Finding peaks on Scchr14 (minCount=19.1), total tags positions = 74890
        Finding peaks on Scchr15 (minCount=19.1), total tags positions = 107270
        Finding peaks on Scchr16 (minCount=19.1), total tags positions = 97603
        Finding peaks on Scmito (minCount=19.1), total tags positions = 7397
        Tags Used for cluster (less clonal tags) = 1224068.0 / 1224068.0
Expected tags per peak = 20.138359 (tbp = 0.100692)
        Threshold       Peak Count      Expected Peak Count     FDR
        44      2126.000        0.362   0.000170
        43      2173.000        0.789   0.000363
        42      2226.000        1.702   0.000764
        41      2281.000        3.604   0.001580
        40      2341.000        7.478   0.003195
        39      2394.000        15.173  0.006338
        38      2486.000        30.074  0.012098
        37      2564.000        58.193  0.022696
        36      2667.000        109.8540.041190
        35      2781.000        202.2050.072709
        34      2908.000        362.7100.124728
```

33	3049.000	633.6930.207836	
32	3213.000	1077.743	0.335432
31	3437.000	1783.341	0.518866
30	3652.000	2869.509	0.785736
29	3916.000	4487.557	1.145954
28	4224.000	6817.622	1.614020
27	4548.000	10057.283	2.211364
26	4985.000	14400.780	2.888823
25	5468.000	20008.544	3.659207
24	6020.000	26970.089	4.480081
23	6648.000	35266.557	5.304837
22	7285.000	44741.945	6.141653
21	7924.000	55093.253	6.952707
20	8612.000	65887.440	7.650655
19	8612.000	76607.457	8.895432
18	8612.000	86721.504	10.069845
17	8612.000	95761.621	11.119557
16	8612.000	103392.930	12.005682
15	8612.000	109456.034	12.709711
14	8612.000	113972.118	13.234106
13	8612.000	117111.656	13.598660
12	8612.000	119138.336	13.833992
11	8612.000	120345.989	13.974221
10	8612.000	121005.635	14.050817
9	8612.000	121333.192	14.088852
8	8612.000	121479.580	14.105850
7	8612.000	121537.733	14.112603
6	8612.000	121557.946	14.114950
5	8612.000	121563.969	14.115649
4	8612.000	121565.464	14.115823
3	8612.000	121565.761	14.115857
2	8612.000	121565.805	14.115862
1	8612.000	121565.810	14.115863
0	8612.000	121565.810	14.115863

0.10% FDR Threshold set at 42.0 (poisson pvalue ~ 1.40e-05)
2226 peaks passed threshold
Differential Peaks: 669 of 2226 (30.05% passed)
Local Background Filtering: 457 of 669 (68.31% passed)
Clonal filtering: 457 of 457 (100.00% passed)
Total Peaks identified = 457

My experimental design consisted of two individual ChIP experiments followed by sequencing, to give a list of peaks with a high degree of confidence. Each individual peak file was then compared to its biological replicate peak file to give

three individual peak files outlining significant peaks unique to each run and significant peaks overlapping in both runs. To generate these files the HOMER script mergePeaks was used with the parameter –d 100 indicating that peaks within the distance of 100 base pairs will be reported as a merged single peak.

1. Run the mergePeaks program found in the bin file of the downloaded HOMER program on the two or more peak files of interest: mergePeaks –d 100 <peak file 1> <peak file 2> -prefix <prefix name for new output files>

```
GENAW80171CCAGV:bin agmason$ mergePeaks -d 100
/Users/agmason/exampleChIpseq/peaks.txt /Users/agmason/Individual\ Tag\
directories\ and\ merged\ peaks/Sgf73Myc1/peaks.txt -prefix example
        Max distance to merge: 100 bp
        Merging peaks...
        Comparing /Users/agmason/exampleChIpseq/peaks.txt (457 total) and
/Users/agmason/exampleChIpseq/peaks.txt (457 total)
        Comparing /Users/agmason/exampleChIpseq/peaks.txt (457 total) and
/Users/agmason/Individual Tag directories and merged peaks/Sgf73Myc1/peaks.txt
(507 total)
        Comparing /Users/agmason/Individual Tag directories and merged
peaks/Sgf73Myc1/peaks.txt (507 total) and
/Users/agmason/exampleChIpseq/peaks.txt (457 total)
        Comparing /Users/agmason/Individual Tag directories and merged
peaks/Sgf73Myc1/peaks.txt (507 total) and /Users/agmason/Individual Tag directories
and merged peaks/Sgf73Myc1/peaks.txt (507 total)
/Users/agmason/exampleChIpseq/peaks.txt         /Users/agmason/Individual Tag
directories and merged peaks/Sgf73Myc1/peaks.txt         Total     Name
        X       118     /Users/agmason/Individual Tag directories and merged
peaks/Sgf73Myc1/peaks.txt
X               68      /Users/agmason/exampleChIpseq/peaks.txt
X       X       389
        /Users/agmason/exampleChIpseq/peaks.txt|/Users/agmason/Individual Tag
directories and merged peaks/Sgf73Myc1/peaks.txt
```

The output from the mergePeaks scripts generates a file with the average peak location, which is indicated by a chromosome number Scchr followed by a Arabic number and the bases on the chromosome that the peak spans. These peaks can be

annotated by hand by going to the genomic region indicated by the peak and looking for the closest genes to that peak. For other downstream applications such as annotating peaks through the HOMER program annotatePeaks.pl or uploading peaks to the UCSC genome browser, the chromosome numbers once again need to be converted to the form chr followed by a roman numeral. To do this conversion on the merged peaks the python script listed below was run:

Python renamer script 32 lines:

```
1       import sys
2
3       infile = sys.argv[1]
4
5
6       renamer = {
7       "Scchr01" : "chrI",
8       "Scchr02" : "chrII",
9       "Scchr03" : "chrIII",
10      "Scchr04" : "chrIV",
11      "Scchr05" : "chrV",
12      "Scchr06" : "chrVI",
13      "Scchr07" : "chrVII",
14      "Scchr08" : "chrVIII",
15      "Scchr09" : "chrIX",
16      "Scchr10" : "chrX",
17      "Scchr11" : "chrXI",
18      "Scchr12" : "chrXII",
19      "Scchr13" : "chrXIII",
20      "Scchr14" : "chrXIV",
21      "Scchr15" : "chrXV",
22      "Scchr16" : "chrXVI",
23
24      }
25
26      file_handle = open(infile)
27      print file_handle.next().strip()
28      for line in file_handle:
29          line = line.strip().split()
30          line[1] = renamer[line[1]]
31          print "\t".join(line)
32
```

This script is run through terminal command line as follows: python renamer.py <merged peak file> > <new file name post running script>

```
GENAW80171CCAGV:~ agmason$ python renamer.py
/Users/agmason/exampleChIpseq/example__Users_agmason_Individual\ Tag\
directories\ and\ merged\ peaks_Sgf73Myc1_peaks.txt>
exampleChIPseq_peaks_renamed.txt
```

After running this script a new .txt file is generated with corrected chromosome names. The HOMER program annotatePeaks.pl can then be run to search for the closest 5' transcription start site (TSS) to the peak. The annotatePeaks program is useful for a very large number of peaks but it should be noted that many peaks are close to the TSS and in the promoter region for multiple genes and therefore some information is lost by using the annotatePeaks.pl program.

1. Run the annotatePeaks.pl program found in the bin file of the downloaded HOMER program on the corrected merged peaks file: annotatePeaks.pl <renamed merged peak file> <genome peaks are from> > <desired name of annotated file.txt>

```
GENAW80171CCAGV:~ agmason$ annotatePeaks.pl
exampleChIPseq_peaks_renamed.txt sacCer3 > annnotatedexample_peaks.txt
Peak file = exampleChIPseq_peaks_renamed.txt
Genome = sacCer3
Organism = yeast
Peak/BED file conversion summary:
        BED/Header formatted lines: 0
        peakfile formatted lines: 389
Peak File Statistics:
        Total Peaks: 389
        Redundant Peak IDs: 0
        Peaks lacking information: 0 (need at least 5 columns per peak)
        Peaks with misformatted coordinates: 0 (should be integer)
        Peaks with misformatted strand: 0 (should be either +/- or 0/1)
Peak file looks good!
Reading Positions...
Finding Closest TSS...
Annotating:.....
        Annotation      Number of peaks         Total size (bp)   Log2 Enrichment
```

```
        TTS     8.0     2928641         -3.561
        Exon    2.0     3365692         -5.762
        Intron  0.0     11601   -8.604
        Intergenic      4.0     197523  -0.671
        Promoter        375.0   5561076         1.064
NOTE: If this part takes more than 2 minutes, there is a good chance
        your machine ran out of memory: consider hitting ctrl+C and rerunning
        the command with "-noann"
Annotating:.....
        Annotation      Number of peaks         Total size (bp)   Log2 Enrichment
        TTS     8.0     2928641         -3.561
        Exon    2.0     3365692         -5.762
        Intron  0.0     11601   -8.604
        Intergenic      4.0     197523  -0.671
        Promoter        375.0   5561076         1.064
Counting Tags in Peaks from each directory...
Organism: yeast
Loading Gene Informaiton...
Outputing Annotation File...
Done annotating peaks file
```

The annotated .txt file can be opened in Microsoft Excel for further data analysis carried out. This annotated file gives a list of genes that have occupancy by our protein of interest.

To find pathways which were overrepresented in the gene list obtained from the occupancy data I used the program DAVID[59,60] (The **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery v6.7) http://david.abcc.ncifcrf.gov/. For visualization of GO terms WebGestalt [61,62] was used used (http://bioinfo.vanderbilt.edu/webgestalt/) to look into GO annotations information is also available through DAVID but I prefer the output from WebGestalt as its output is relayed in a more organized manner.

Additional analysis can be done using HOMER including looking for enriched motifs present among ChIP-seq peaks by using the program findMotifsGenome.pl

1. Run the findMotifsGenome.pl program found in the bin file of the downloaded HOMER program on the merged peaks corrected peak file: findMotifsGenome.pl The general code to run this script is: findMotifsGenome.pl<peak file> <genome> <output directory> -size # I used –size 100 meaning the program will analyze a 100bp region at the peak location.

```
GENAW80171CCAGV:~ agmason$ findMotifsGenome.pl
/Users/agmason/exampleChIpseq/exampleChIPseq_peaks_renamed.txt  sacCer3
examplemotifs -size 100

 Position file = /Users/agmason/exampleChIpseq/exampleChIPseq_peaks_renamed.txt
 Genome = sacCer3
 Output Directory = examplemotifs
 Fragment size set to 100
 Found mset for "yeast", will check against yeast motifs
 Peak/BED file conversion summary:
        BED/Header formatted lines: 0
        peakfile formatted lines: 118
Peak File Statistics:
        Total Peaks: 118
        Redundant Peak IDs: 0
        Peaks lacking information: 0 (need at least 5 columns per peak)
        Peaks with misformatted coordinates: 0 (should be integer)
        Peaks with misformatted strand: 0 (should be either +/- or 0/1)
Peak file looks good!
 Background files for 100 bp fragments found.
 Extracting sequences from file:
/users/agmason/homer//data/genomes/sacCer3///genome.fa
 Looking for peak sequences in a single file
(/users/agmason/homer//data/genomes/sacCer3///genome.fa)
 Extracting 2 sequences from chrI
 Extracting 13 sequences from chrII
 Extracting 4 sequences from chrIII
 Extracting 11 sequences from chrIV
 Extracting 4 sequences from chrIX
 Extracting 8 sequences from chrV
 Extracting 1 sequences from chrVI
 Extracting 11 sequences from chrVII
 Extracting 3 sequences from chrVIII
 Extracting 7 sequences from chrX
 Extracting 5 sequences from chrXI
 Extracting 12 sequences from chrXII
 Extracting 11 sequences from chrXIII
```

Extracting 7 sequences from chrXIV
Extracting 11 sequences from chrXV
Extracting 8 sequences from chrXVI
Not removing redundant sequences
Sequences processed:
Auto detected maximum sequence length of 101 bp
118 total
Frequency Bins: 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.6 0.7 0.8

Freq	Bin	Count
0.3	2	1
0.35	3	10
0.4	4	24
0.45	5	24
0.5	6	27
0.6	7	29
0.7	8	3

Total sequences set to 50000
Choosing background that matches in CpG/GC Content...

Bin	# Targets	# Background	Background Weight
2	1	423	0.999
3	10	4227	1.000
4	24	10145	1.000
5	24	19471	0.521
6	27	11544	0.989
7	29	3920	3.127
8	3	151	8.398

Assembling sequence file...
Normalizing lower order oligos using homer2
Reading input files...
49999 total sequences read
Autonormalization: 1-mers (4 total)

A	27.58%	27.69%	0.996
C	22.42%	22.31%	1.005
G	22.42%	22.31%	1.005
T	27.58%	27.69%	0.996

Autonormalization: 2-mers (16 total)

AA	10.28%	8.68%	1.184
CA	6.01%	6.92%	0.869
GA	6.00%	6.31%	0.951
TA	5.22%	5.78%	0.903
AC	5.24%	5.66%	0.927
CC	5.50%	5.20%	1.058
GC	5.67%	5.16%	1.099
TC	6.00%	6.31%	0.951
AG	5.73%	6.16%	0.931
CG	5.22%	4.04%	1.292
GG	5.50%	5.20%	1.058
TG	6.01%	6.92%	0.869
AT	6.36%	7.19%	0.884

```
        CT      5.73%   6.16%   0.931
        GT      5.24%   5.66%   0.927
        TT      10.28%  8.68%   1.184
Autonormalization: 3-mers (64 total)
Normalization weights can be found in file: examplemotifs/seq.autonorm.tsv
Converging on autonormalization solution:
.....................................................................
Final normalization:    Autonormalization: 1-mers (4 total)
        A       27.58%  27.72%  0.995
        C       22.42%  22.28%  1.006
        G       22.42%  22.28%  1.006
        T       27.58%  27.72%  0.995
Autonormalization: 2-mers (16 total)
        AA      10.28%  9.84%   1.044
        CA      6.01%   6.26%   0.961
        GA      6.00%   6.15%   0.975
        TA      5.22%   5.45%   0.958
        AC      5.24%   5.32%   0.985
        CC      5.50%   5.33%   1.031
        GC      5.67%   5.48%   1.035
        TC      6.00%   6.15%   0.975
        AG      5.73%   5.92%   0.969
        CG      5.22%   4.78%   1.093
        GG      5.50%   5.33%   1.031
        TG      6.01%   6.26%   0.961
        AT      6.36%   6.63%   0.959
        CT      5.73%   5.92%   0.969
        GT      5.24%   5.32%   0.985
        TT      10.28%  9.84%   1.044
Autonormalization: 3-mers (64 total)
Finished preparing sequence/group files
------------------------------------------------------------
Known motif enrichment
Reading input files...
49999 total sequences read
11 motifs loaded
Cache length = 11180
Using binomial scoring
Checking enrichment of 11 motif(s)
|0%                          50%                         100%|
==============================================================
Preparing HTML output with sequence logos...
        1 of 11 (1e-3) RLR1?/SacCer-Promoters/Homer
        2 of 11 (1e-2) GAGA-repeat/SacCer-Promoters/Homer
------------------------------------------------------------
De novo motif finding (HOMER)
Scanning input files...
Parsing sequences...
|0%                          50%                         100%|
```

```
==========================================================
Total number of Oligos: 32896
Autoadjustment for sequence coverage in background: 1.00x
Oligos: 32896 of 34497 max
Tree  : 66988 of 172485 max
Optimizing memory usage...
Cache length = 11180
Using binomial scoring
Global Optimization Phase: Looking for enriched oligos with up to 2 mismatches...
Screening oligos 32896 (allowing 0 mismatches):
|0%                    50%                    100%|
==========================================================
        73.76% skipped, 26.24% checked (8632 of 32896), of those checked:
        73.76% not in target, 0.00% increased p-value, 0.00% high p-value
Screening oligos 32896 (allowing 1 mismatches):
|0%                    50%                    100%|
==========================================================
        73.76% skipped, 26.24% checked (8632 of 32896), of those checked:
        0.00% not in target, 18.87% increased p-value, 11.75% high p-value
Screening oligos 32896 (allowing 2 mismatches):
|0%                    50%                    100%|
==========================================================
        94.44% skipped, 5.56% checked (1829 of 32896), of those checked:
        0.00% not in target, 4.52% increased p-value, 0.00% high p-value
Reading input files...
49999 total sequences read
Cache length = 11180
Using binomial scoring
Local Optimization Phase:
1 of 25 Initial Sequence: TTTTTTTT... (-25.574)
        Round 1: -35.97 TTTTTTTT T:161.0(74.59%),B:18722.4(37.39%),P:1e-15
        Round 2: -35.97 TTTTTTTT T:161.0(74.59%),B:18722.4(37.39%),P:1e-15
        =Final=: -12.98 TTTTTTTT T:27.0(22.88%),B:3450.6(8.63%),P:1e-5
        Performing exhaustive masking of motif...
        Reprioritizing potential motifs...
2 of 25 Initial Sequence: CCCGTTTC... (-10.441)
        Round 1: -10.44 CCCGTTTC T:28.0(21.20%),B:3672.1(8.78%),P:1e-4
        Round 2: -10.44 CCCGTTTC T:28.0(21.20%),B:3672.1(8.78%),P:1e-4
        =Final=: -12.68 CCCGTTTC T:27.0(22.88%),B:3506.1(8.77%),P:1e-5
        Performing exhaustive masking of motif...
        Reprioritizing potential motifs...
3 of 25 Initial Sequence: AAGGGAAG... (-9.240)
        Round 1: -12.67 AAGGGAAG T:40.0(28.85%),B:5468.6(12.78%),P:1e-5
        Round 2: -13.99 AAGGGAAG T:19.0(14.93%),B:1424.2(3.50%),P:1e-6
        Round 3: -13.99 AAGGGAAG T:19.0(14.93%),B:1424.2(3.50%),P:1e-6
        =Final=: -14.08 AAGGGAAG T:17.0(14.41%),B:1389.7(3.48%),P:1e-6
        Performing exhaustive masking of motif...
        Reprioritizing potential motifs...
4 of 25 Initial Sequence: CACTGCGC... (-8.244)
```

```
        Round 1: -8.24 CACTGCGC T:106.0(59.43%),B:22476.1(43.00%),P:1e-3
        Round 2: -8.24 CACTGCGC T:106.0(59.43%),B:22476.1(43.00%),P:1e-3
        =Final=: -6.78 CACTGCGC T:63.0(53.39%),B:15628.5(39.09%),P:1e-2
        Performing exhaustive masking of motif...
        Reprioritizing potential motifs...
5 of 25 Initial Sequence: GAGCGGGG... (-8.211)
        Round 1: -14.09 GAGCGGGG T:13.0(10.47%),B:672.4(1.67%),P:1e-6
        Round 2: -14.09 GAGCGGGG T:13.0(10.47%),B:672.4(1.67%),P:1e-6
        =Final=: -15.98 GAGCGGGG T:13.0(11.02%),B:672.4(1.68%),P:1e-6
        Performing exhaustive masking of motif...
        Reprioritizing potential motifs...
Remaining seeds don't look promising (After initial 5 motifs, logp -6.638 > -8.072)
Finalizing Enrichment Statistics (new in v3.4)
Reading input files...
49999 total sequences read
Cache length = 11180
Using binomial scoring
Checking enrichment of 5 motif(s)
|0%                    50%                    100%|
=================================================================
Output in file: examplemotifs/homerMotifs.motifs8
Scanning input files...
Parsing sequences...
|0%                    50%                    100%|
=================================================================
Total number of Oligos: 511667
Autoadjustment for sequence coverage in background: 1.00x
Oligos: 511667 of 551887 max
Tree  : 1043324 of 2759435 max
Optimizing memory usage...
Cache length = 11180
Using binomial scoring
Global Optimization Phase: Looking for enriched oligos with up to 2 mismatches...
Screening oligos 511667 (allowing 0 mismatches):
|0%                    50%                    100%|
=================================================================
        97.97% skipped, 2.03% checked (10380 of 511667), of those checked:
        97.97% not in target, 0.00% increased p-value, 0.00% high p-value
Screening oligos 511667 (allowing 1 mismatches):
|0%                    50%                    100%|
=================================================================
        97.97% skipped, 2.03% checked (10380 of 511667), of those checked:
        0.00% not in target, 1.95% increased p-value, 0.00% high p-value
Screening oligos 511667 (allowing 2 mismatches):
|0%                    50%                    100%|
=================================================================
        99.92% skipped, 0.08% checked (424 of 511667), of those checked:
        0.00% not in target, 0.03% increased p-value, 0.00% high p-value
Reading input files...
```

```
49999 total sequences read
Cache length = 11180
Using binomial scoring
Local Optimization Phase:
1 of 25 Initial Sequence: TTTTTTTTTT... (-45.201)
        Round 1: -47.10 TTTTTTTTTT T:101.0(57.67%),B:8160.9(18.46%),P:1e-20
        Round 2: -47.10 TTTTTTTTTT T:101.0(57.67%),B:8160.9(18.46%),P:1e-20
        =Final=: -15.02 TTTTTTTTTT T:25.0(21.19%),B:2706.0(6.77%),P:1e-6
        Performing exhaustive masking of motif...
        Reprioritizing potential motifs...
2 of 25 Initial Sequence: TGGGAAAGGG... (-12.541)
        Round 1: -27.52 TGGGAAAGGG T:22.0(17.08%),B:866.2(2.14%),P:1e-11
        Round 2: -27.52 TGGGAAAGGG T:22.0(17.08%),B:866.2(2.14%),P:1e-11
        =Final=: -27.65 TGGGAAAGGG T:20.0(16.95%),B:850.3(2.13%),P:1e-12
        Performing exhaustive masking of motif...
        Reprioritizing potential motifs...
3 of 25 Initial Sequence: CCCAAAAGGG... (-9.290)
        Round 1: -9.29 CCCAAAAGGG T:32.0(23.84%),B:4755.0(11.21%),P:1e-4
        Round 2: -9.29 CCCAAAAGGG T:32.0(23.84%),B:4755.0(11.21%),P:1e-4
        =Final=: -4.48 CCCAAAAGGG T:5.0(4.24%),B:453.0(1.13%),P:1e-1
        Performing exhaustive masking of motif...
        Reprioritizing potential motifs...
4 of 25 Initial Sequence: ACACGAAATG... (-9.143)
        Round 1: -18.96 ACACGAAATG T:12.0(9.71%),B:336.2(0.84%),P:1e-8
        Round 2: -24.69 ACACGAAATG T:13.0(10.47%),B:254.6(0.63%),P:1e-10
        Round 3: -25.74 ACAGGAAAKG T:12.0(9.71%),B:175.0(0.44%),P:1e-11
        Round 4: -25.74 ACAGGAAAKG T:12.0(9.71%),B:175.0(0.44%),P:1e-11
        =Final=: -25.74 ACAGGAAAKG T:11.0(9.32%),B:174.6(0.44%),P:1e-11
        Performing exhaustive masking of motif...
        Reprioritizing potential motifs...
5 of 25 Initial Sequence: TCTTCTAGAA... (-8.264)
        Round 1: -17.43 TCTTCTAGAA T:9.0(7.37%),B:157.8(0.39%),P:1e-7
        Round 2: -18.59 TATTCTAGAA T:8.0(6.58%),B:86.7(0.21%),P:1e-8
        Round 3: -18.59 TATTCTAGAA T:8.0(6.58%),B:86.7(0.21%),P:1e-8
        =Final=: -12.65 TATTCTAGAA T:5.0(4.24%),B:74.6(0.19%),P:1e-5
        Performing exhaustive masking of motif...
        Reprioritizing potential motifs...
Remaining seeds don't look promising (After initial 5 motifs, logp -7.976 > -8.072)
Finalizing Enrichment Statistics (new in v3.4)
Reading input files...
49999 total sequences read
Cache length = 11180
Using binomial scoring
Checking enrichment of 5 motif(s)
|0%                          50%                         100%|
================================================================
Output in file: examplemotifs/homerMotifs.motifs10
-blen automatically set to 2
Scanning input files...
```

```
Parsing sequences...
|0%                    50%                   100%|
==========================================================================
Total number of Oligos: 2924748
Autoadjustment for sequence coverage in background: 1.00x
Oligos: 2924748 of 4499369 max
Tree  : 6971548 of 22496845 max
Optimizing memory usage...
Cache length = 11180
Using binomial scoring
Global Optimization Phase: Looking for enriched oligos with up to 2 mismatches...
Screening oligos 2924748 (allowing 0 mismatches):
|0%                    50%                   100%|
==========================================================================
        99.64% skipped, 0.36% checked (10426 of 2924748), of those checked:
        99.64% not in target, 0.00% increased p-value, 0.00% high p-value
Screening oligos 2924748 (allowing 1 mismatches):
|0%                    50%                   100%|
==========================================================================
        99.64% skipped, 0.36% checked (10426 of 2924748), of those checked:
        0.00% not in target, 0.35% increased p-value, 0.00% high p-value
Screening oligos 2924748 (allowing 2 mismatches):
|0%                    50%                   100%|
==========================================================================
        100.00% skipped, 0.00% checked (142 of 2924748), of those checked:
        0.00% not in target, 0.00% increased p-value, 0.00% high p-value
Reading input files...
49999 total sequences read
Cache length = 11180
Using binomial scoring
Local Optimization Phase:
1 of 25 Initial Sequence: TAAAAAAAAAAA... (-41.912)
        Round 1: -52.66 TAAAAAAAAAAA
T:119.0(63.68%),B:9349.4(20.85%),P:1e-22
        Round 2: -52.66 TAAAAAAAAAAA
T:119.0(63.68%),B:9349.4(20.85%),P:1e-22
        =Final=: -16.15 TAAAAAAAAAAA T:27.0(22.88%),B:2933.7(7.34%),P:1e-
7
        Performing exhaustive masking of motif...
        Reprioritizing potential motifs...
2 of 25 Initial Sequence: TTGGGAAAGGGA... (-12.503)
        Round 1: -21.00 TTGGGAAAGGGA T:5.0(4.17%),B:5.2(0.01%),P:1e-9
        Round 2: -21.00 TTGGGAAAGGGA T:5.0(4.17%),B:5.2(0.01%),P:1e-9
        =Final=: -25.97 TTGGGAAAGGGA T:5.0(4.24%),B:5.2(0.01%),P:1e-11
        Performing exhaustive masking of motif...
        Reprioritizing potential motifs...
3 of 25 Initial Sequence: CCATTTCTCATG... (-10.205)
        Round 1: -23.95 CCAYTTCTCATG T:13.0(10.47%),B:271.5(0.68%),P:1e-
10
```

```
        Round 2: -23.95 CCAYTTCTCATG T:13.0(10.47%),B:271.5(0.68%),P:1e-
10
        =Final=: -26.80 CCAYTTCTCATG T:13.0(11.02%),B:271.1(0.68%),P:1e-11
        Performing exhaustive masking of motif...
        Reprioritizing potential motifs...
 4 of 25 Initial Sequence: ACATTCCGACAA... (-8.209)
        Round 1: -15.56 ACATTCCGWCAA T:7.0(5.78%),B:81.3(0.20%),P:1e-6
        Round 2: -15.56 ACATTCCGWCAA T:7.0(5.78%),B:81.3(0.20%),P:1e-6
        =Final=: -15.86 ACATTCCGWCAA T:6.0(5.08%),B:76.4(0.19%),P:1e-6
        Performing exhaustive masking of motif...
        Reprioritizing potential motifs...
 5 of 25 Initial Sequence: GTGCCATAGCAA... (-7.406)
        Round 1: -17.22 GTACCATTGCAA T:4.0(3.35%),B:3.8(0.01%),P:1e-7
        Round 2: -17.22 GTACCATTGCAA T:4.0(3.35%),B:3.8(0.01%),P:1e-7
        =Final=: -22.14 GTACCATTGCAA T:4.0(3.39%),B:3.8(0.01%),P:1e-9
        Performing exhaustive masking of motif...
        Reprioritizing potential motifs...
Remaining seeds don't look promising (After initial 5 motifs, logp -7.406 > -8.072)
Finalizing Enrichment Statistics (new in v3.4)
Reading input files...
49999 total sequences read
Cache length = 11180
Using binomial scoring
Checking enrichment of 5 motif(s)
|0%                          50%                         100%|
=================================================================
Output in file: examplemotifs/homerMotifs.motifs12
(Motifs in homer2 format)
Determining similar motifs... 15 reduced to 10 motifs
Outputing HTML and sequence logos for motif comparison...
        1 of 10 (1e-11) similar to MA0282.1_CEP3/Jaspar
        2 of 10 (1e-11) similar to MA0401.1_SWI4/Jaspar
        3 of 10 (1e-9) similar to RFX1/RFX1_YPD/[](Harbison)/Yeast
        4 of 10 (1e-7) similar to RLR1?/SacCer-Promoters/Homer
        5 of 10 (1e-6) similar to MA0406.1_TEC1/Jaspar
        6 of 10 (1e-5) similar to MA0305.1_GCR2/Jaspar
        7 of 10 (1e-5) similar to MA0431.1_YML081W/Jaspar
        8 of 10 (1e-5) similar to MA0266.1_ABF2/Jaspar
        9 of 10 (1e-3) similar to SUT1?/SacCer-Promoters/Homer
        10 of 10 (1e-1) similar to NDD1/NDD1_YPD/14-MCM1(Harbison)/Yeast
Job finished - if results look good, please send beer to ..
Cleaning up tmp files...
```

Methods Table 3-1: Primers used for ChIP-PCR occupancy validation

Primer Name	Sequence	Occupancy Target
chrVII 976969 -F	ACCCCGATTACCCCTCATCT	5' FMP43
chrVII 976969 -R	TTGAGCCTTTCGGTTTTGCG	
chrXIV 559344-F	ATGAGTGACCTATGACGCCC	5' IDH1
chrXIV 559344-F	GCGCCGAGTTTAGGTGAATC	
chrV:153026- F	ACAACCAAGGCATCCCACTT	5' MNN1
chrV:153026- R	GCACGCACTTTCCACCATTT	
chrII:697587- F	TTGGTGCACACACAAAGCTC	5' YBR238
chrII:697587- R	CCTGCAGTGAACTGCTCGTA	
chrXIII:754426 F	GCAGGGTTTAGAGGGGAACG	5' RPL20A
chrXIII:754426 R	TAAGCTCACTTCGGTGTGGT	
chrXIII:124351 F	GAGAAGCCCTCTCCAGAACG	5' RPL6A
chrXIII:124351 F	AACCAGATTCGAAAAGCGGC	
chr16:645466- F	GTGCGTGCGTGATTGGTAAT	5' YPR036W-A
chr16:645466- R	CTGCGGAGGGTGGTTTTC	
chrVII:772110- F	CCGGAAAATTGCGACACT	5' BTN2 and 3' VPS62
chrVII:772110- R	ATCCTGGGCGGCTGTTAG	
chrXIII:388445- F	TTACTTTCCGGGTGCGAAT	5' FET3 and 5' AAC1
chrXIII:388445- R	CAAATGGGTGCACTTTTGAA	
chrXII:97654- F	ATTTCGTCAAAACCTCGAAGC	5' POM33
chrXII:97654- R	CCGAGAAGTTCTTCCGATTCAC	
chrVII:310724- F	CCTATTTTTCCTCTGGCGCGT	5' RPL28
chrVII:310724- R	ACAAGCGAAAAACTGCGAGG	
chrXII:282717- F	CCCTCCGAAACTAGTTAGCACA	5' RPL10
chrXII:282717- R	TGGTTGCTCCACATCAACATC	
chrIV:117378- F	GGTGAACCCCACGACAACTT	5' RPL35A
chrIV:117378- R	GATGCTGGAATCCTGCCCAA	
off target chrVIII F	AGGCCATCCCATCATTACGAG	Off target
off target chrVIII R	TGCAGTTTCTTAAGATCACCGTT	
off target chrXII F	AGAGGCCCATCAGCATCCTA	Off target
off target chrXII R	CAATTCTTGGTCGGGGACGA	

Methods Table 3-2: Strains used in studies

Strain number	Genotype	MAT
LPy6494	WT	α
LPy18835	*sgf73Δ*	α
LPy18851	*ubp8Δ*	α
LPy18905	Ubp8-13Myc	α
LPy18909	Sgf73-13Myc #4	a
LPy18910	Sgf73-13Myc #5	α
LPy18919	*sir2Δ fob1Δ*	α
LPy18899	Sir2OE	α
LPy17446	*pdr5Δ*	α
LPy20489	*pdr5Δ sgf73Δ*	α
LPy20496	Idh1-GFP	α
LPy20497	Idh1-GFP *sgf73Δ*	α
LPy20487	Fmp43-GFP	α
LPy20486	Fmp43-GFP *sgf73Δ*	α
LPy20499	Mnn1-GFP	α
LPy20500	Mnn1-GFP *sgf73Δ*	α
LPy20524	Mpc2-GFP	α
LPy20525	Mpc2-GFP *sgf73Δ*	α

Methods Table 3-3: Primers used for yeast gene expression analysis by qPCR

Primer	Sequence
SSA1 F	ACAAGGACACCGTCACCAAG
SSA1 R	GGAAACCGCCTGGAGCAC
EFB1 F	GACTCTTTCCCAGCTGCCTC
EFB1 R	AGCAGCCTTCTTAGCGTTGT
EFB1 ex F	GGCATCCACCGATTTCTCCA
EFB1 ex R	AGCTTCAGCGTCAGCTTCTT
RPS23B F	TGAACTCCGCTAGAAAGTTGC
RPS23B R	TGAGAAGAACCACCGAATGGA
RPL33A F	TCAAGATCGAAGGTGTCGCT
RPL33A R	AACGACACCAGAGTTACCGT
RPL33A ex F	ATGGCTGAATCCCATAGATTGTA
RPL33A ex R	AGCGACACCTTCGATCTTGA
RPL2A F	CAACCCAGACGAAAACAAGACT
RPL2A R	TCTACCACCACCGGCAATG
RPL10 F	TGCTCGTATCTGTGCCAACA
RPL10 R	CCCCAAGCACCTCTCATACC
RPS0B F	AGAACCTACGGTCAAAGAGCTG
RPS0B R	TCCTTAATGGCCTGAGCGTC
RPS0B ex F	CGTTCAAGTCCACCAAGAACC
RPS0B ex R	CACCAGTGTGAGCAGCAAAT
RPL42A F	AGACCTGTCGTAAGCACACT
RPL42A R	AACGTCTCTTACCTTGGGCA
RPS5 F	ACAATGAACAGCCTTGAGTTCT
RPS5 R	TGCAGGGCCATTCTCATCTG
RPL6A F	CGAAGACGTCGCTGCTCTAA
RPL6A R	GGAACTAGAGAGGCACGCAA
RPL6A ex F	TGAGTGCCCAAAAAGCTCC
RPL6A ex R	GCCTTTCTGGTCTTCTTTAGAGC
RPL8A F	CCACGTTGTTGCCTTGATCG
RPL8A R	GCAACGGCAGAGGTCTTTTG
HSP78 F	GATCCCAATCAGCAACCGGA
HSP78 R	TTTACCGACACCAGCTCGAC
RPL28 ex F	GTCACGTCTCAGCCGGTAAA
RPL28 ex R	GATGTTGACCACCGGCCATA
YPR036W-A F	CCAGTCTCTCAGCCAAGCTC
YPR036W-A R	GCATCCAAAAGAGCGGCATT
BTN2 F	CGAAGGTGGCATCAACGAAC
BTN2 R	GGTTCTTTCGCTTTCTCCGC
VPS62 F	ATGGGGGTGGTAGTGCCTAT
VPS62 R	GCATGTTGACCGACTGATGC
FET3 F	TACCATTCACACACGGACGG
FET3 R	TCAAGTCCGTGACCAAGTCG
SSA2 F	GCTGCTATGAACCCAGCTAAC
SSA2 R	CACCTTGGACTTCTGGGTCA
POM33 F	TAAGTCGCTCGTGCAAAGGT
POM33 R	TGCGGGTGAAAAATTGCTGG
LST8 F	CTGCTGGCTATGACCACACA
LST8 R	TTTTGATGGCCTGCTGTTGC
SIS1 F	AATCCTCAAACGGGCCGTAG

Methods Table 3-3: Primers used for yeast gene expression analysis by qPCR continued.

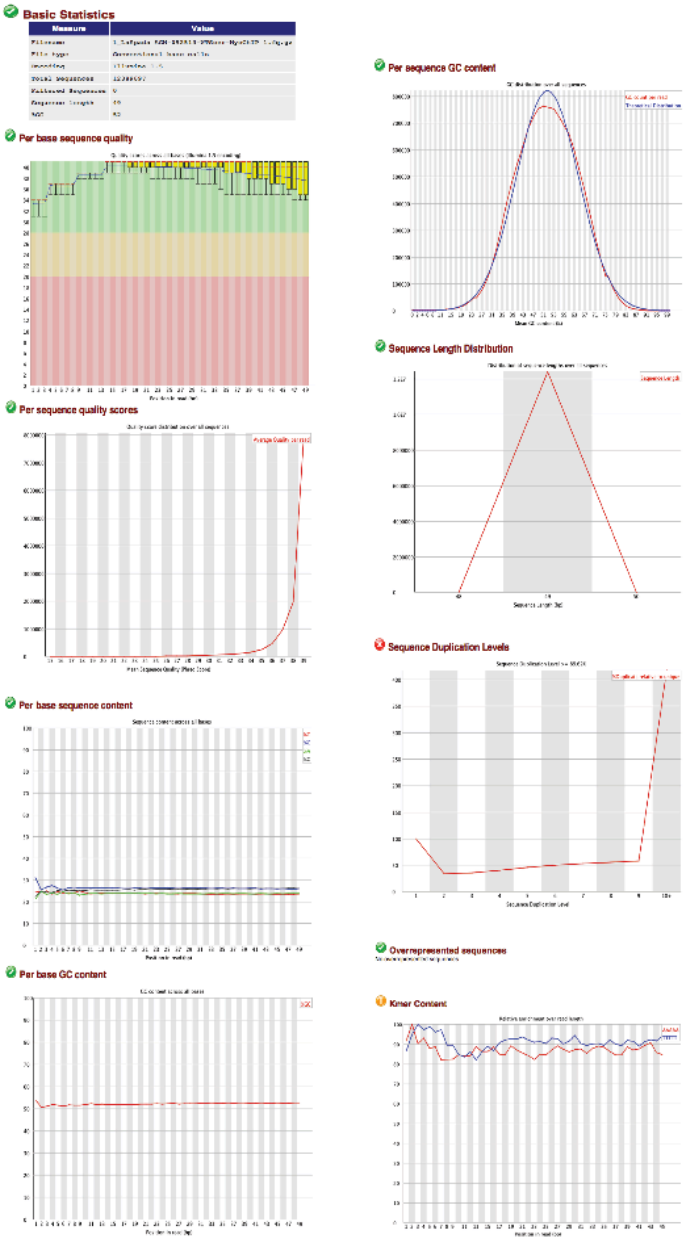
Primer	Sequence
SIS1 R	GGGTTGGACAGGCTGTACTC
RPL35A F	GGCCGGTGTTAAAGCTTACG
RPL35A R	CGGTCAAGACACAGGCGATA
RPS0A ex F	CGTTCAAGTTCACCAAGAACCA
RPS0A ex R	GCCCTTTGACCGAAAGTTCT
RPL23B ex F	CTTTAGGTTTACCAGTTGGTGCT
RPL23B ex R	AGGCGGCTGGTAATCTGTTC
RPL27B ex F	AGCAGGTAAAGTTGCTGTCGT
RPL27B ex R	TAGAACCCTCATCGTGTGGC
RPS23A ex F	AGAAGAAACAACCGTTGGGC
RPS23A ex R	GGCATGAGAAGAACCACCGA
RPS23A ex FB	TCTTTAGGTTTACCAGTTGGTGCTA
RPS23A ex RB	GGCGGCTGGTAATCTGTTCA
FES1 F	AGAACTACGTGCTGCTGCTT
FES1 R	ACGTCGAGTGGCTTTGTCTT
RPL20A F	GCTGCCAGACACAGAGCTA
RPL20A R	TGTGAGGCAATGGGAATTTCAAG
PDR5 F	ATACCACAGACCCCGAAGGA
PDR5 R	TAGCTTCACGGCTTGCTTCA
SOD1 F	TTGTAGGCAGAAGCGTCGTT
SOD1 R	GTCTTGGACCGGCATTACCA
HSF1 F	ATAATGACACCGAGCACGCA
HSF1 R	CATCTACCGTGAGGAAGGGC
SLX8 F	GTATGGTGCAGCCAAGGACT
SLX8 R	GGACAACAAAAGACGTGCCC
AHP1 F	CGGTGACGGTGTTTACTGGA
AHP1 R	CGACTGAGGAAACGGTCACA
CTR1 F	TGCTGTTGCCGATACCACTT
CTR1 R	ATTCTCAAGGATGTCGCGCT
LOT5 F	CCAACCCACAAGACACCAGT
LOT5 R	ACAAAACGCTGCTGTAGGGA
UGO1 F	ACGTCAAGGGCACAGATACG
UGO1 R	CCGTGCGGATCTACTACACC
AAC1 F NEW	GCTTCAAGAGGACTGCGACT
AAC1 R NEW	TGAAGGCAAAATTCAGCGCC
RPL28 F	GCAGCCGGTTACGGTAAGAT
RPL28 R	ACCACCAGCAGCTCTGATTT
RPL23B F	GGACAATGCTGGTGTCATCG
RPL23B R	AACACCAGAGTTGGAAGCAA
RPS0A F	GCCGGTAGATTCACTCCAGG
RPS0A R	TCAGTTAGAGCAATGACAGGGA
RPL27B F	GCCACACGATGAGGGTTCTA
RPL27B R	TTCTTGGCACCGTGCTTCTT
RPS23B F	AGTCTTCTCCATTCGGTGGT
RPS23B R	CGGCTAGCAAGACTTCGTCA
RPS23A F	GGGCCGAAAACAACTACAAGAA
RPS23A R	CCAAGACGATACCCTTGGCA
RPL2A EX F	ATGGGTAGAGTTATTCGTAACC

Methods Table 3-3: Primers used for yeast gene expression analysis by qPCR continued.

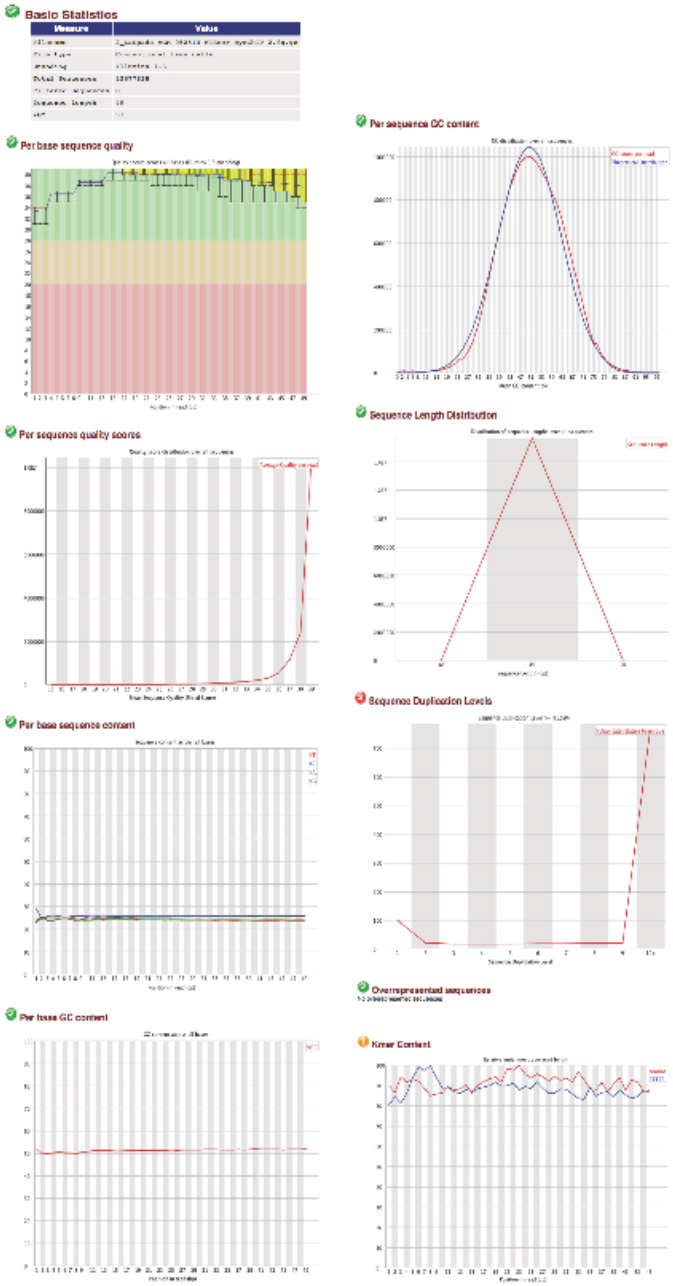
Primer	Sequence
RPL2A EX R	GCACCTTGTCTCAATCTGGT
RPL42A EX F	ATGGTTAACGTTCCAAAGACC
RPL42A EX R	AACGTCTCTTACCTTGGGCA
RPL35A EX F	ATGGCCGGTGTTAAAGCTTACG
RPL35A EX R	CAAGTCAACCAATTGAGAAGCCA
RPL20A EX F	ATGGCTCACTTTAAAGAATACC
RPL20A EX R	GTAACGAGACTTGGCAATAACTT
FMP43 F	CTGGTCTTCGCAGGGCTAAA
FMP43 R	CCAACGCGTCCAAATCAGTG
IDH1 F	GGTGACGGTGTTGGGAAAGA
IDH1 R	ACCTGTTTGGTCAGCAGGAG
MNN1 F	AATACAGCCCACTGCACACA
MNN1 R	AGCTCCGGAATACACTCCCT
YBR238C F	TTGGGCGGCCTTAAGAGATG
YBR238C R	ACAAATGCTCTGGTGCTGGA
PHO89 F	ATCGCCGTTGGTGGTATTGT
PHO89 R	ACCAAAGCGAGGAGCATTCA
MPC2 F	GCTGCTATCTACTGCGCTGA
MPC2 R	TAGCCAGCGGTCAGACAAAG
SCR1 F	CGCGGCTAGACACGGATT
SCR1 R	GCACGGTGCGGAATAGAGAA

Methods Table 3-4: Primers used for human gene expression analysis by qPCR

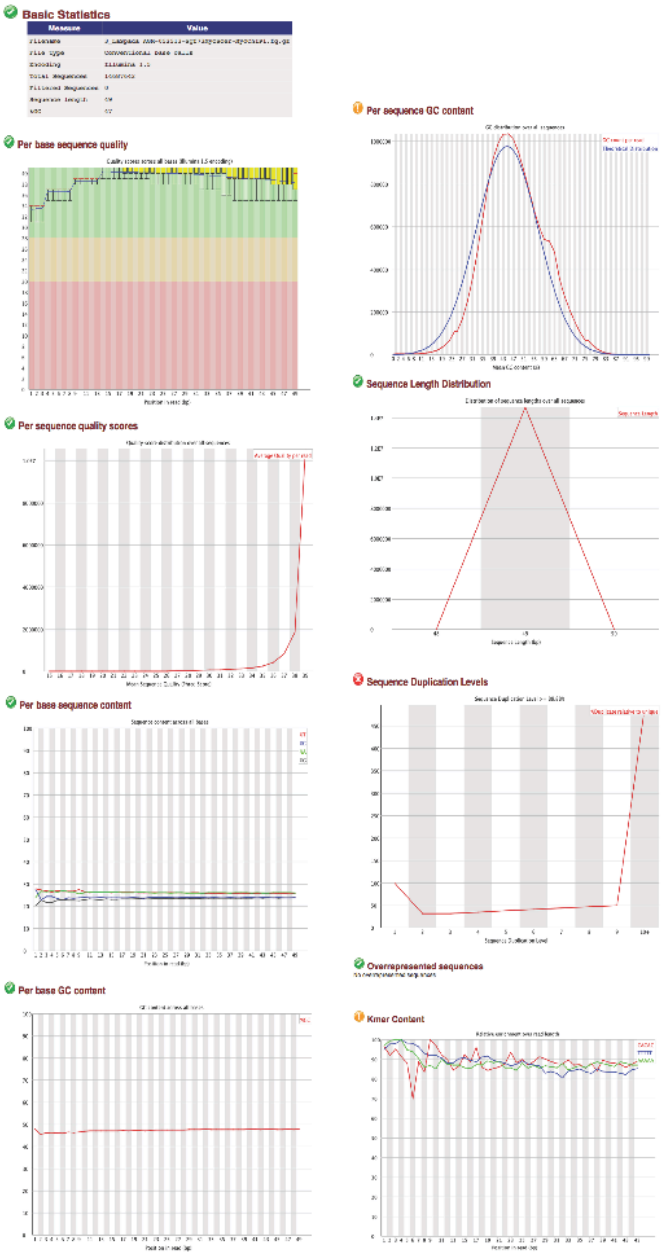
Primer	Sequence	Yeast Ortholog
TMEM33 F	GACAGCTGCCACTACCTGTT	POM33
TMEM33 R	TATTTGAGCCCCTTGCGTCA	
RPL23hu F	GATGTCGAAGCGAGGACGTG	RPL23B
RPL23hu R	CTGTTCAGCCGTCCCTTGAT	
RPL27hu F	ATGATGGCACCTCAGATCGC	RPL27B
RPL27hu R	ACAGAGTACCTTGTGGGCATT	
RPL10Lhu F	GGCGTAGTTAAGAGAGCGCA	RPL10
RPL10Lhu R	CCCTCGGCAGAAACGAGATT	
RPL8hu F	CTGACGCCGTGTTTCCTCTT	RPL2A
RPL8hu R	GGTCGTGGATGATGTCCTTGA	
RPS23hu F	AGGATGGGCAAGTGTCGTG	RPS23A/B
RPS23hu R	TAGGGCTGTGCCCAAATGAG	
MPC2 F	TTTTCCTCACGTCCCACAACA	FMP43
MPC2 R	CACTTTATCGAGGAGCCGGT	
IDH3B F	ATCAAAGTTGGCAAGGTGCG	IDH1
IDH3B R	GAGGCACAAGGTCTCTTCCC	
SLC20A2 F	AAATCAGCGTGTTCCTGCCT	PHO89
SLC20A2 R	AAAGCTGTGATGTCTCCTCAAGT	
RPL6 hu F	GGCATCCAGTCGACACCTAC	RPL6A
RPL6 hu R	TTCACCCGCCATCCTCTTTG	
RPL18A hu F	CAAGTCCCGCTTCTGGTACT	RPL20A
RPL18A hu R	ACCCATGTCTCGGTAGCACT	
RPL35 hu F	CTCTAAGATCCGAGTCGTCCG	RPL35A
RPL35 hu R	CAGGTTCTCCTCGTGCTTGT	
HEPH hu F	ACCTGCCTTGGTCTGAGTTG	FET3
HEPH hu R	CCATGGCCCACATTACTTTTGT	



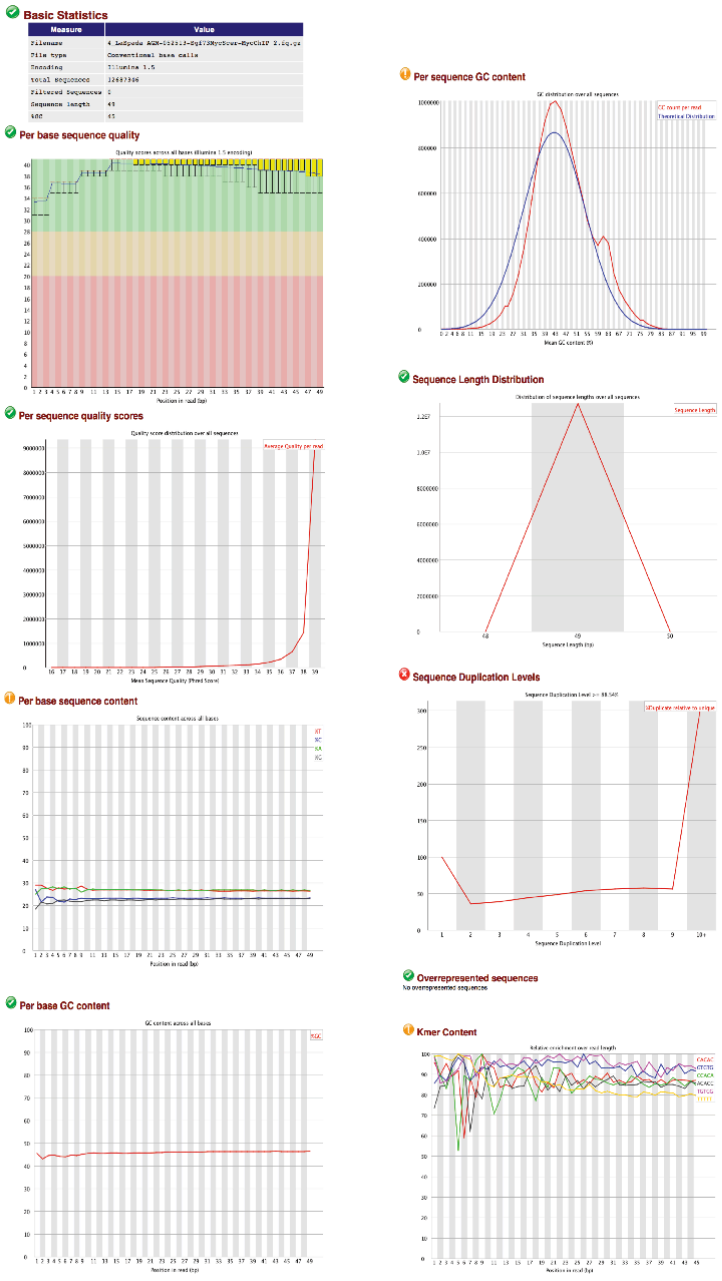
Supplementary figure 3-1: FastQC report for library 1, no tag myc ChIP #1
FastQC gives a quality control analysis for all raw reads sequenced from the library. It gives: the average read length and the quality of each base for each sequencing position; quality scores for each sequence; the percentage of each base at each sequence position; the percentage of G or C's at each base position; the G/C percentage in each sequence; the distribution of sequence length; the sequence duplication level in the library, how many times a sequence was sequenced; if there is an overrepresented sequence in the library; and 5-mer sequence enrichment in the library.



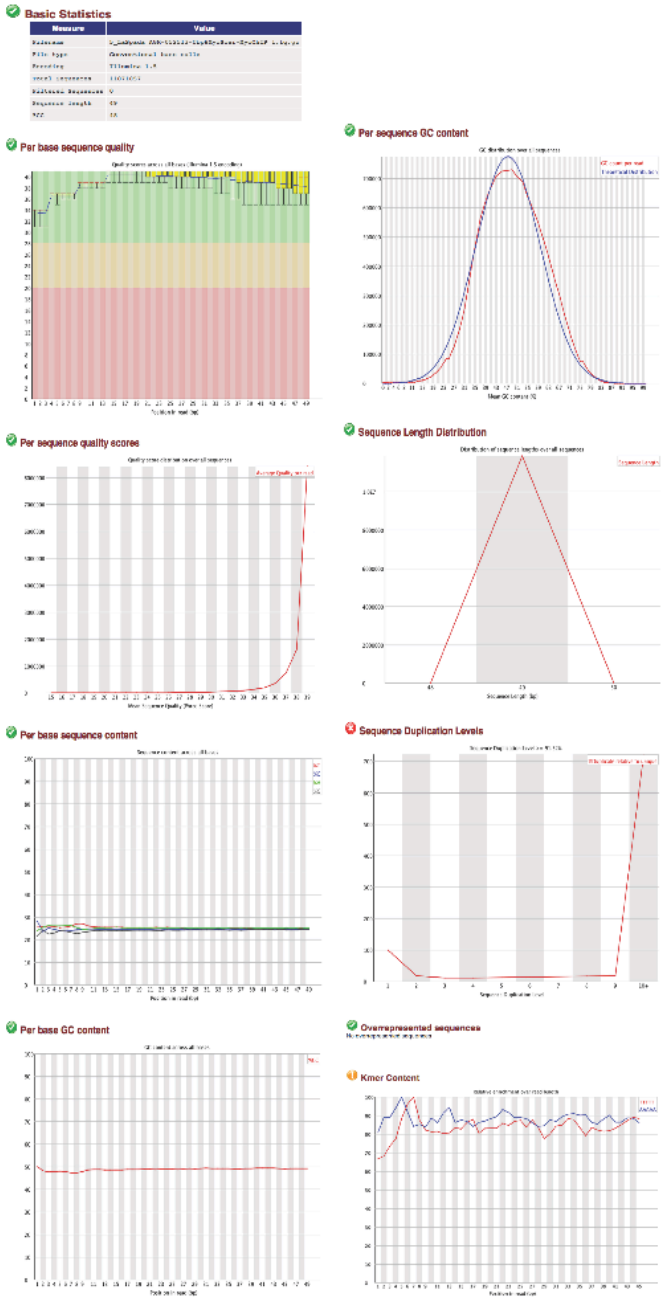
Supplementary figure 3-2: FastQC report for library 2, no tag myc ChIP #2
FastQC gives a quality control analysis for all raw reads sequenced from the library. It gives: the average read length and the quality of each base for each sequencing position; quality scores for each sequence; the percentage of each base at each sequence position; the percentage of G or C's at each base position; the G/C percentage in each sequence; the distribution of sequence length; the sequence duplication level in the library, how many times a sequence was sequenced; if there is an overrepresented sequence in the library; and 5-mer sequence enrichment in the library.



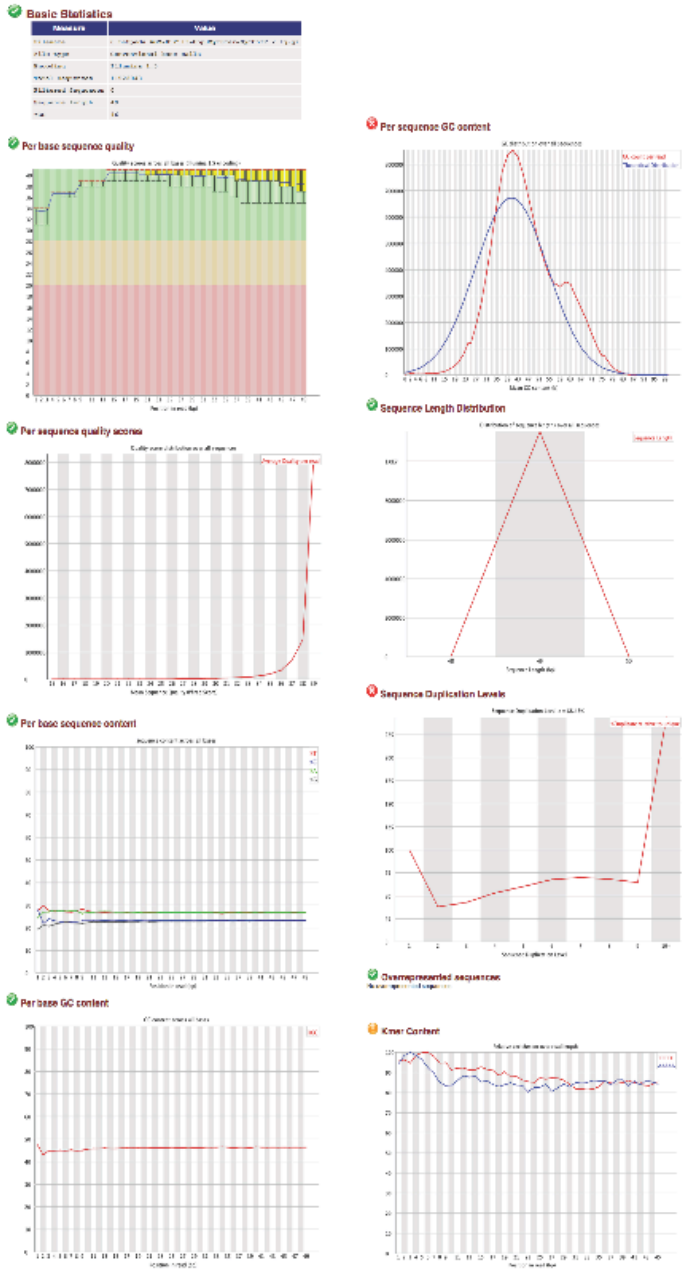
Supplementary figure 3-3: FastQC report for library 3, Sgf73-myc myc ChIP #1
FastQC gives a quality control analysis for all raw reads sequenced from the library. It gives: the average read length and the quality of each base for each sequencing position; quality scores for each sequence; the percentage of each base at each sequence position; the percentage of G or C's at each base position; the G/C percentage in each sequence; the distribution of sequence length; the sequence duplication level in the library, how many times a sequence was sequenced; if there is an overrepresented sequence in the library; and 5-mer sequence enrichment in the library.



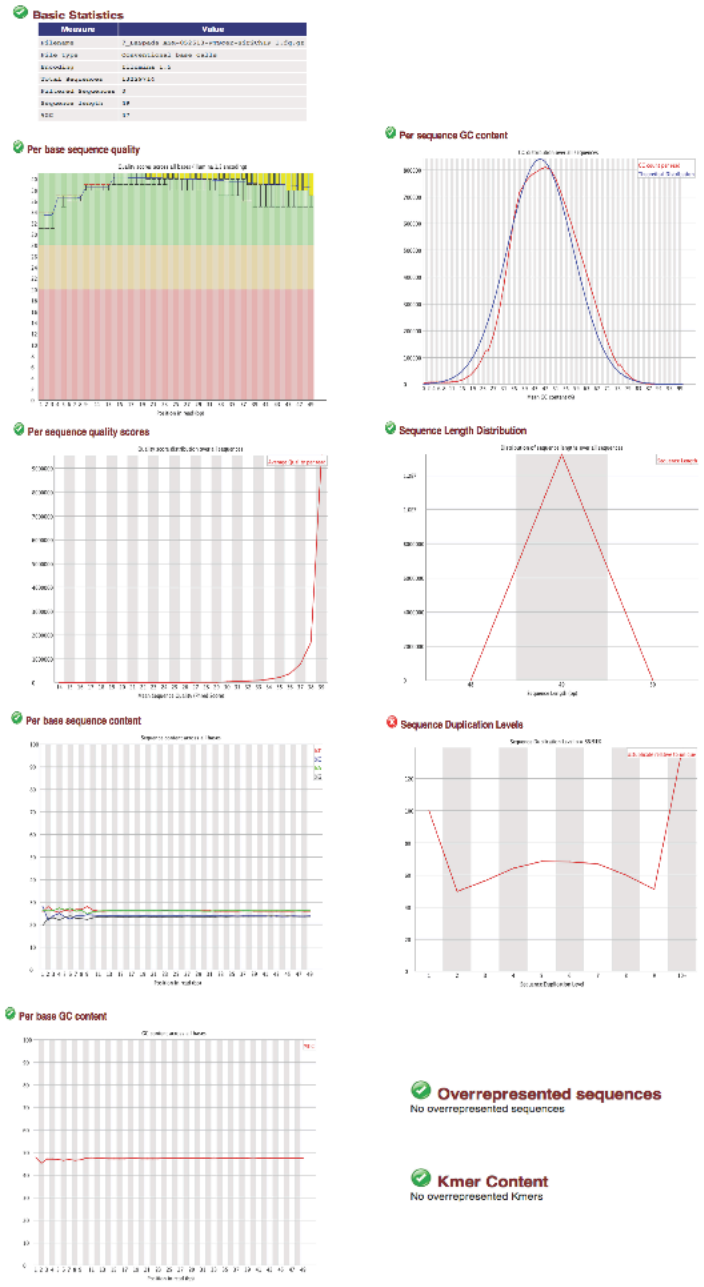
Supplementary figure 3-4: FastQC report for library 4, Sgf73-myc myc ChIP #2
FastQC gives a quality control analysis for all raw reads sequenced from the library. It gives: the average read length and the quality of each base for each sequencing position; quality scores for each sequence; the percentage of each base at each sequence position; the percentage of G or C's at each base position; the G/C percentage in each sequence; the distribution of sequence length; the sequence duplication level in the library, how many times a sequence was sequenced; if there is an overrepresented sequence in the library; and 5-mer sequence enrichment in the library.



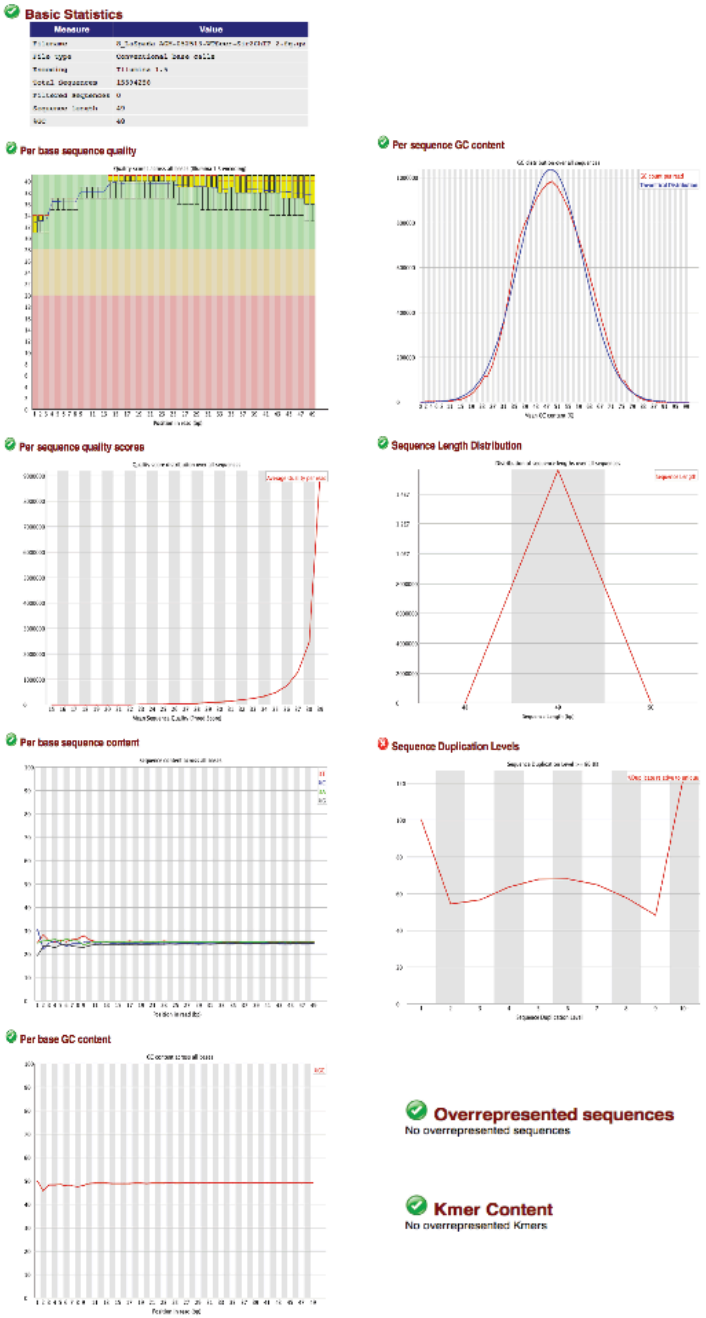
Supplementary figure 3-5: FastQC report for library 5, Ubp8-myc myc ChIP #1
FastQC gives a quality control analysis for all raw reads sequenced from the library. It gives: the average read length and the quality of each base for each sequencing position; quality scores for each sequence; the percentage of each base at each sequence position; the percentage of G or C's at each base position; the G/C percentage in each sequence; the distribution of sequence length; the sequence duplication level in the library, how many times a sequence was sequenced; if there is an overrepresented sequence in the library; and 5-mer sequence enrichment in the library.



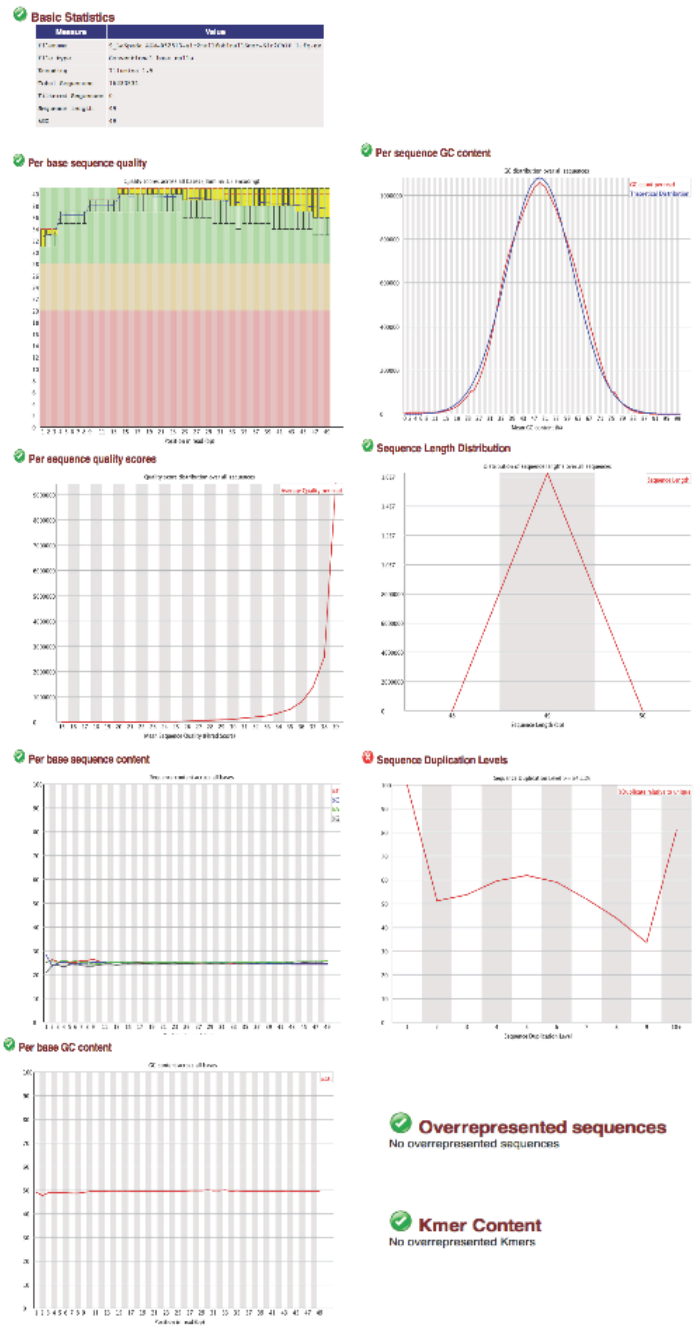
Supplementary figure 3-6: FastQC report for library 6, Ubp8-myc myc ChIP #2
FastQC gives a quality control analysis for all raw reads sequenced from the library. It gives: the average read length and the quality of each base for each sequencing position; quality scores for each sequence; the percentage of each base at each sequence position; the percentage of G or C's at each base position; the G/C percentage in each sequence; the distribution of sequence length; the sequence duplication level in the library, how many times a sequence was sequenced; if there is an overrepresented sequence in the library; and 5-mer sequence enrichment in the library.



Supplementary figure 3-7: FastQC report for library 7, WT Sir2 ChIP #1
FastQC gives a quality control analysis for all raw reads sequenced from the library. It gives: the average read length and the quality of each base for each sequencing position; quality scores for each sequence; the percentage of each base at each sequence position; the percentage of G or C's at each base position; the G/C percentage in each sequence; the distribution of sequence length; the sequence duplication level in the library, how many times a sequence was sequenced; if there is an overrepresented sequence in the library; and 5-mer sequence enrichment in the library.



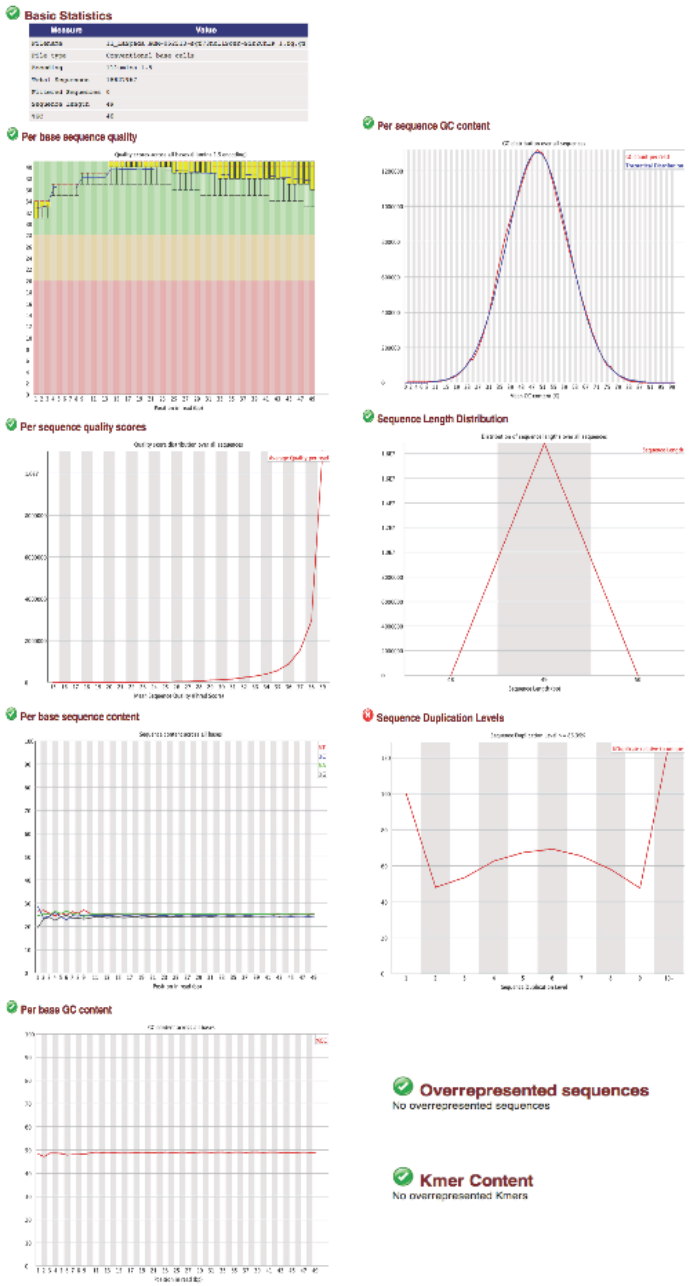
Supplementary figure 3-8: FastQC report for library 8, WT Sir2 ChIP #2
FastQC gives a quality control analysis for all raw reads sequenced from the library. It gives: the average read length and the quality of each base for each sequencing position; quality scores for each sequence; the percentage of each base at each sequence position; the percentage of G or C's at each base position; the G/C percentage in each sequence; the distribution of sequence length; the sequence duplication level in the library, how many times a sequence was sequenced; if there is an overrepresented sequence in the library; and 5-mer sequence enrichment in the library.



Supplementary figure 3-9: FastQC report for library 9, *sir2Δ fob1Δ* Sir2 ChIP #1
FastQC gives a quality control analysis for all raw reads sequenced from the library. It gives: the average read length and the quality of each base for each sequencing position; quality scores for each sequence; the percentage of each base at each sequence position; the percentage of G or C's at each base position; the G/C percentage in each sequence; the distribution of sequence length; the sequence duplication level in the library, how many times a sequence was sequenced; if there is an overrepresented sequence in the library; and 5-mer sequence enrichment in the library.

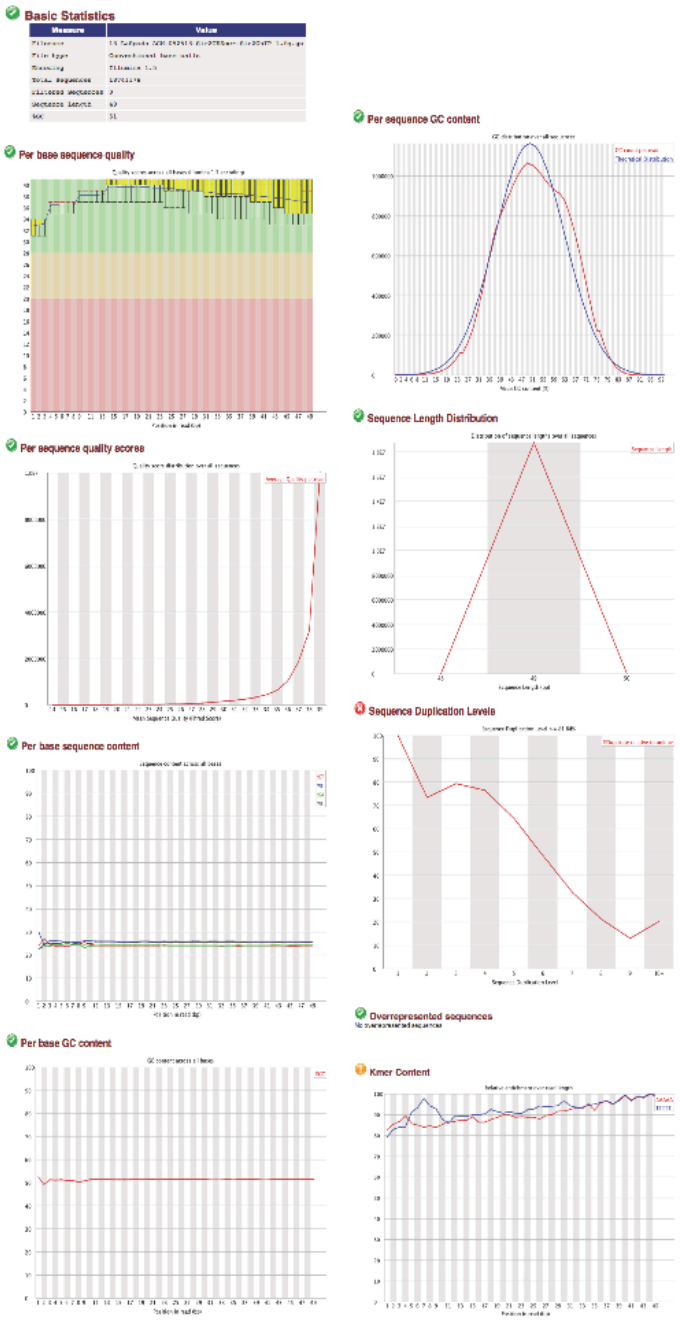
Supplementary figure 3-10: FastQC report for library 10, *sir2Δ fob1Δ* Sir2 ChIP#2

FastQC gives a quality control analysis for all raw reads sequenced from the library. It gives: the average read length and the quality of each base for each sequencing position; quality scores for each sequence; the percentage of each base at each sequence position; the percentage of G or C's at each base position; the G/C percentage in each sequence; the distribution of sequence length; the sequence duplication level in the library, how many times a sequence was sequenced; if there is an overrepresented sequence in the library; and 5-mer sequence enrichment in the library.

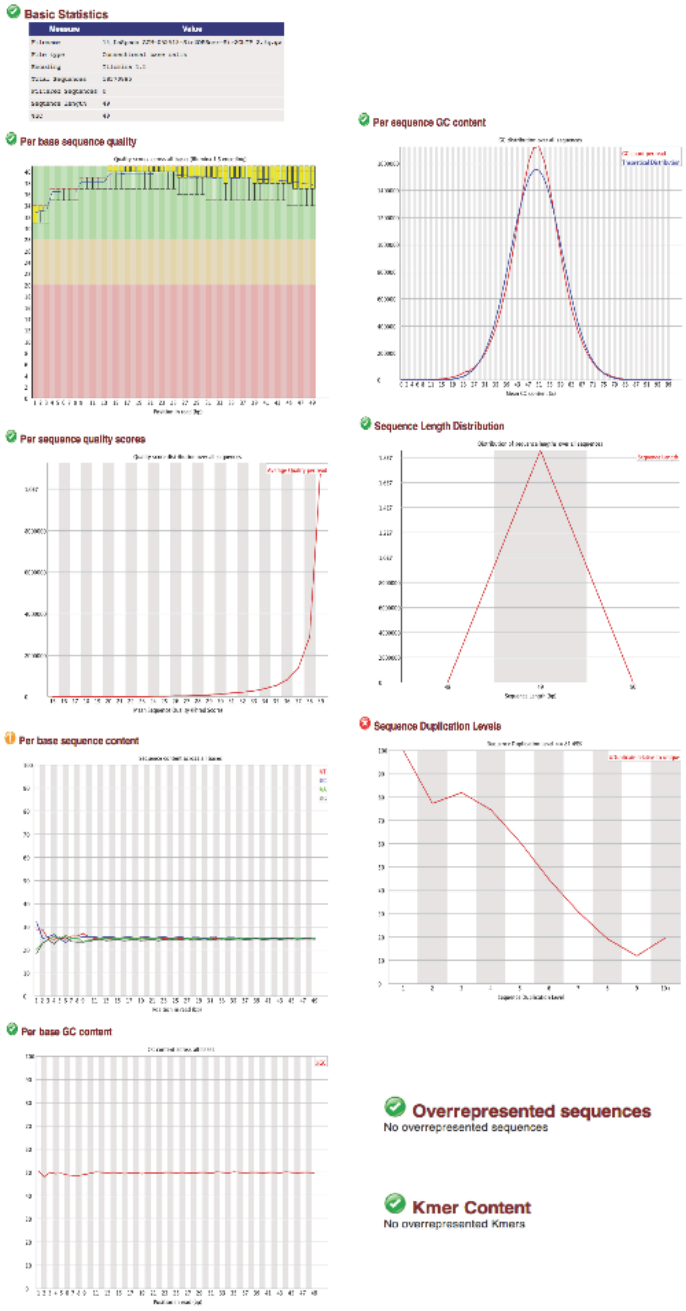


Supplementary figure 3-11: FastQC report for library 11, *sgf73Δ* Sir2 ChIP #1
FastQC gives a quality control analysis for all raw reads sequenced from the library. It gives: the average read length and the quality of each base for each sequencing position; quality scores for each sequence; the percentage of each base at each sequence position; the percentage of G or C's at each base position; the G/C percentage in each sequence; the distribution of sequence length; the sequence duplication level in the library, how many times a sequence was sequenced; if there is an overrepresented sequence in the library; and 5-mer sequence enrichment in the library.

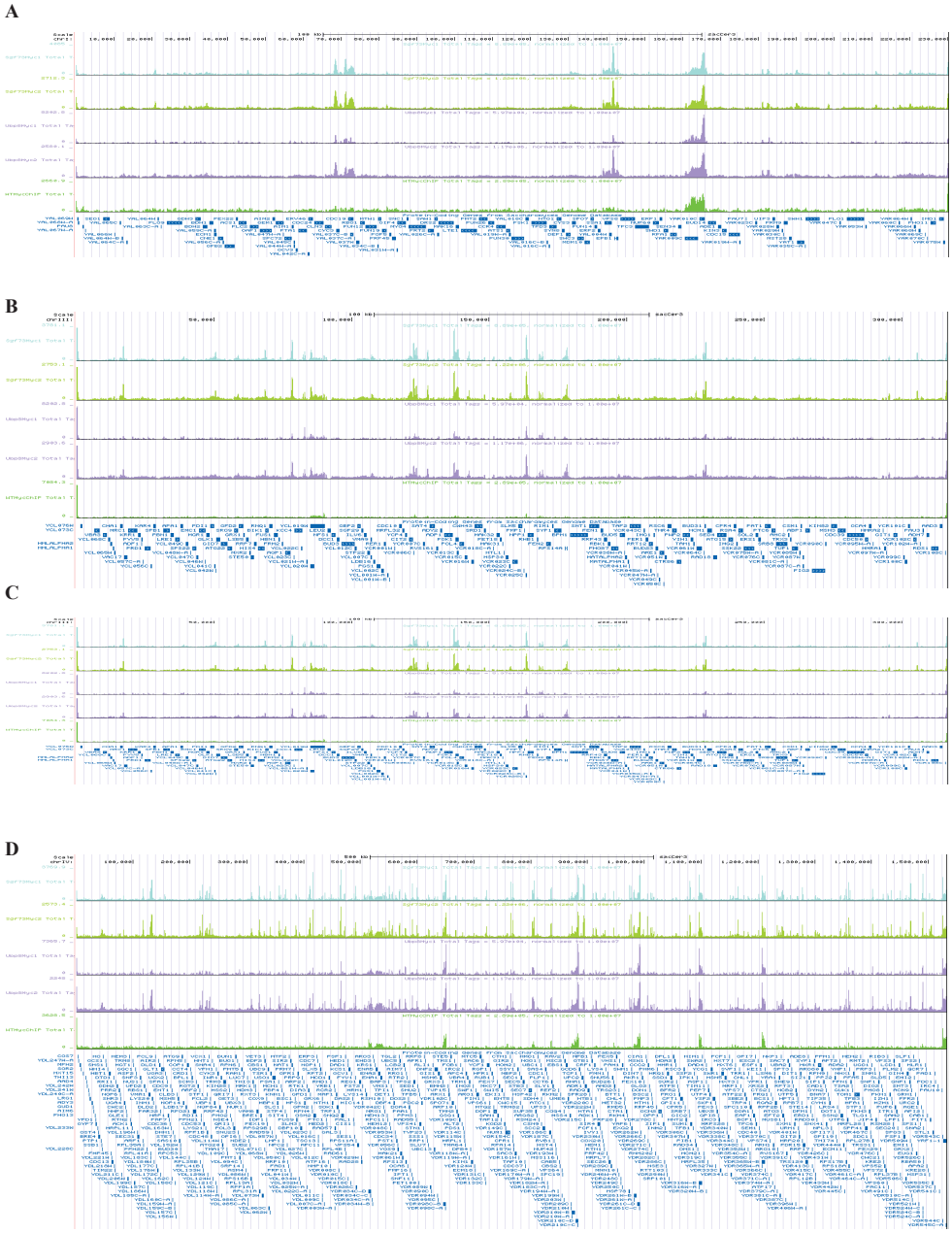
Supplementary figure 3-12: FastQC report for library 12, *sgf73Δ* Sir2 ChIP #2
FastQC gives a quality control analysis for all raw reads sequenced from the library. It gives: the average read length and the quality of each base for each sequencing position; quality scores for each sequence; the percentage of each base at each sequence position; the percentage of G or C's at each base position; the G/C percentage in each sequence; the distribution of sequence length; the sequence duplication level in the library, how many times a sequence was sequenced; if there is an overrepresented sequence in the library; and 5-mer sequence enrichment in the library.



Supplementary figure 3-13: FastQC report for library 13, *SIR2*OE Sir2 ChIP #1
FastQC gives a quality control analysis for all raw reads sequenced from the library. It gives: the average read length and the quality of each base for each sequencing position; quality scores for each sequence; the percentage of each base at each sequence position; the percentage of G or C's at each base position; the G/C percentage in each sequence; the distribution of sequence length; the sequence duplication level in the library, how many times a sequence was sequenced; if there is an overrepresented sequence in the library; and 5-mer sequence enrichment in the library.

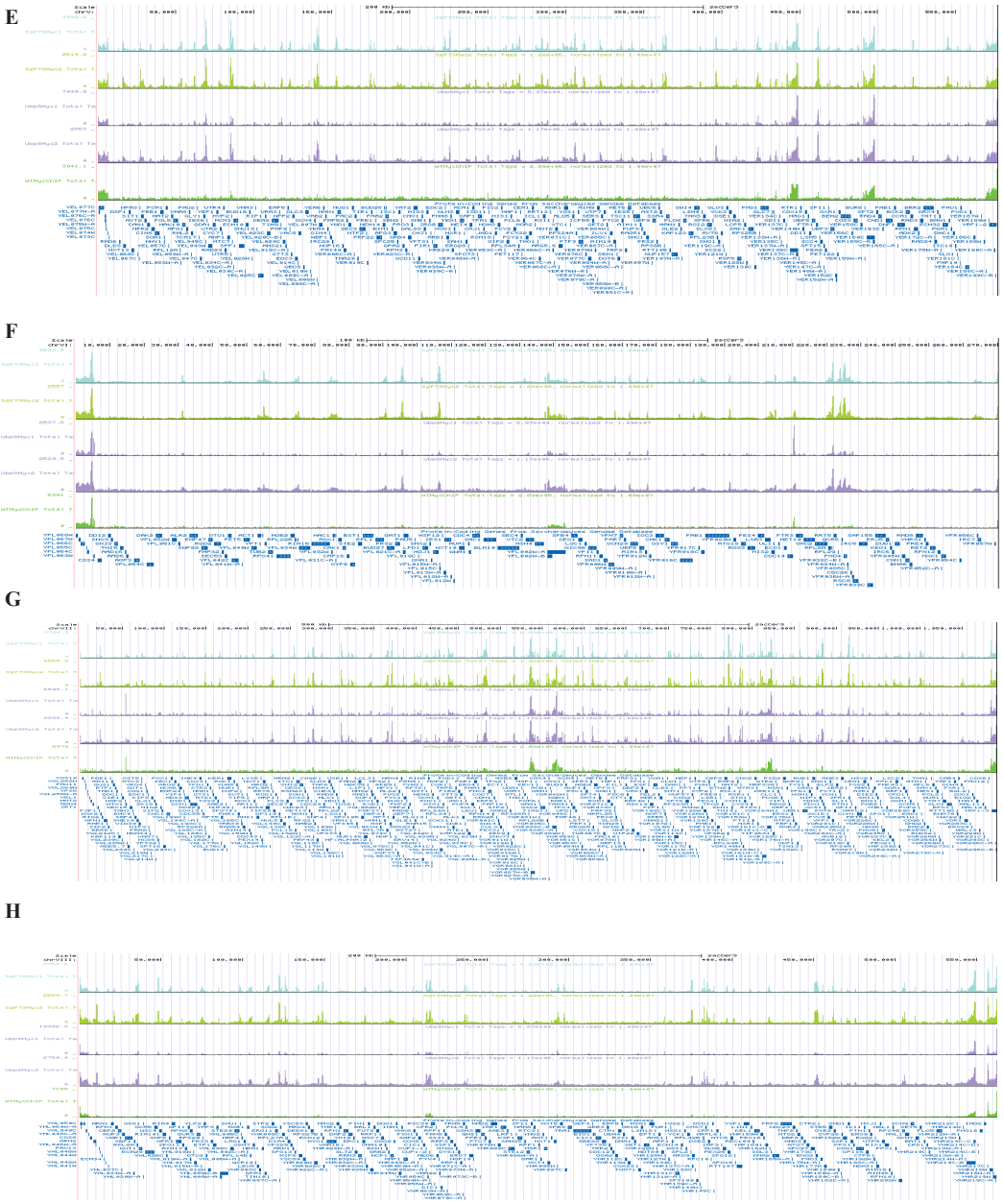


Supplementary figure 3-14: FastQC report for library 14, *SIR2*OE Sir2 ChIP #2
FastQC gives a quality control analysis for all raw reads sequenced from the library. It gives: the average read length and the quality of each base for each sequencing position; quality scores for each sequence; the percentage of each base at each sequence position; the percentage of G or C's at each base position; the G/C percentage in each sequence; the distribution of sequence length; the sequence duplication level in the library, how many times a sequence was sequenced; if there is an overrepresented sequence in the library; and 5-mer sequence enrichment in the library.

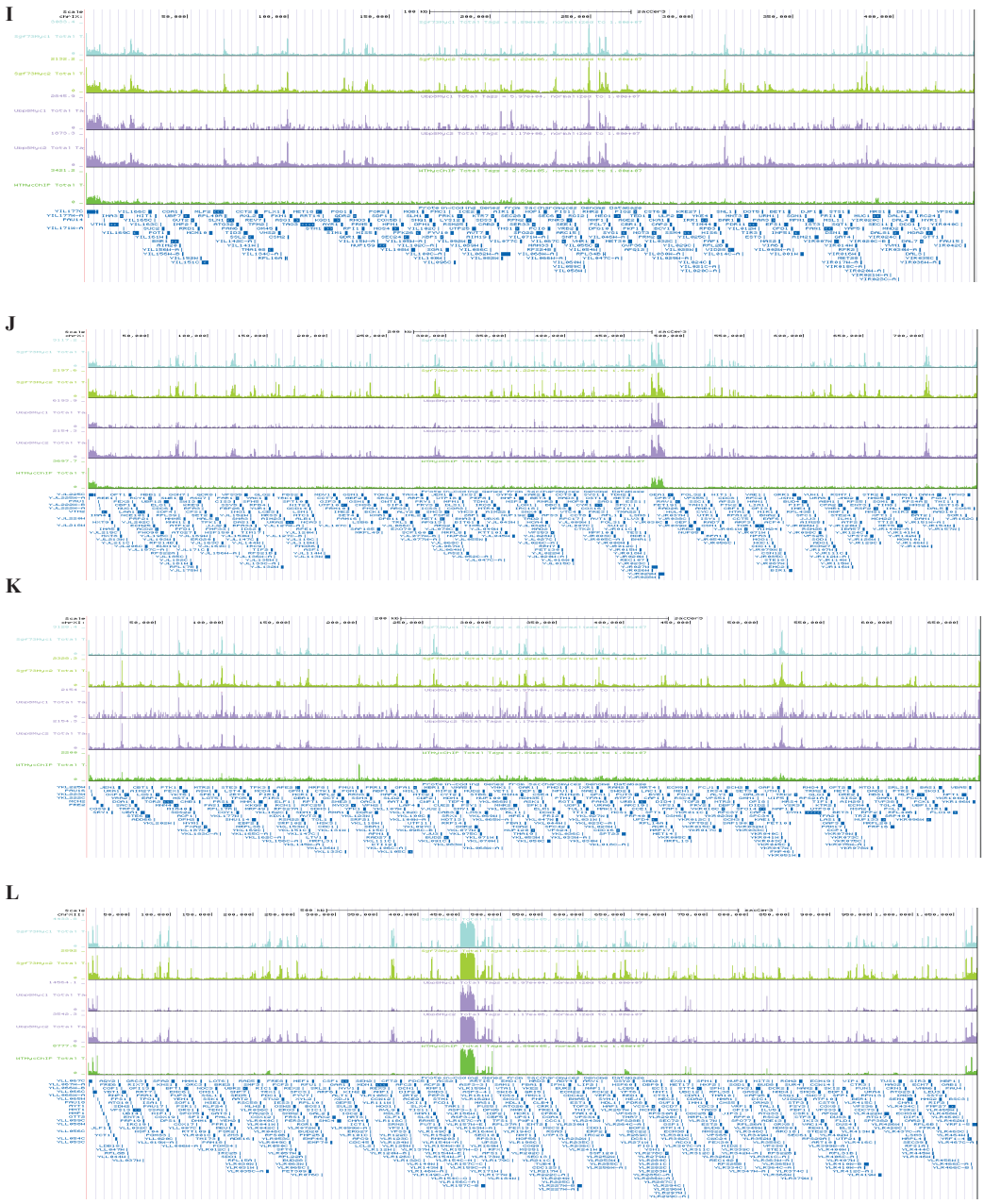


Supplementary figure 3-15: UCSC genome browser sequence traces for myc ChIP

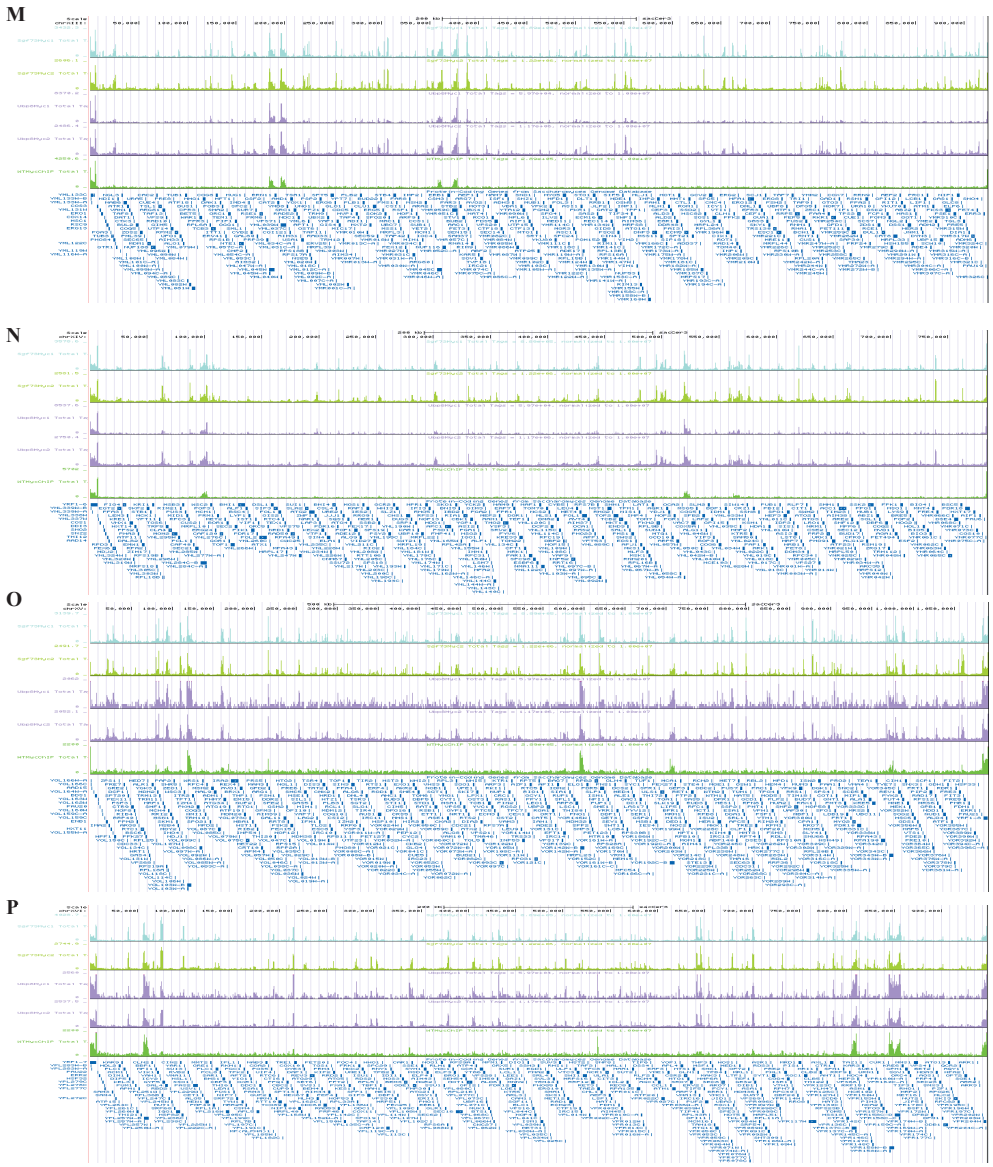
5 traces are listed in each subfigure, which represent the sequences mapped to each chromosome. The height of the peak (cluster of reads) represents the number of reads mapped to the region. Note that the scale for each track is different as it is based on the highest peak in the run. The top light blue track is Sgf73-myc myc ChIP replicate #1, the second light green track is Sgf73-myc myc ChIP replicate #2, the third purple track is Ubp8-myc myc ChIP replicate #1, the fourth purple track is Ubp8-myc myc ChIP replicate #2, and the bottom dark green track is the merged WT no tag myc ChIP replicate #1 and 2. A) Chromosome I B) Chromosome II C) Chromosome III D) Chromosome IV



Supplementary figure 3-15: UCSC genome browser sequence traces for myc ChIP, continued E) Chromosome V F) Chromosome VI G) Chromosome VII H) Chromosome VIII



Supplementary figure 3-15: UCSC genome browser sequence traces for myc ChIP, continued I) Chromosome IX J) Chromosome X K) Chromosome XI L) Chromosome XII



Supplementary figure 3-15: UCSC genome browser sequence traces for myc ChIP, continued M) Chromosome XIII N) Chromosome XIV O) Chromosome XV P) Chromosome XVI

Supplementary Table 3-1: Significant Sgf73 occupied peaks

Table lists the chromosome number and the start and stop point of each significant peak. The peak score refers to the number of uniquely mapped reads to that peak location. The table also list the distance to the nearest transcriptional start site (TSS) what promoter ID that TSS is associated with. The table also includes the gene name and description the TSS is associated with if available.

Chr #	Start	End	Peak Score	Distance to TSS	Nearest PromoterID	Gene Name	Gene Description
chrI	141707	141838	3693	-341	YAL005C	SSA1	Hsp70 family ATPase SSA1
chrXVI	645455	645577	3649.449951	-434	YPR036W-A	YPR036W-A	hypothetical protein
chrXVI	75080	75186	3532.5	-414	YPL250C	ICY2	Icy2p
chrIV	465299	465407	3402.100098	30	YDR010C		
chrVII	772109	772212	3341.25	-294	YGR142W	BTN2	Btn2p
chrIII	137335	137440	3245.25	-359	YCR012W	PGK1	phosphoglycerate kinase
chrV	69152	69273	3229.25	53	YEL045C		
chrXII	97654	97755	3146.049805	-219	YLL024C	SSA2	Hsp70 family chaperone SSA2
chrVII	914977	915084	3095.399902	-211	YGR211W	ZPR1	Zpr1p
chrIV	416997	417100	3093.399902	-340	YDL020C	RPN4	Rpn4p
chrXII	370597	370697	2938.949951	-144	YLR112W		
chrIV	1490147	1490262	2916.050049	-214	YDR524C-A		
chrXIII	388438	388553	2899.699951	235	YMR058W	FET3	Fet3 Ferro-O2-oxidoreductase
chrXIV	619847	619948	2890.5	-170	YNL006W	LST8	Lst8p
chrII	477329	477444	2832.100098	-285	YBR118W	TEF2	Tef2p
chrVII	371312	371427	2816.899902	-7	YGL072C		
chrVII	483464	483589	2808.100098	266	YGL007W		
chrIX	249160	249275	2802.5	-774	YIL056W	VHR1	Vhr1p
chrXII	368368	368489	2778.350098	-304	YLR108C	YLR108C	hypothetical protein
chrXV	619323	619443	2765.75	-457	YOR153W	PDR5	ATP-binding cassette multidrug transporter PDR5
chrXI	518624	518748	2752.850098	-120	YKR040C		
chrXV	83246	83361	2713.899902	-383	YOL126C	MDH2	malate dehydrogenase MDH2
chrXVI	76423	76536	2697.600098	-240	YPL249C-A	RPL36B	ribosomal 60S subunit protein L36B
chrIX	386713	386823	2683.350098	-1067	YIR018C-A	YIR018C-A	hypothetical protein
chrVII	884045	884150	2639.049805	-287	YGR192C	TDH3	glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) TDH3
chrXV	797080	797207	2588.899902	-350	YOR246C	ENV9	Env9p
chrVII	609975	610076	2582.850098	-539	YGR060W	ERG25	methylsterol monooxygenase
chrVIII	122344	122456	2534.450195	365	YHR007C-A	YHR007C-A	hypothetical protein
chrII	444793	444899	2486.050049	-153	YBR101C	FES1	Fes1p

Supplementary Table 3-1: Significant Sgf73 occupied peaks continued

Chr #	Start	End	Peak Score	Distance to TSS	Nearest PromoterID	Gene Name	Gene Description
chrXV	877838	877959	2474.350098	-213	YOR298C-A	MBF1	Mbf1p
chrXII	86985	87090	2469.25	-366	YLL027W	ISA1	Isa1p
chrXIII	306644	306749	2452.200195	-793	YMR017W	SPO20	Spo20p
chrXIV	87501	87612	2443.75	-341	YNL289W	PCL1	Pcl1p
chrVII	785546	785668	2433.800049	-170	YGR146C-A	YGR146C-A	hypothetical protein
chrXIII	632111	632218	2425.75	-191	YMR186W	HSC82	Hsp90 family chaperone HSC82
chrXVI	641341	641530	2407.850098	-773	YPR035W	GLN1	glutamate--ammonia ligase
chrIV	132841	132988	2401.200195	-523	YDL182W	LYS20	homocitrate synthase LYS20
chrVI	221769	221872	2330.100098	-402	YFR031C-A	RPL2A	ribosomal 60S subunit protein L2A
chrVII	809750	809868	2316.850098	-389	YGR161C	RTS3	Rts3p
chrIV	1080572	1080682	2313.25	-428	YDR309C	GIC2	Gic2p
chrIII	78427	78535	2311.699951	-562	YCL025C	AGP1	Agp1p
chrXVI	679114	679269	2307.100098	240	YPR064W		
chrXIV	739430	739557	2297.049805	-458	YNR060W	FRE4	Fre4p
chrXVI	40478	40599	2293.699951	-505	YPL265W	DIP5	Dip5p
chrIV	765359	765481	2291.649902	-265	YDR151C	CTH1	Cth1p
chrVI	225173	225280	2282.100098	-457	YFR033C	QCR6	ubiquinol--cytochrome-c reductase subunit 6
chrXV	109715	109850	2261.600098	-515	YOL109W	ZEO1	Zeo1p
chrXI	99841	99943	2238.050049	-779	YKL182W	FAS1	tetrafunctional fatty acid synthase subunit FAS1
chrXVI	785815	785935	2232	-333	YPR124W	CTR1	Ctr1p
chrXI	381883	381993	2218.75	-80	YKL032C	IXR1	Ixr1p
chrVII	310724	310830	2204.850098	-190	YGL103W	RPL28	ribosomal 60S subunit protein L28
chrXII	253425	253546	2199.549805	-376	YLR056W	ERG3	C-5 sterol desaturase
chrIV	806329	806433	2185.649902	-240	YDR171W	HSP42	Hsp42p
chrX	521723	521839	2181.350098	-179	YJR045C	SSC1	Hsp70 family ATPase SSC1
chrXI	258357	258458	2181.050049	506	YKL097C		
chrXVI	98796	98904	2175.449951	-225	YPL240C	HSP82	Hsp90 family chaperone HSP82
chrXIII	551398	551502	2158.399902	-243	YMR142C	RPL13B	ribosomal 60S subunit protein L13B
chrXIII	362278	362398	2112.649902	289	YMR045C	YMR045C	gag-pol fusion protein
chrXIII	302883	303024	2109.600098	-468	YMR015C	ERG5	C-22 sterol desaturase
chrV	306028	306150	2074.899902	-234	YER074W	RPS24A	ribosomal 40S subunit protein S24A
chrIII	122610	122748	2055.049805	324	YCR006C		
chrI	71087	71225	2043.899902	-630	YAL038W	CDC19	pyruvate kinase CDC19

Supplementary Table 3-1: Significant Sgf73 occupied peaks continued

Chr #	Start	End	Peak Score	Distance to TSS	Nearest PromoterID	Gene Name	Gene Description
chrIV	221994	222116	2043.699951	-254	YDL133C-A	RPL41B	ribosomal 60S subunit protein L41B
chrXIII	915350	915470	2029.850098	-872	YMR319C	FET4	Fet4p
chrII	140573	140678	2027.449951	-365	YBL042C	FUI1	Fui1p
chrII	415546	415648	2027.050049	-336	YBR084C-A	RPL19A	ribosomal 60S subunit protein L19A
chrV	225532	225632	2020	-307	YER037W	PHM8	Phm8p
chrVII	598973	599078	2016.849976	-392	YGR055W	MUP1	Mup1p
chrIV	976598	976701	1996.599976	580	YDR260C	SWM1	Swm1p
chrII	221556	221691	1995.699951	293	YBL005W-B	YBL005W-B	gag-pol fusion protein
chrVII	726667	726769	1994.349976	-239	YGR117C	YGR117C	hypothetical protein
chrIV	894079	894197	1987.550049	364	YDR215C		
chrII	332549	332653	1984.5	-230	YBR048W	RPS11B	ribosomal 40S subunit protein S11B
chrXIV	495272	495401	1984.149902	-335	YNL069C	RPL16B	ribosomal 60S subunit protein L16B
chrXI	164629	164731	1972.899902	-295	YKL152C	GPM1	phosphoglycerate mutase GPM1
chrIII	228865	228974	1965.25	-391	YCR065W	HCM1	Hcm1p
chrII	168066	168211	1955	-285	YBL027W	RPL19B	ribosomal 60S subunit protein L19B
chrXIII	253438	253555	1952.350098	-224	YML007C-A	YML007C-A	hypothetical protein
chrIV	591702	591833	1930	-423	YDR072C	IPT1	inositolphosphotransferase
chrXIII	408541	408658	1925.349976	-555	YMR070W	MOT3	Mot3p
chrXIV	576314	576456	1920.25	-335	YNL031C	HHT2	Hht2p
chrXII	645883	646012	1898.299927	6	YLR255C		
chrVII	149217	149324	1897.75	246	YGL188C		
chrXIV	63173	63298	1897.200073	-292	YNL302C	RPS19B	ribosomal 40S subunit protein S19B
chrXIII	225563	225673	1892.600098	-253	YML025C	YML6	mitochondrial 54S ribosomal protein YmL6
chrXII	809351	809459	1890.649902	-592	YLR342W	FKS1	Fks1p
chrXVI	453531	453639	1888.549927	150	YPL056C	LCL1	Lcl1p
chrVII	23427	23542	1886.400024	-451	YGL253W	HXK2	hexokinase 2
chrXII	1019008	1019117	1878.75	-154	YLR441C	RPS1A	ribosomal 40S subunit protein S1A
chrVI	106456	106562	1876.299927	94	YFL015W-A		
chrIV	769194	769295	1871.199951	-244	YDR155C	CPR1	peptidylprolyl isomerase CPR1
chrIV	1468880	1469013	1857.5	-454	YDR510W	SMT3	SUMO family protein SMT3
chrXII	88325	88433	1855.5	-244	YLL026W	HSP104	chaperone ATPase HSP104
chrXV	987633	987812	1826.649902	1060	YOR348C	PUT4	Put4p
chrXII	568200	568307	1820.25	-314	YLR214W	FRE1	Fre1p

Supplementary Table 3-1: Significant Sgf73 occupied peaks continued

Chr #	Start	End	Peak Score	Distance to TSS	Nearest PromoterID	Gene Name	Gene Description
chrVII	254368	254492	1819.050049	-211	YGL135W	RPL1B	ribosomal 60S subunit protein L1B
chrXV	968087	968210	1809.400024	326	YOR343C		
chrXI	327452	327569	1799.650024	-23	YKL060C	FBA1	fructose-bisphosphate aldolase FBA1
chrXVI	432008	432169	1799.650024	193	YPL062W		
chrXIII	540470	540591	1797.449951	-474	YMR135C	GID8	glucose-induced degradation complex subunit GID8
chrIX	254323	254427	1793.899902	-168	YIL054W	YIL054W	hypothetical protein
chrVIII	48291	48430	1792.599976	-392	YHL029C	OCA5	Oca5p
chrXV	27930	28050	1785.900024	906	YOL155W-A	YOL155W-A	hypothetical protein
chrXVI	794676	794786	1764.150024	-233	YPR131C	NAT3	Nat3p
chrIV	892501	892609	1760.400024	-320	YDR214W	AHA1	Aha1p
chrIV	1165017	1165208	1759.149902	-452	YDR345C	HXT3	Hxt3p
chrX	291143	291259	1751.350098	-427	YJL079C	PRY1	Pry1p
chrII	408680	408781	1741.449951	-439	YBR083W	TEC1	Tec1p
chrIV	974377	974480	1740.800049	-185	YDR258C	HSP78	chaperone ATPase HSP78
chrIV	321879	321984	1733.550049	-295	YDL075W	RPL31A	ribosomal 60S subunit protein L31A
chrIV	550979	551096	1732.649902	-461	YDR046C	BAP3	Bap3p
chrXIV	559344	559470	1728.550049	-405	YNL037C	IDH1	isocitrate dehydrogenase (NAD(+)) IDH1
chrXI	67856	67969	1724.5	-446	YKL201C	MNN4	Mnn4p
chrIV	357221	357333	1721.399902	-518	YDL055C	PSA1	mannose-1-phosphate guanylyltransferase
chrXII	737858	737998	1715.099976	-380	YLR304C	ACO1	aconitate hydratase ACO1
chrXVI	653977	654090	1714.699951	-133	YPR043W	RPL43A	ribosomal 60S subunit protein L43A
chrXV	253915	254041	1710.600098	-319	YOL039W	RPP2A	ribosomal protein P2A
chrVII	516402	516507	1696.400024	-489	YGR014W	MSB2	Msb2p
chrV	396441	396542	1692.199951	-278	YER117W	RPL23B	ribosomal 60S subunit protein L23B
chrIX	257361	257495	1688.699951	-365	YIL052C	RPL34B	ribosomal 60S subunit protein L34B
chrV	364279	364401	1678	-249	YER103W	SSA4	Hsp70 family chaperone SSA4
chrXII	241894	241996	1676	-287	YLR048W	RPS0B	ribosomal 40S subunit protein S0B
chrXIII	880621	880722	1673.449951	-439	YMR305C	SCW10	Scw10p
chrXIV	444493	444598	1658.099976	-230	YNL096C	RPS7B	ribosomal 40S subunit protein S7B
chrXIV	331036	331137	1652.699951	-236	YNL162W	RPL42A	ribosomal 60S subunit protein L42A
chrXIII	661832	661940	1646.300049	-758	YMR199W	CLN1	Cln1p
chrXV	232167	232269	1644.25	-463	YOL052C-A	DDR2	Ddr2p

Supplementary Table 3-1: Significant Sgf73 occupied peaks continued

Chr #	Start	End	Peak Score	Distance to TSS	Nearest PromoterID	Gene Name	Gene Description
chrVII	920127	920250	1631.900024	-387	YGR214W	RPS0A	ribosomal 40S subunit protein S0A
chrX	651426	651538	1628.449951	-419	YJR123W	RPS5	Rps5p
chrXIII	124367	124483	1627.349976	-253	YML073C	RPL6A	ribosomal 60S subunit protein L6A
chrXV	779633	779774	1625.900024	-167	YOR235W		
chrX	157391	157495	1625.800049	-170	YJL136C	RPS21B	Rps21bp
chrXII	837671	837775	1623.75	-366	YLR354C	TAL1	sedoheptulose-7-phosphate:D-glyceraldehyde-3-phosphate transaldolase TAL1
chrVII	557749	557863	1622.199951	241	YGR035W-A	YGR035W-A	hypothetical protein
chrX	75554	75706	1620.550049	-303	YJL189W	RPL39	Rpl39p
chrIV	491198	491321	1614.149902	242	YDR024W		
chrXIII	26153	26277	1612.949951	203	YML122C		
chrV	141172	141289	1601.400024	-412	YEL008C-A		
chrXII	941022	941126	1596.400024	-409	YLR410W-A	YLR410W-A	gag protein
chrXII	1012184	1012302	1595.599976	-220	YLR437C	DIF1	Dif1p
chrVII	398060	398174	1594.449951	-499	YGL056C	SDS23	Sds23p
chrXVI	297165	297293	1591.349976	-324	YPL135W	ISU1	Isu1p
chrXIV	481044	481146	1584.050049	-296	YNL077W	APJ1	Apj1p
chrIII	17042	17170	1578.25	-184	YCL063W	VAC17	Vac17p
chrVII	438598	438715	1571.650024	-435	YGL030W	RPL30	ribosomal 60S subunit protein L30
chrIX	68403	68520	1566	-247	YIL148W	RPL40A	ubiquitin-ribosomal 60S subunit protein L40A fusion protein
chrIV	234460	234589	1565.099976	-403	YDL127W	PCL2	Pcl2p
chrXII	704439	704551	1564.5	186	YLR279W		
chrX	236586	236708	1561.900024	-291	YJL101C	GSH1	Gsh1p
chrV	362398	362512	1560.900024	-645	YER102W	RPS8B	ribosomal 40S subunit protein S8B
chrVIII	450904	451015	1560.849976	194	YHR173C	YHR173C	hypothetical protein
chrXV	424438	424542	1557.900024	127	YOR050C		
chrXIII	224035	224154	1554.300049	-266	YML026C	RPS18B	ribosomal 40S subunit protein S18B
chrII	89622	89733	1553.099976	-121	YBL071C-B	YBL071C-B	hypothetical protein
chrVIII	462119	462233	1547.650024	-326	YHR179W	OYE2	Oye2p
chrII	392568	392681	1542.949951	-331	YBR077C	SLM4	Slm4p
chrV	222426	222534	1542.5	-159	YER035W	EDC2	Edc2p
chrXII	84210	84327	1535.75	-536	YLL028W	TPO1	Tpo1p
chrIV	117431	117534	1530.599976	-182	YDL191W	RPL35A	ribosomal 60S subunit protein L35A

Supplementary Table 3-1: Significant Sgf73 occupied peaks continued

Chr #	Start	End	Peak Score	Distance to TSS	Nearest PromoterID	Gene Name	Gene Description
chrIX	22172	22282	1528.800049	-3674	YIL172C	IMA3	Ima3p
chrXIV	763158	763262	1523	-618	YNR069C	BSC5	Bsc5p
chrIII	178331	178439	1519.449951	-165	YCR031C	RPS14A	ribosomal 40S subunit protein S14A
chrX	123110	123216	1518.600098	-215	YJL158C	CIS3	Cis3p
chrII	236086	236201	1518.600098	-349	YBL002W	HTB2	Htb2p
chrIV	130611	130753	1513.25	-198	YDL184C	RPL41A	ribosomal 60S subunit protein L41A
chrXV	901425	901527	1482.649902	-282	YOR312C	RPL20B	ribosomal 60S subunit protein L20B
chrII	407335	407438	1473.5	-217	YBR082C	UBC4	E2 ubiquitin-conjugating protein UBC4
chrXII	1028449	1028574	1472.050049	-343	YLR448W	RPL6B	ribosomal 60S subunit protein L6B
chrII	217192	217304	1467.349976	-119	YBL006C	LDB7	Ldb7p
chrXIV	503192	503298	1465.800049	-479	YNL065W	AQR1	Aqr1p
chrII	555857	555975	1464.099976	-633	YBR158W	AMN1	Amn1p
chrXVI	281806	281916	1462.199951	-261	YPL143W	RPL33A	ribosomal 60S subunit protein L33A
chrXII	282749	282850	1457.149902	-128	YLR075W	RPL10	ribosomal 60S subunit protein L10
chrV	284803	284951	1445.849976	-655	YER064C	VHR2	Vhr2p
chrIII	123589	123699	1445.400024	-641	YCR006C		
chrVIII	505877	505989	1438.25	-386	YHR204W	MNL1	Mnl1p
chrVIII	384814	384915	1433.649902	-646	YHR143W	DSE2	Dse2p
chrXIII	754426	754526	1431.550049	179	YMR242C	RPL20A	ribosomal 60S subunit protein L20A
chrIV	1278743	1278855	1429.699951	-411	YDR406W	PDR15	ATP-binding cassette multidrug transporter PDR15
chrXVI	425314	425465	1427.5	-293	YPL068C	YPL068C	hypothetical protein
chrIV	600048	600162	1425.600098	-688	YDR077W	SED1	Sed1p
chrXVI	502824	503000	1421.25	118	YPL025C		
chrXV	312689	312792	1416.849976	-373	YOL007C	CSI2	Csi2p
chrXI	231062	231179	1408.650024	-1107	YKL109W	HAP4	Hap4p
chrIX	127827	127946	1406.699951	-265	YIL123W	SIM1	Sim1p
chrXII	897920	898023	1393.75	-299	YLR387C	REH1	Reh1p
chrVII	384781	384917	1392.150024	-198	YGL063C-A		
chrXIII	751433	751534	1386.550049	-243	YMR240C	CUS1	Cus1p
chrXIII	675196	675318	1386.199951	-491	YMR205C	PFK2	6-phosphofructokinase subunit beta
chrXV	904003	904112	1374.300049	-401	YOR314W-A		
chrIV	1355690	1355800	1370.25	-192	YDR447C	RPS17B	ribosomal 40S subunit protein S17B
chrXI	326149	326290	1369.650024	448	YKL061W	BLI1	Bli1p

Supplementary Table 3-1: Significant Sgf73 occupied peaks continued

Chr #	Start	End	Peak Score	Distance to TSS	Nearest PromoterID	Gene Name	Gene Description
chrXVI	302905	303019	1368.25	-159	YPL131W	RPL5	ribosomal 60S subunit protein L5
chrXI	108988	109092	1362.650024	-229	YKL180W	RPL17A	ribosomal 60S subunit protein L17A
chrXVI	188724	188846	1362.5	-272	YPL189C-A	COA2	Coa2p
chrV	174817	174926	1358.800049	-377	YER011W	TIR1	Tir1p
chrIV	1401480	1401586	1358	-237	YDR471W	RPL27B	ribosomal 60S subunit protein L27B
chrXIII	653544	653670	1355.350098	-427	YMR195W	ICY1	Icy1p
chrV	153026	153167	1352.5	-424	YER001W	MNN1	Mnn1p
chrV	545185	545334	1348.849976	-352	YER177W	BMH1	Bmh1p
chrXV	1004227	1004345	1347.349976	-851	YOR355W	GDS1	Gds1p
chrXII	125194	125303	1343.649902	-286	YLL012W	YEH1	Yeh1p
chrXVI	67103	67213	1342.699951	-544	YPL256C	CLN2	Cln2p
chrXIII	510965	511080	1341.699951	53	YMR122C		
chrVII	856795	856928	1337.550049	-560	YGR180C	RNR4	ribonucleotide-diphosphate reductase subunit RNR4
chrIV	568935	569047	1333.399902	-288	YDR058C	TGL2	Tgl2p
chrXIV	662059	662194	1331.699951	-752	YNR016C	ACC1	acetyl-CoA carboxylase ACC1
chrII	604264	604410	1331.399902	-171	YBR189W	RPS9B	ribosomal 40S subunit protein S9B
chrII	60934	61043	1330.300049	-249	YBL087C	RPL23A	ribosomal 60S subunit protein L23A
chrII	466265	466410	1314.550049	-567	YBR112C	CYC8	Cyc8p
chrVIII	35959	36080	1295.150024	6	YHL033C	RPL8A	ribosomal 60S subunit protein L8A
chrVII	366078	366180	1292.449951	-133	YGL076C	RPL7A	ribosomal 60S subunit protein L7A
chrXII	65478	65595	1285.399902	-329	YLL039C	UBI4	ubiquitin
chrX	623064	623166	1282.199951	-101	YJR104C	SOD1	Sod1p
chrVII	650916	651020	1281.75	-351	YGR086C	PIL1	Pil1p
chrXI	533346	533477	1275.349976	53	YKR052C	MRS4	Mrs4p
chrX	159992	160104	1272.5	-201	YJL133C-A	YJL133C-A	hypothetical protein
chrVIII	293514	293634	1257.25	631	YHR095W		
chrV	322007	322114	1252.150024	-626	YER081W	SER3	phosphoglycerate dehydrogenase SER3
chrIV	1434497	1434597	1251.599976	-279	YDR490C	PKH1	Pkh1p
chrIV	369019	369137	1242.050049	-693	YDL047W	SIT4	Sit4p
chrXV	1028799	1028904	1232.300049	-226	YOR369C	RPS12	ribosomal 40S subunit protein S12
chrX	173127	173231	1230.699951	-812	YJL130C	URA2	Ura2p
chrXV	762189	762320	1224.400024	-170	YOR226C	ISU2	Isu2p
chrXII	665102	665237	1219.899902	-167	YLR259C	HSP60	chaperone ATPase HSP60

Supplementary Table 3-1: Significant Sgf73 occupied peaks continued

Chr #	Start	End	Peak Score	Distance to TSS	Nearest PromoterID	Gene Name	Gene Description
chrVII	766278	766386	1219.899902	606	YGR139W		
chrXIII	852188	852303	1216.949951	-385	YMR291W	TDA1	Tda1p
chrXV	384815	384916	1211.849976	265	YOR029W		
chrIV	721855	721970	1199.599976	-615	YDR133C		
chrIX	232664	232776	1198	-293	YIL068W-A		
chrXI	428546	428673	1192.449951	-58	YKL008C	LAC1	sphingosine N-acyltransferase LAC1
chrXIV	302459	302564	1192.449951	-169	YNL178W	RPS3	ribosomal 40S subunit protein S3
chrXI	463166	463280	1189.900024	-527	YKR011C	YKR011C	hypothetical protein
chrIV	1362422	1362534	1184.699951	-297	YDR451C	YHP1	Yhp1p
chrXII	39519	39635	1181.400024	-106	YLL051C	FRE6	Fre6p
chrXII	950693	950810	1160.050049	-278	YLR412C-A	YLR412C-A	hypothetical protein
chrX	69033	69168	1150.949951	-238	YJL194W	CDC6	AAA family ATPase CDC6
chrII	292037	292149	1150.050049	-228	YBR025C	OLA1	Ola1p
chrXI	431245	431385	1146.300049	-591	YKL006W	RPL14A	ribosomal 60S subunit protein L14A
chrVII	702960	703075	1142.449951	350	YGR107W		
chrV	335515	335681	1141.800049	-329	YER088W-B		
chrII	452252	452366	1134.650024	-340	YBR105C	VID24	glucose-induced degradation complex subunit VID24
chrX	525859	525976	1129.050049	-418	YJR048W	CYC1	Cyc1p
chrVII	987318	987437	1125.25	-672	YGR249W	MGA1	Mga1p
chrIV	556658	556799	1123.25	-256	YDR050C	TPI1	triose-phosphate isomerase TPI1
chrXVI	491689	491791	1122.900024	-376	YPL032C	SVL3	Svl3p
chrVII	271917	272031	1122.400024	-546	YGL125W	MET13	methylenetetrahydrofolate reductase (NAD(P)H) MET13
chrII	159876	159979	1120.800049	-234	YBL033C	RIB1	GTP cyclohydrolase II
chrVII	166154	166290	1109.099976	-1130	YGL178W	MPT5	Mpt5p
chrV	16082	16239	1109	-195	YEL071W	DLD3	Dld3p
chrXIII	347180	347300	1099.400024	-723	YMR037C	MSN2	Msn2p
chrII	614216	614316	1099.300049	-93	YBR196C-A	YBR196C-A	hypothetical protein
chrXIV	198220	198344	1097.25	-339	YNL241C	ZWF1	glucose-6-phosphate dehydrogenase
chrXIII	886612	886747	1096.300049	-107	YMR306C-A		
chrIV	1450918	1451035	1095.849976	-123	YDR500C	RPL37B	ribosomal 60S subunit protein L37B
chrII	681997	682097	1095.650024	1043	YBR231C	SWC5	Swc5p
chrXII	710591	710726	1087.599976	-522	YLR286C	CTS1	Cts1p
chrXI	614597	614727	1085.099976	-777	YKR092C	SRP40	Srp40p

Supplementary Table 3-1: Significant Sgf73 occupied peaks continued

Chr #	Start	End	Peak Score	Distance to TSS	Nearest PromoterID	Gene Name	Gene Description
chrIX	139279	139386	1082.550049	-420	YIL118W	RHO3	Rho3p
chrXVI	576184	576292	1078.900024	-314	YPR009W	SUT2	Sut2p
chrXVI	132373	132512	1075.449951	-327	YPL222C-A		
chrII	504543	504651	1073.800049	-257	YBR135W	CKS1	Cks1p
chrXIV	301795	301969	1073.75	-779	YNL179C		
chrXV	912014	912121	1072.099976	369	YOR318C		
chrIV	341325	341441	1063.099976	-236	YDL060W	TSR1	Tsr1p
chrXV	370984	371085	1054	-190	YOR020C	HSP10	Hsp10p
chrXIII	798076	798196	1048.75	-211	YMR265C	YMR265C	hypothetical protein
chrXV	520721	520840	1042.699951	-573	YOR107W	RGS2	Rgs2p
chrVII	1050308	1050489	1042.550049	-440	YGR279C	SCW4	Scw4p
chrV	123235	123349	1028.849976	-363	YEL017C-A	PMP2	Pmp2p
chrXIV	373780	373912	1024.349976	-265	YNL134C	YNL134C	hypothetical protein
chrVIII	21182	21339	1021.299988	-289	YHL040C	ARN1	Arn1p
chrXV	770214	770346	1020.050049	-520	YOR230W	WTM1	Wtm1p
chrIX	138333	138436	1018.699951	-507	YIL119C	RPI1	Rpi1p
chrX	195448	195572	1017.599976	-637	YJL116C	NCA3	Nca3p
chrXVI	404647	404764	1008.400024	-245	YPL081W	RPS9A	ribosomal 40S subunit protein S9A
chrXVI	770913	771021	1006.900024	-686	YPR119W	CLB2	Clb2p
chrXI	589676	589779	1005.299988	-422	YKR079C	TRZ1	Trz1p
chrXI	340303	340486	990.5	152	YKL052C	ASK1	Ask1p
chrIV	543739	543844	987.949951	-422	YDR043C	NRG1	Nrg1p
chrII	564951	565081	981.300049	-215	YBR162W-A	YSY6	Ysy6p
chrXIII	483588	483705	974.150024	-438	YMR108W	ILV2	acetolactate synthase catalytic subunit
chrXIII	353193	353302	972.599976	-624	YMR043W	MCM1	Mcm1p
chrXIII	119315	119431	972	-475	YML075C	HMG1	hydroxymethylglutaryl-CoA reductase (NADPH) HMG1
chrIV	1011529	1011702	966.950012	345	YDR274C		
chrIV	386910	387016	962.75	-1376	YDL037C	BSC1	Bsc1p
chrXII	789896	790027	957.800049	-715	YLR332W	MID2	Mid2p
chrXVI	280095	280222	954.099976	-322	YPL144W	POC4	Poc4p
chrVII	894412	894522	952.199951	-226	YGR198W	YPP1	Ypp1p
chrIV	307945	308070	943	-218	YDL083C	RPS16B	ribosomal 40S subunit protein S16B
chrIV	1429139	1429239	941.849976	-209	YDR487C	RIB3	3,4-dihydroxy-2-butanone-4-phosphate synthase RIB3
chrXIII	861206	861315	937.949951	-369	YMR296C	LCB1	serine C-palmitoyltransferase LCB1

Supplementary Table 3-1: Significant Sgf73 occupied peaks continued

Chr #	Start	End	Peak Score	Distance to TSS	Nearest PromoterID	Gene Name	Gene Description
chrX	348871	348999	937.849976	243	YJL047C-A	YJL047C-A	hypothetical protein
chrXIII	334967	335077	937.25	-276	YMR032W	HOF1	Hof1p
chrXII	674057	674169	930.200012	-169	YLR264C-A	YLR264C-A	hypothetical protein
chrII	798953	799073	930.150024	-491	YBR296C	PHO89	Pho89p
chrXV	40431	40532	929.699951	-267	YOL152W	FRE7	Fre7p
chrXIII	472975	473081	929.150024	236	YMR103C		
chrXVI	257130	257284	925.349976	-440	YPL156C	PRM4	Prm4p
chrV	342558	342662	921.050049	1	YER091C-A		
chrIV	909559	909677	915.299988	-436	YDR222W	YDR222W	hypothetical protein
chrXII	677793	677938	907.099976	-141	YLR266C	PDR8	Pdr8p
chrXII	962972	963077	903.049988	-761	YLR420W	URA4	dihydroorotase
chrXV	925249	925358	898.549988	-263	YOR324C	FRT1	Frt1p
chrVII	798239	798396	898.400024	-226	YGR155W	CYS4	cystathionine beta-synthase CYS4
chrXV	505408	505544	896.449951	-318	YOR096W	RPS7A	ribosomal 40S subunit protein S7A
chrIII	305739	305839	887.75	-322	YCR102C	YCR102C	hypothetical protein
chrXV	55235	55352	882.650024	-189	YOL143C	RIB4	lumazine synthase RIB4
chrVIII	422640	422780	880.699951	-362	YHR162W	MPC2	Mpc2p
chrXIV	427113	427255	880.549988	-430	YNL104C	LEU4	2-isopropylmalate synthase LEU4
chrV	431236	431336	878.5	-837	YER132C	PMD1	Pmd1p
chrXIII	872758	872867	868.849976	-187	YMR302C	YME2	Yme2p
chrVI	210605	210712	867.25	-279	YFR029W	PTR3	Ptr3p
chrXII	498646	498781	862.150024	-234	YLR167W	RPS31	ubiquitin-ribosomal 40S subunit protein S31 fusion protein
chrXIII	70213	70317	859.699951	127	YML100W-A	YML100W-A	hypothetical protein
chrXV	304909	305060	857.599976	-365	YOL011W	PLB3	Plb3p
chrXIII	500872	500979	857.049988	-237	YMR116C	ASC1	Asc1p
chrXVI	228035	228157	856	-218	YPL170W	DAP1	Dap1p
chrXV	825818	825919	853.049988	64	YOR268C	YOR268C	hypothetical protein
chrXIII	91151	91274	850.5	197	YML089C		
chrVIII	56055	56168	846.349976	-538	YHL024W	RIM4	Rim4p
chrXIV	652073	652181	841.799988	-338	YNR014W	YNR014W	hypothetical protein
chrXII	329437	329555	835.549988	-258	YLR094C	GIS3	Gis3p
chrVIII	133016	133134	833.549988	1028	YHR014W	SPO13	Spo13p
chrIX	420848	420950	832.549988	-163	YIR034C	LYS1	saccharopine dehydrogenase (NAD+, L-lysine-forming)
chrVIII	204331	204444	829.549988	-220	YHR048W	YHK8	Yhk8p

Supplementary Table 3-1: Significant Sgf73 occupied peaks continued

Chr #	Start	End	Peak Score	Distance to TSS	Nearest PromoterID	Gene Name	Gene Description
chrIV	930707	930822	826.450012	-365	YDR234W	LYS4	homoaconitate hydratase LYS4
chrXIII	608340	608452	825.849976	-185	YMR172C-A		
chrVIII	396939	397053	823.5	-255	YHR150W	PEX28	Pex28p
chrXII	818597	818710	822.75	-659	YLR344W	RPL26A	ribosomal 60S subunit protein L26A
chrVII	287365	287475	822.349976	1031	YGL118C		
chrX	36479	36595	818.799988	-288	YJL212C	OPT1	Opt1p
chrV	100118	100221	817.299988	-600	YEL027W	VMA3	Vma3p
chrV	242276	242386	808.549988	-830	YER045C	ACA1	Aca1p
chrVIII	512259	512369	801.300049	-418	YHR206W	SKN7	Skn7p
chrIX	84947	85104	800.950012	-28	YIL141W		
chrV	77702	77825	793.700012	-290	YEL040W	UTR2	Utr2p
chrIV	1345429	1345544	790.549988	-161	YDR442W		
chrXII	404252	404356	786.599976	-242	YLR130C	ZRT2	low-affinity Zn(2+) transporter ZRT2
chrXVI	689557	689659	786.549988	-499	YPR072W	NOT5	CCR4-NOT core subunit NOT5
chrXII	514958	515073	775.25	-247	YLR180W	SAM1	methionine adenosyltransferase SAM1
chrXI	490612	490756	770.650024	-680	YKR027W	BCH2	Bch2p
chrXIII	902554	902657	767.550049	-195	YMR315W	YMR315W	hypothetical protein
chrX	607973	608113	763.5	-262	YJR094W-A	RPL43B	Rpl43bp
chrXIV	350004	350108	757.950012	-300	YNL149C	PGA2	Pga2p
chrXV	891829	891941	756.799988	-453	YOR306C	MCH5	Mch5p
chrXII	263015	263122	756.400024	-126	YLR061W	RPL22A	ribosomal 60S subunit protein L22A
chrIX	177610	177748	746.25	2	YIL100C-A		
chrXVI	901680	901785	743.599976	-312	YPR184W	GDB1	bifunctional 4-alpha-glucanotransferase/amylo-alpha-1,6-glucosidase
chrIV	156087	156194	735	-178	YDL170W	UGA3	Uga3p
chrVII	677825	677971	734.450012	-797	YGR097W	ASK10	Ask10p
chrXIII	434405	434528	726.299988	-322	YMR083W	ADH3	alcohol dehydrogenase ADH3
chrXIII	170147	170278	721.550049	-190	YML052W	SUR7	Sur7p
chrXIII	37803	37919	721.550049	-335	YML116W	ATR1	Atr1p
chrII	326763	326869	717.049988	-757	YBR044C	TCM62	Tcm62p
chrXV	69173	69287	702.75	-476	YOL136C	PFK27	Pfk27p
chrIV	1256987	1257106	671.550049	-199	YDR390C	UBA2	E1 ubiquitin-activating protein UBA2
chrXV	832474	832645	668.450012	-254	YOR272W	YTM1	Ytm1p
chrXII	759161	759298	667.450012	-251	YLR312W-A	MRPL15	mitochondrial 54S ribosomal protein YmL15

Supplementary Table 3-1: Significant Sgf73 occupied peaks continued

Chr #	Start	End	Peak Score	Distance to TSS	Nearest PromoterID	Gene Name	Gene Description
chrXI	93849	93958	663.349976	-596	YKL185W	ASH1	Ash1p
chrVII	640560	640694	662.849976	-93	YGR079W	YGR079W	hypothetical protein
chrX	445081	445204	658.849976	-281	YJR004C	SAG1	Sag1p
chrXIII	747899	748001	657.200012	1597	YMR238W	DFG5	Dfg5p
chrXIII	667303	667407	640.450012	-182	YMR202W	ERG2	C-8 sterol isomerase ERG2
chrIX	370414	370527	631.150024	-234	YIR007W	YIR007W	hydrolase
chrXII	855615	855728	626.099976	-26	YLR364C-A		
chrIX	268002	268153	625	-232	YIL046W-A	YIL046W-A	hypothetical protein
chrII	101261	101378	616.349976	-163	YBL064C	PRX1	Prx1p
chrIV	562775	562956	614.349976	-538	YDR054C	CDC34	SCF E2 ubiquitin-protein ligase catalytic subunit CDC34
chrXVI	527104	527208	612.349976	-273	YPL015C	HST2	Hst2p
chrXVI	22600	22735	606.650024	-271	YPL274W	SAM3	Sam3p
chrII	697587	697702	604.550049	-342	YBR238C	YBR238C	hypothetical protein
chrVII	156307	156416	602.050049	-349	YGL184C	STR3	cystathionine beta-lyase STR3
chrVII	976969	977114	599.450012	-295	YGR243W	FMP43	Fmp43p
chrV	106956	107071	597.400024	-247	YEL024W	RIP1	ubiquinol--cytochrome-c reductase catalytic subunit RIP1
chrVIII	401139	401249	596.400024	-240	YHR152W	SPO12	Spo12p
chrVII	939166	939273	581.700012	-227	YGR221C	TOS2	Tos2p
chrI	198540	198708	579.550049	3163	YAR047C		
chrXIII	572879	572981	578.599976	401	YMR158C-A	YMR158C-A	hypothetical protein
chrII	662612	662714	566.300049	-414	YBR218C	PYC2	pyruvate carboxylase 2
chrII	628508	628660	559.599976	-584	YBR203W	COS111	Cos111p
chrXV	417976	418126	549.400024	-370	YOR047C	STD1	Std1p
chrX	268453	268577	518.800049	-284	YJL088W	ARG3	Arg3p
chrXIII	209141	209285	508	-312	YML034W	SRC1	Src1p
chrVII	206467	206572	472.200012	-514	YGL158W	RCK1	Rck1p
chrXVI	930516	930645	468.150024	-796	YPR196W	YPR196W	hypothetical protein
chrXVI	328957	329058	468.150024	-277	YPL117C	IDI1	isopentenyl-diphosphate delta-isomerase IDI1
chrIV	1056080	1056234	441.549988	-394	YDR297W	SUR2	sphingosine hydroxylase
chrIV	1087152	1087255	396.099976	-378	YDR312W	SSF2	Ssf2p

Supplementary Table 3-2: Significant Ubp8 occupied peaks

Table lists the chromosome number and the start and stop point of each significant peak. The peak score refers to the number of uniquely mapped reads to that peak location. The table also list the distance to the nearest transcriptional start site (TSS) what promoter ID that TSS is associated with. The table also includes the gene name and description the TSS is associated with if available.

Chr #	Start	End	Peak Score	Distance to TSS	Nearest PromoterID	Gene Name	Gene Description
chrXIII	388459	388581	3352.800049	210	YMR057C		
chrXVI	645462	645583	3236.450195	-428	YPR036W-A	YPR036W-A	hypothetical protein
chrV	461865	461992	2893.149902	-188	YER145C	FTR1	Ftr1p
chrVI	221775	221941	2402.199951	-440	YFR031C-A	RPL2A	ribosomal 60S subunit protein L2A
chrI	141702	141833	2399.399902	-336	YAL005C	SSA1	Hsp70 family ATPase SSA1
chrVII	371321	371451	2227.649902	-24	YGL072C		
chrVII	483514	483620	2220.600098	307	YGL007W		
chrXV	83225	83359	2114.149902	-372	YOL126C	MDH2	malate dehydrogenase MDH2
chrVII	310714	310834	2092.899902	-193	YGL103W	RPL28	ribosomal 60S subunit protein L28
chrXIV	619816	619945	2087.149902	-187	YNL006W	LST8	Lst8p
chrIV	1490153	1490268	2081.550049	-220	YDR524C-A		
chrXII	97654	97764	2034.700073	-224	YLL024C	SSA2	Hsp70 family chaperone SSA2
chrVII	772097	772234	2023.300049	-289	YGR142W	BTN2	Btn2p
chrXV	978840	978990	1990.699951	-846	YOR344C	TYE7	Tye7p
chrXII	282701	282834	1911.150024	-160	YLR075W	RPL10	ribosomal 60S subunit protein L10
chrXII	368432	368586	1885.699951	-272	YLR109W	AHP1	Ahp1p
chrVII	914963	915083	1867.25	-218	YGR211W	ZPR1	Zpr1p
chrVII	609915	610073	1864.400024	-570	YGR060W	ERG25	methylsterol monooxygenase
chrXV	109795	109963	1864.349976	-418	YOL109W	ZEO1	Zeo1p
chrXI	518633	518786	1824.75	-143	YKR040C		
chrXIII	124345	124457	1783.550049	-229	YML073C	RPL6A	ribosomal 60S subunit protein L6A
chrXVI	785825	785925	1743.849976	-333	YPR124W	CTR1	Ctr1p
chrXI	99878	100000	1739.5	-732	YKL182W	FAS1	tetrafunctional fatty acid synthase subunit FAS1
chrXV	619336	619437	1685.699951	-454	YOR153W	PDR5	ATP-binding cassette multidrug transporter PDR5
chrXIII	754429	754546	1677.100098	190	YMR242W-A	YMR242W-A	hypothetical protein
chrVII	726584	726764	1651.649902	-195	YGR117C	YGR117C	hypothetical protein
chrXIV	739397	739581	1604.800049	-462	YNR060W	FRE4	Fre4p
chrXII	241879	241998	1592.050049	-294	YLR048W	RPS0B	ribosomal 40S subunit protein S0B
chrVIII	35926	36081	1573.550049	22	YHL033C	RPL8A	ribosomal 60S subunit protein L8A
chrIV	1401447	1401578	1570.650024	-258	YDR471W	RPL27B	ribosomal 60S subunit protein L27B
chrXVI	281807	281920	1565.050049	-259	YPL143W	RPL33A	ribosomal 60S subunit protein L33A
chrXVI	794680	794788	1556.5	-231	YPR132W	RPS23B	ribosomal 40S subunit protein S23B
chrXII	941027	941134	1525.25	-403	YLR410W-A	YLR410W-A	gag protein

Supplementary Table 3-2: Significant Ubp8 occupied peaks continued

Chr #	Start	End	Peak Score	Distance to TSS	Nearest PromoterID	Gene Name	Gene Description
chrX	651431	651544	1522.449951	-414	YJR123W	RPS5	Rps5p
chrXIV	331018	331136	1492.650024	-245	YNL162W	RPL42A	ribosomal 60S subunit protein L42A
chrII	444786	444903	1452.900024	-151	YBR101C	FES1	Fes1p
chrIV	117331	117526	1430.25	-236	YDL191W	RPL35A	ribosomal 60S subunit protein L35A
chrXI	258381	258514	1426	466	YKL097C		
chrV	396407	396542	1397.649902	-295	YER117W	RPL23B	ribosomal 60S subunit protein L23B
chrVII	398008	398149	1369.199951	-460	YGL056C	SDS23	Sds23p
chrVII	920044	920244	1312.5	-431	YGR214W	RPS0A	ribosomal 40S subunit protein S0A
chrIV	974392	974530	1268.449951	-218	YDR258C	HSP78	chaperone ATPase HSP78
chrXI	533356	533534	1159.199951	19	YKR052C	MRS4	Mrs4p
chrIV	976552	976687	1142.199951	610	YDR260C	SWM1	Swm1p

Supplementary Table 3-3: Significant Sir2 occupied peaks

Table lists the chromosome number and the start and stop point of each significant peak. The peak score refers to the number of uniquely mapped reads to that peak location. The table also includes the gene name or region the peak is in or closest to.

***SIR2*OE Sir2-ChIP significant peaks:**				
chr #	**start**	**end**	**Peak Score**	**Annotation**
chr14	6571	6740	3608.3	In TEL14L and YNL338W
chr01	230044	230208	3165.6	In TEL01R
chr15	1	165	2590	In TEL15L
chr04	1525303	1525472	2036.6	In TEL04R and YDR544C
chr01	1	143	1948	In TEL01L
chr07	1	152	1261.8	In TEL07L
chr03	14646	14815	752.7	In HML locus
chr11	656	825	686.2	In TEL11L and YKL225W
chr03	292709	292878	553.4	In HMR locus
chr11	302516	302685	553.4	In and 5' YKLCdelta5
WT Sir2-ChIP significant peaks:				
chr #	**start**	**end**	**Peak Score**	**Annotation**
chr15	1059099	1059324	1297	5' FIT2 and 3' YDR381W-A
chr05	491070	491295	1193.3	5' YER158C, 5' YER158W-A, 3'BUR6
chr12	416740	416965	1141.4	5' RKM5, 5' TIS11
chr07	277170	277395	1037.6	5' RPS2, 5' MON1 , 3' ygl123C-A
chr01	1	221	933.9	In TEL01L
chr05	152626	152851	933.9	5' MNN1
chr07	598745	598970	882	5' MUP1, 3'YGR054W
chr05	461912	462137	830.1	5'FTR1, 5' LSM5, 3' yer145c-a
chr15	79973	80198	830.1	5' RPL25, 5'YGk3
chr02	642940	643165	778.2	In tc(GCA)b, 3' ybr209w, 3' ybrctau2
***sgf73Δ* Sir2 ChIP significant peaks :**				
chr #	**start**	**end**	**Peak Score**	**Annotation**
chr14	6573	6749	3811.8	In TEL14L and YNL338W
chr01	230026	230197	2901.9	In TEL01R
chr13	5449	5625	2803.5	In TEL13L
chr15	1	166	2680.6	In TEL15L
chr01	1	153	2188.7	In TEL01L
chr02	813022	813178	2065.8	In TEL02R
chr04	1525298	1525474	1918.2	In TEL04R and YDR544C
chr07	1	141	1278.8	In TEL07L
chr07	465	641	885.3	In TEL07L

References:

1. Huisinga, K.L. & Pugh, B.F. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. *Mol Cell* **13**, 573-85 (2004).

2. Koutelou, E., Hirsch, C.L. & Dent, S.Y. Multiple faces of the SAGA complex. *Curr Opin Cell Biol* **22**, 374-82 (2010).

3. Helmlinger, D., Hardy, S., Sasorith, S., Klein, F., Robert, F., Weber, C., Miguet, L., Potier, N., Van-Dorsselaer, A., Wurtz, J.M., Mandel, J.L., Tora, L. & Devys, D. Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. *Hum Mol Genet* **13**, 1257-65 (2004).

4. Timmers, H.T. & Tora, L. SAGA unveiled. *Trends Biochem Sci* **30**, 7-10 (2005).

5. Kohler, A., Schneider, M., Cabal, G.G., Nehrbass, U. & Hurt, E. Yeast Ataxin-7 links histone deubiquitination with gene gating and mRNA export. *Nat Cell Biol* **10**, 707-15 (2008).

6. Lee, K.K., Swanson, S.K., Florens, L., Washburn, M.P. & Workman, J.L. Yeast Sgf73/Ataxin-7 serves to anchor the deubiquitination module into both SAGA and Slik(SALSA) HAT complexes. *Epigenetics Chromatin* **2**, 2 (2009).

7. Zhao, Y., Lang, G., Ito, S., Bonnet, J., Metzger, E., Sawatsubashi, S., Suzuki, E., Le Guezennec, X., Stunnenberg, H.G., Krasnov, A., Georgieva, S.G., Schule, R., Takeyama, K., Kato, S., Tora, L. & Devys, D. A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. *Mol Cell* **29**, 92-101 (2008).

8. Bonnet, J., Wang, Y.H., Spedale, G., Atkinson, R.A., Romier, C., Hamiche, A., Pijnappel, W.W., Timmers, H.T., Tora, L., Devys, D. & Kieffer, B. The structural plasticity of SCA7 domains defines their differential nucleosome-binding properties. *EMBO Rep* **11**, 612-8 (2010).

9. Kaeberlein, M., McVey, M. & Guarente, L. The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms. *Genes Dev* **13**, 2570-80 (1999).

10. Blander, G. & Guarente, L. The Sir2 family of protein deacetylases. *Annu Rev Biochem* **73**, 417-35 (2004).

11. Martinez, E., Palhan, V.B., Tjernberg, A., Lymar, E.S., Gamper, A.M., Kundu, T.K., Chait, B.T. & Roeder, R.G. Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors in vivo. *Mol Cell Biol* **21**, 6782-95 (2001).

12. David, G., Abbas, N., Stevanin, G., Durr, A., Yvert, G., Cancel, G., Weber, C., Imbert, G., Saudou, F., Antoniou, E., Drabkin, H., Gemmill, R., Giunti, P., Benomar, A., Wood, N., Ruberg, M., Agid, Y., Mandel, J.L. & Brice, A. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat Genet* **17**, 65-70 (1997).

13. Stevanin, G., Durr, A. & Brice, A. Clinical and molecular advances in autosomal dominant cerebellar ataxias: from genotype to phenotype and physiopathology. *Eur J Hum Genet* **8**, 4-18 (2000).

14. Lebre, A.S. & Brice, A. Spinocerebellar ataxia 7 (SCA7). *Cytogenet Genome Res* **100**, 154-63 (2003).

15. Kharchenko, P.V., Tolstorukov, M.Y. & Park, P.J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* **26**, 1351-9 (2008).

16. Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P. & Pringle, J.R. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* **14**, 953-61 (1998).

17. Amerik, A.Y., Li, S.J. & Hochstrasser, M. Analysis of the deubiquitinating enzymes of the yeast Saccharomyces cerevisiae. *Biol Chem* **381**, 981-92 (2000).

18. Alamgir, M., Erukova, V., Jessulat, M., Azizi, A. & Golshani, A. Chemical-genetic profile analysis of five inhibitory compounds in yeast. *BMC Chem Biol* **10**, 6 (2010).

19. Kaeberlein, M., Kirkland, K.T., Fields, S. & Kennedy, B.K. Sir2-independent life span extension by calorie restriction in yeast. *PLOS Biology* **2**, 1381-1387 (2004).

20. Delaney, J.R., Sutphin, G.L., Dulken, B., Sim, S., Kim, J.R., Robison, B., Schleit, J., Murakami, C.J., Carr, D., An, E.H., Choi, E., Chou, A., Fletcher, M., Jelic, M., Liu, B., Lockshon, D., Moller, R.M., Pak, D.N., Peng, Q., Peng, Z.J., Pham, K.M., Sage, M., Solanky, A., Steffen, K.K., Tsuchiya, M., Tsuchiyama, S., Johnson, S., Raabe, C., Suh, Y., Zhou, Z., Liu, X., Kennedy, B.K. & Kaeberlein, M. Sir2 deletion prevents lifespan extension in 32 long-lived mutants. *Aging Cell* (2011).

21. Darst, R.P., Garcia, S.N., Koch, M.R. & Pillus, L. Slx5 promotes transcriptional silencing and is required for robust growth in the absence of Sir2. *Mol Cell Biol* **28**, 1361-72 (2008).

22. Shukla, A., Bajwa, P. & Bhaumik, S.R. SAGA-associated Sgf73p facilitates formation of the preinitiation complex assembly at the promoters either in a HAT-dependent or independent manner in vivo. *Nucleic Acids Res* **34**, 6225-32 (2006).

23. Shukla, A., Stanojevic, N., Duan, Z., Sen, P. & Bhaumik, S.R. Ubp8p, a histone deubiquitinase whose association with SAGA is mediated by Sgf11p, differentially regulates lysine 4 methylation of histone H3 in vivo. *Mol Cell Biol* **26**, 3339-52 (2006).

24. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).

25. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. & Glass, C.K. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-89 (2010).

26. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. & Haussler, D. The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).

27. Muhlrad, D. & Parker, R. Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. *RNA* **5**, 1299-307 (1999).

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

29. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).

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

31. Kaeberlein, M., Powers, R.W., 3rd, Steffen, K.K., Westman, E.A., Hu, D., Dang, N., Kerr, E.O., Kirkland, K.T., Fields, S. & Kennedy, B.K. Regulation of yeast replicative life span by TOR and Sch9 in response to nutrients. *Science* **310**, 1193-6 (2005).

32. Steffen, K.K., MacKay, V.L., Kerr, E.O., Tsuchiya, M., Hu, D., Fox, L.A., Dang, N., Johnston, E.D., Oakes, J.A., Tchao, B.N., Pak, D.N., Fields, S., Kennedy, B.K. & Kaeberlein, M. Yeast life span extension by depletion of 60s ribosomal subunits is mediated by Gcn4. *Cell* **133**, 292-302 (2008).

33. Xiao, L. & Grove, A. Coordination of Ribosomal Protein and Ribosomal RNA Gene Expression in Response to TOR Signaling. *Curr Genomics* **10**, 198-205 (2009).

34. Tkach, J.M., Yimit, A., Lee, A.Y., Riffle, M., Costanzo, M., Jaschob, D., Hendry, J.A., Ou, J., Moffat, J., Boone, C., Davis, T.N., Nislow, C. & Brown, G.W. Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. *Nat Cell Biol* **14**, 966-76 (2012).

35. Cupp, J.R. & McAlister-Henn, L. Cloning and characterization of the gene encoding the IDH1 subunit of NAD(+)-dependent isocitrate dehydrogenase from Saccharomyces cerevisiae. *J Biol Chem* **267**, 16417-23 (1992).

36. Graham, T.R., Seeger, M., Payne, G.S., MacKay, V.L. & Emr, S.D. Clathrin-dependent localization of alpha 1,3 mannosyltransferase to the Golgi complex of Saccharomyces cerevisiae. *J Cell Biol* **127**, 667-78 (1994).

37. Lussier, M., Sdicu, A.M. & Bussey, H. The KTR and MNN1 mannosyltransferase families of Saccharomyces cerevisiae. *Biochim Biophys Acta* **1426**, 323-34 (1999).

38. Yip, C.L., Welch, S.K., Klebl, F., Gilbert, T., Seidel, P., Grant, F.J., O'Hara, P.J. & MacKay, V.L. Cloning and analysis of the Saccharomyces cerevisiae MNN9 and MNN1 genes required for complex glycosylation of secreted proteins. *Proc Natl Acad Sci U S A* **91**, 2723-7 (1994).

39. Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S. & O'Shea, E.K. Global analysis of protein localization in budding yeast. *Nature* **425**, 686-91 (2003).

40. Martinez, P. & Persson, B.L. Identification, cloning and characterization of a derepressible Na+-coupled phosphate transporter in Saccharomyces cerevisiae. *Mol Gen Genet* **258**, 628-38 (1998).

41. Teytelman, L., Ozaydin, B., Zill, O., Lefrancois, P., Snyder, M., Rine, J. & Eisen, M.B. Impact of chromatin structures on DNA processing for genomic analyses. *PLoS One* **4**, e6700 (2009).

42. Kosugi, A., Koizumi, Y., Yanagida, F. & Udaka, S. MUP1, high affinity methionine permease, is involved in cysteine uptake by Saccharomyces cerevisiae. *Biosci Biotechnol Biochem* **65**, 728-31 (2001).

43. Isnard, A.D., Thomas, D. & Surdin-Kerjan, Y. The study of methionine uptake in Saccharomyces cerevisiae reveals a new family of amino acid permeases. *J Mol Biol* **262**, 473-84 (1996).

44. Stearman, R., Yuan, D.S., Yamaguchi-Iwai, Y., Klausner, R.D. & Dancis, A. A permease-oxidase complex involved in high-affinity iron uptake in yeast. *Science* **271**, 1552-7 (1996).

45. De Silva, D.M., Askwith, C.C., Eide, D. & Kaplan, J. The FET3 gene product required for high affinity iron transport in yeast is a cell surface ferroxidase. *J Biol Chem* **270**, 1098-101 (1995).

46. Askwith, C., Eide, D., Van Ho, A., Bernard, P.S., Li, L., Davis-Kaplan, S., Sipe, D.M. & Kaplan, J. The FET3 gene of S. cerevisiae encodes a multicopper oxidase required for ferrous iron uptake. *Cell* **76**, 403-10 (1994).

47. Yuan, D.S., Stearman, R., Dancis, A., Dunn, T., Beeler, T. & Klausner, R.D. The Menkes/Wilson disease gene homologue in yeast provides copper to a ceruloplasmin-like oxidase required for iron uptake. *Proc Natl Acad Sci U S A* **92**, 2632-6 (1995).

48. de Silva, D., Davis-Kaplan, S., Fergestad, J. & Kaplan, J. Purification and characterization of Fet3 protein, a yeast homologue of ceruloplasmin. *J Biol Chem* **272**, 14208-13 (1997).

49. Venters, B.J., Wachi, S., Mavrich, T.N., Andersen, B.E., Jena, P., Sinnamon, A.J., Jain, P., Rolleri, N.S., Jiang, C., Hemeryck-Walsh, C. & Pugh, B.F. A comprehensive genomic binding map of gene and chromatin regulatory proteins in Saccharomyces. *Mol Cell* **41**, 480-92 (2011).

50. Herzig, S., Raemy, E., Montessuit, S., Veuthey, J.L., Zamboni, N., Westermann, B., Kunji, E.R. & Martinou, J.C. Identification and functional expression of the mitochondrial pyruvate carrier. *Science* **337**, 93-6 (2012).

51. Otero, J.M., Papadakis, M.A., Udatha, D.B., Nielsen, J. & Panagiotou, G. Yeast biological networks unfold the interplay of antioxidants, genome and phenotype, and reveal a novel regulator of the oxidative stress response. *PLoS One* **5**, e13606 (2010).

52. Bricker, D.K., Taylor, E.B., Schell, J.C., Orsak, T., Boutron, A., Chen, Y.C., Cox, J.E., Cardon, C.M., Van Vranken, J.G., Dephoure, N., Redin, C., Boudina, S., Gygi, S.P., Brivet, M., Thummel, C.S. & Rutter, J. A

mitochondrial pyruvate carrier required for pyruvate uptake in yeast, Drosophila, and humans. *Science* **337**, 96-100 (2012).

53. Colca, J.R., McDonald, W.G., Cavey, G.S., Cole, S.L., Holewa, D.D., Brightwell-Conrad, A.S., Wolfe, C.L., Wheeler, J.S., Coulter, K.R., Kilkuskie, P.M., Gracheva, E., Korshunova, Y., Trusgnich, M., Karr, R., Wiley, S.E., Divakaruni, A.S., Murphy, A.N., Vigueira, P.A., Finck, B.N. & Kletzien, R.F. Identification of a mitochondrial target of thiazolidinedione insulin sensitizers (mTOT)--relationship to newly identified mitochondrial pyruvate carrier proteins. *PLoS One* **8**, e61551 (2013).

54. Cai, L., Sutter, B.M., Li, B. & Tu, B.P. Acetyl-CoA induces cell growth and proliferation by promoting the acetylation of histones at growth genes. *Mol Cell* **42**, 426-37 (2011).

55. Downey, M., Knight, B., Vashisht, A.A., Seller, C.A., Wohlschlegel, J.A., Shore, D. & Toczyski, D.P. Gcn5 and sirtuins regulate acetylation of the ribosomal protein transcription factor ifh1. *Curr Biol* **23**, 1638-48 (2013).

56. Cai, L., McCormick, M.A., Kennedy, B.K. & Tu, B.P. Integration of multiple nutrient cues and regulation of lifespan by ribosomal transcription factor ifh1. *Cell Rep* **4**, 1063-71 (2013).

57. Hinnebusch, A.G. Translational regulation of GCN4 and the general amino acid control of yeast. *Annu Rev Microbiol* **59**, 407-50 (2005).

58. Zhao, Y., Sohn, J.H. & Warner, J.R. Autoregulation in the biosynthesis of ribosomes. *Mol Cell Biol* **23**, 699-707 (2003).

59. Miyoshi, K., Tsujii, R., Yoshida, H., Maki, Y., Wada, A., Matsui, Y., Toh, E.A. & Mizuta, K. Normal assembly of 60 S ribosomal subunits is required for the signaling in response to a secretory defect in Saccharomyces cerevisiae. *J Biol Chem* **277**, 18334-9 (2002).

60. Mushegian, A.R., Vishnivetskiy, S.A. & Gurevich, V.V. Conserved phosphoprotein interaction motif is functionally interchangeable between ataxin-7 and arrestins. *Biochemistry* **39**, 6809-13 (2000).

61. Scheel, H., Tomiuk, S. & Hofmann, K. Elucidation of ataxin-3 and ataxin-7 function by integrative bioinformatics. *Hum Mol Genet* **12**, 2845-52 (2003).

62. Rauh, D., Fischer, F., Gertz, M., Lakshminarasimhan, M., Bergbrede, T., Aladini, F., Kambach, C., Becker, C.F., Zerweck, J., Schutkowski, M. & Steegborn, C. An acetylome peptide microarray reveals specificities and deacetylation substrates for all human sirtuin isoforms. *Nat Commun* **4**, 2327 (2013).

63. Garcia, S.N. & Pillus, L. A unique class of conditional sir2 mutants displays distinct silencing defects in Saccharomyces cerevisiae. *Genetics* **162**, 721-36 (2002).

64. Huang da, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).

65. Zhang, B., Kirov, S. & Snoddy, J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* **33**, W741-8 (2005).

66. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res* **41**, W77-83 (2013).

Chapter 4: Future directions

The function of ataxin 7 in normal cells and SCA7 expanded is multifaceted. In this thesis, a multi-pronged approach was taken to shed light on this complexity. Using yeast as a model organism we studied the ataxin 7 yeast ortholog Sgf73. We explored the impact of Sgf73 on replicative lifespan, identifying a link between Sgf73 and Sir2. In addition using a ChIP-seq approach we identified 389 genomic regions bound by Sgf73 and found that there was an enrichment of binding in the region 5' to many ribosomal protein encoding genes. These studies have created a solid basis for further study of the ataxin 7 normal function in both yeast and mammalian model organisms.

Future studies in yeast:

Ribosomal protein regulation:

Our studies have highlighted the role of Sgf73 in the regulation of ribosomal protein (RP) gene transcription. We find that in *sgf73Δ* mutants there is a reduction in the transcript levels of RP genes. It was previously known that SAGA binds at RP genes and that there is a reduction of SAGA occupancy at a few RP genes upon ChIP of SAGA structural domain protein Spt7 in *sgf73Δ* mutants[1]. More recent reports show that Gcn5, the SAGA member harboring histone acetyltransferase activity, acetylates the RP gene transcription co-activator Ifh1 (Interacts with ForkHead) to stabilize and reduce RP gene transcription. Ifh1 is subsequently de-acetylated by Sir2 [2,3]. Interplay between the dynamic natures of Gcn5 and Sir2 activities on Ifh1 regulates RP gene transcription. As we have shown that there is a physical interaction between Sgf73 and

Sir2 (chapter 2), we hypothesize that Sgf73 is a key component of the modulation of Ifh1 acetylation through Sir2 binding.

In our model (Figure 4-1) deletion of *SGF73* results in inefficient recruitment of Sir2 to Ifh1, causing Ifh1 to remain in its less active, stably acetylated form. This then reduces RP transcription in *sgf73Δ* mutants in basal growth conditions. This model is supported by our findings that there are reduced RP transcript levels in *sgf73Δ* mutants (Figure 3-14).

Many signaling pathways, including the Target Of Rapamycin (TOR) kinase pathway, regulate RP gene transcription (reviewed in [4]). Rap1[5,6] and Fhl1[7] bind upstream of RP genes and Ifh1[8] is subsequently recruited to activate RP gene transcription. Ifh1 binding is modulated by TOR as treatment with the TOR inhibitor rapamycin causes a loss of Ifh1 binding upstream of RP genes but does not change Rap1 and Fhl1 binding[8].

We find that Sgf73 also plays a role in ribosomal biogenesis under cellular stressor rapamycin. Upon treatment with rapamycin, RP gene transcription is reduced in WT and to a lesser extent in *sgf73Δ* mutant cells. Elevated levels of RP gene transcription in *sgf73*Δ under rapamycin treatment may provide a growth advantage under stress conditions. Under rapamycin stress Ifh1 is found in its deacetylated form[2] in WT cells, suggesting complete dissociation from Fhl1 and thus lack of RP gene transcription. Again, the muted response to rapamycin-induced down regulation of RP gene transcripts we detected agrees with our model. In the absence of *SGF73* there is inefficient recruitment of Sir2 to deacetylate Ifh1.

Essential to validate our predicted mechanistic model of RP gene transcription regulation by Sgf73 is to measure Ifh1 acetylation levels in *sgf73Δ* mutants and compare them to WT. We hypothesize that acetylation will be higher in *sgf73Δ* mutants in comparison to WT cells.

Mitochondrial candidates:

Aside from the striking enrichment of RP genes occupied by Sgf73, our studies led us to additional high confidence targets linked to RLS. Targets *FMP43*, *IDH1*, *MNN1*, and *YBR238C* are either directly linked to mitochondrial function or in the case of *MNN1* indirectly linked, though modification of proteins moving through the secretory pathway that affect mitochondrial function. In our studies we found that there was a significant reduction of *FMP43*, *MNN1*, and *YBR238C* in *sgf73Δ* mutants at both the RNA and protein levels (Chapter 3).

The mitochondria are the main producers of energy in the cell, through the tightly coupled tricarboxylic acid (TCA) cycle and electron transport chain, producing chemical energy in the form of ATP. *FMP43* is a mitochondrial pyruvate carrier [9 10], and reduced levels likely limit the transport of pyruvate into the mitochondria. Pyruvate in the mitochondria is converted into oxaloacetate, a molecule that feeds directly into the TCA cycle, or alternatively, is converted into acetyl-CoA, which acts as an acceptor in the TCA cycle. *IDH1*, a subunit of a mitochondrial NAD(+)-dependent isocitrate dehydrogenase, catalyzes the oxidation of isocitrate to alpha-ketoglutarate in the TCA cycle [11]. We detected a trend towards increased levels of *IDH1*transcript and protein levels in *sgf73Δ* mutants.

The combination of limiting pyruvate and increasing isocitrate dehydrogenase is predicted to alter the TCA cycle. It would be interesting to explore the rate of the TCA cycle in *sgf73Δ* mutants to address if cellular energy production is different, and how Sgf73 occupation on TCA cycle- related genes affects metabolism .

YBR238C is a mitochondrial membrane protein of unknown function. It is known to be transcriptionally up-regulated by TOR[12]. When we inhibited TOR with rapamycin treatment, indeed *YBR238C* transcript levels plummeted like RP genes, but increased transcript levels persisted in *sgf73Δ* mutants compared to WT ($p = 0.06$). *YBR238C* has a yeast paralog, *RMD9,* in which they share 45% homology (as predicted DNA sequence)[13]. Rmd9 does not have Sgf73 occupancy but has been shown to be important for the expression of mitochondrial genes encoding respiratory complex subunits through its proposed role in delivering mitochondrial mRNAs to ribosomes[13,14]. Studying how *YBR238C* functions will illuminate if it acts similarly to its paralog, and its connection to ribosomes and the mitochondria involved in Sgf73-dependent regulation and aging.

Another line of investigation that would be informative to uncover our targets' relationships with *SGF73* involves studying *sgf73Δ* mutant phenotypes. Translational inhibitor cycloheximide inhibits *sgf73*Δ growth. Cycloheximide sensitivity in *sgf73Δ* mutants in combination with overexpression, or deletion of candidate targets should be tested.

Unexplored Sgf73 occupancy targets:

Our studies have investigated top candidates that have previously been linked to RLS. To find other strong candidates, we utilized microarray data to identify genes with significant transcript reductions in *sgf73Δ* mutants that were enriched in a specific GO term category. Still, many other targets with Sgf73 occupancy may be important to aging and their roles have yet to be explored. In particular there are a number of genes involved in ion transport, including iron transport. Iron-sulphur-cluster biogenesis has previously been linked to aging in relation to genomic instability [15], and thus altered iron transport may be another attribute that *sgf73Δ* mutants have that contributes to their RLS extension.

Future studies in a mammalian model system:

A goal of our studies on Sgf73 occupancy has been to connect our findings on Sgf73 transcriptional regulation back to the function of the human ortholog ataxin 7, and ultimately to SCA7. Our hypothesis is that targets of Sgf73 are also targets of ataxin 7, and through our Sgf73 occupancy studies we will garner an understanding of transcriptional dysregulation in SCA7 neurodegeneration. A number of Sgf73 occupancy targets have human and mouse homologues. A preliminary study was carried out in human fibroblast cells from a control and two SCA7 affected individuals. Expression of thirteen human homologs of genes occupied by Sgf73 was measured by qRT-PCR.(Figure 3-28). The findings were very encouraging; there was transcriptional dysregulation of these homologous targets in SCA7. To further study the connections between our results in yeast and higher organisms our key targets

should be studied in more detail in a mouse model of SCA7 as well as in patient and control fibroblast cells.

Future studies in yeast and mammalian models:

Levels of histone H2BK123 ubiquitination are elevated in yeast *sgf73*Δ mutants, and substituting this residue with one that cannot be ubiquitinated results in a shortened RLS. Recent studies in Drosophila and HeLa cells seemingly contradict studies in yeast as it was reported that siRNA to ataxin 7 resulted in lower levels of H2BK123 ubiquitination[16]. It is proposed that the deubiquitinase module functions even in the absence of ataxin 7, contrary to what has been observed with Sgf73 absence in the Ubp8 deubiquitinase module[16]. Clearly H2BK123 ubiquitination is important for *sgf73*Δ RLS extension, however the authors raise an important point that is applicable when considering yeast Sgf73 function; in the absence of *SGF73,* other proteins may be inappropriately deubiquitinated by Ubp8. In this case, studies in higher organisms may help to guide future studies in yeast, including determining the location of Ubp8 in the absence of *SGF73* and searching for other Ubp8 targets besides H2BK123Ub. Also, levels of H2BUb should be measured in SCA7 patients to sort out if lowered levels of ataxin 7 by siRNA recapitulate the disease state.

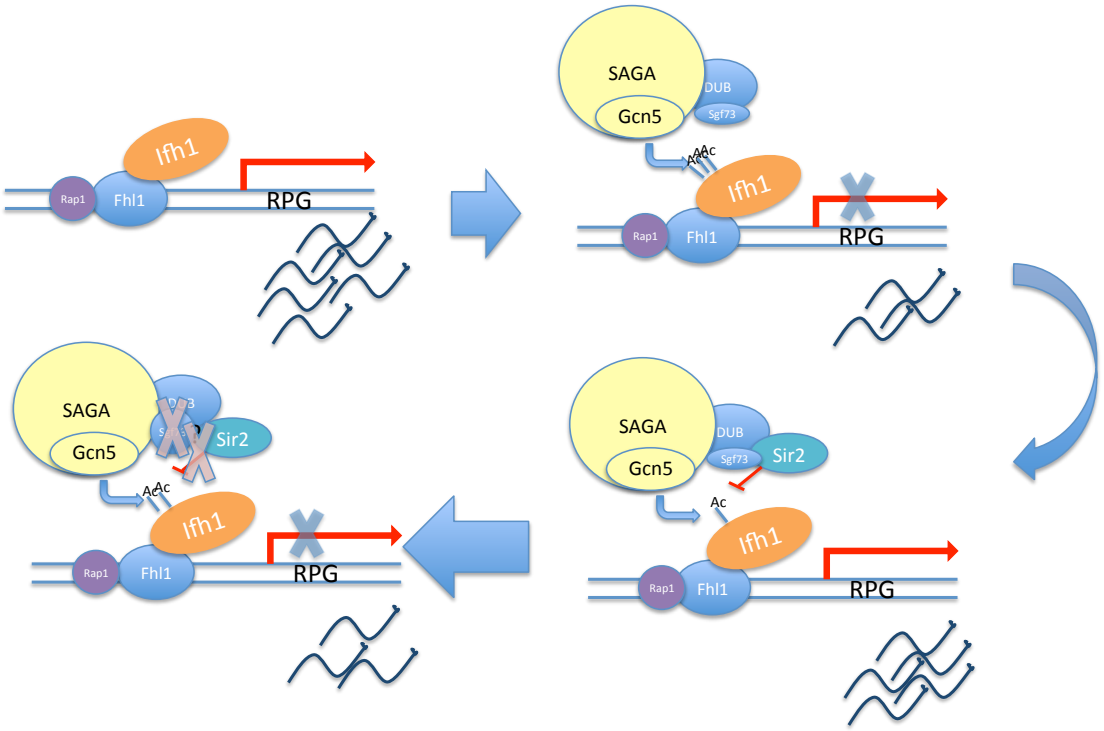


Figure 4-1: Model of Ifh1 regulation of ribosomal protein gene transcription
One mechanism by which ribosomal protein gene (RPG) transcription is regulated through the recruitment of the transcription factor Ifh1that turns on RPG transcription. Acetylation by Gcn5 and deacetylation by Sir2 of Ifh1 modulates transcription of RPGs. Our model proposes that modulation of Ifh1 acetylation is altered in *sgf73Δ* mutants through inefficient recruitment of Sir2 to deacetylate Ifh1 resulting in lower RPG transcription.

References:

1. Cai, L., Sutter, B.M., Li, B. & Tu, B.P. Acetyl-CoA induces cell growth and proliferation by promoting the acetylation of histones at growth genes. *Mol Cell* **42**, 426-37 (2011).

2. Downey, M., Knight, B., Vashisht, A.A., Seller, C.A., Wohlschlegel, J.A., Shore, D. & Toczyski, D.P. Gcn5 and sirtuins regulate acetylation of the ribosomal protein transcription factor ifh1. *Curr Biol* **23**, 1638-48 (2013).

3. Cai, L., McCormick, M.A., Kennedy, B.K. & Tu, B.P. Integration of multiple nutrient cues and regulation of lifespan by ribosomal transcription factor ifh1. *Cell Rep* **4**, 1063-71 (2013).

4. Xiao, L. & Grove, A. Coordination of Ribosomal Protein and Ribosomal RNA Gene Expression in Response to TOR Signaling. *Curr Genomics* **10**, 198-205 (2009).

5. Shore, D. RAP1: a protean regulator in yeast. *Trends Genet* **10**, 408-12 (1994).

6. Lieb, J.D., Liu, X., Botstein, D. & Brown, P.O. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat Genet* **28**, 327-34 (2001).

7. Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.L., Gordon, D.B., Ren, B., Wyrick, J.J., Tagne, J.B., Volkert, T.L., Fraenkel, E., Gifford, D.K. & Young, R.A. Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* **298**, 799-804 (2002).

8. Schawalder, S.B., Kabani, M., Howald, I., Choudhury, U., Werner, M. & Shore, D. Growth-regulated recruitment of the essential yeast ribosomal protein gene activator Ifh1. *Nature* **432**, 1058-61 (2004).

9. Tkach, J.M., Yimit, A., Lee, A.Y., Riffle, M., Costanzo, M., Jaschob, D., Hendry, J.A., Ou, J., Moffat, J., Boone, C., Davis, T.N., Nislow, C. & Brown, G.W. Dissecting DNA damage response pathways by analysing protein

localization and abundance changes during DNA replication stress. *Nat Cell Biol* **14**, 966-76 (2012).

10. Otero, J.M., Papadakis, M.A., Udatha, D.B., Nielsen, J. & Panagiotou, G. Yeast biological networks unfold the interplay of antioxidants, genome and phenotype, and reveal a novel regulator of the oxidative stress response. *PLoS One* **5**, e13606 (2010).

11. Cupp, J.R. & McAlister-Henn, L. Cloning and characterization of the gene encoding the IDH1 subunit of NAD(+)-dependent isocitrate dehydrogenase from Saccharomyces cerevisiae. *J Biol Chem* **267**, 16417-23 (1992).

12. Shamji, A.F., Kuruvilla, F.G. & Schreiber, S.L. Partitioning the transcriptional program induced by rapamycin among the effectors of the Tor proteins. *Curr Biol* **10**, 1574-81 (2000).

13. Nouet, C., Bourens, M., Hlavacek, O., Marsy, S., Lemaire, C. & Dujardin, G. Rmd9p controls the processing/stability of mitochondrial mRNAs and its overexpression compensates for a partial deficiency of oxa1p in Saccharomyces cerevisiae. *Genetics* **175**, 1105-15 (2007).

14. Williams, E.H., Butler, C.A., Bonnefoy, N. & Fox, T.D. Translation initiation in Saccharomyces cerevisiae mitochondria: functional interactions among mitochondrial ribosomal protein Rsm28p, initiation factor 2, methionyl-tRNA-formyltransferase and novel protein Rmd9p. *Genetics* **175**, 1117-26 (2007).

15. Veatch, J.R., McMurray, M.A., Nelson, Z.W. & Gottschling, D.E. Mitochondrial dysfunction leads to nuclear genome instability via an iron-sulfur cluster defect. *Cell* **137**, 1247-58 (2009).

16. Mohan, R.D., Dialynas, G., Weake, V.M., Liu, J., Martin-Brown, S., Florens, L., Washburn, M.P., Workman, J.L. & Abmayr, S.M. Loss of Drosophila Ataxin-7, a SAGA subunit, reduces H2B ubiquitination and leads to neural and retinal degeneration. *Genes Dev* **28**, 259-72 (2014).

Appendix: Sgf73 occupied genes with introns

In the *S. cerevisiae* genome there are 282 open reading frames (ORFs) that contain introns as identified by the Ares Lab Yeast Intron Database [1]. A large number of these genes are ribosomal protein genes (RPGs). RPGs are very highly transcribed genes and account for much of the population of intron containing transcripts that undergo splicing. Thus alterations in ribosomal transcript levels can have global effects on splicing of transcripts of cells. This has been shown by work demonstrating that limiting RP gene transcription increases splicing efficiency of other transcripts by reducing competition for splicing factors[2]. Due to the enrichment of RPGs in the Sgf73 occupancy data (chapter 3) we were interested in investigating the portion of Sgf73 occupied genes that contain introns. Introns are found in 70% of large ribosomal subunit transcripts and 78% of small ribosomal subunits[3].

The Ares intron database lists 282 intron containing genes (http://compbio.soe.ucsc.edu/yeast_introns.html)[1]. These genes with introns account for 4.2% (282/6607) of all ORFs in the *S. cerevisiae* genome. The list of intron containing genes (ICGs) available through the Ares database, as well as a list of ribosomal protein genes (RPGs) with introns was used to analyze intron-containing genes that have Sgf73 occupancy (Supplementary table 3-1). The list of RPGs with introns was obtained by inputting the Ares ICGs list into DAVID (Database for Annotation, Visualization and Integrated Discovery)[4,5], and using their KEGG[6,7] pathway analysis. The KEGG pathway analysis gave an output of 83 intron containing RPGs (Table A1). Recent work in Munding et al[2] suggested this list is missing a

portion of intron containing genes however, the complete list was not available for this analysis.

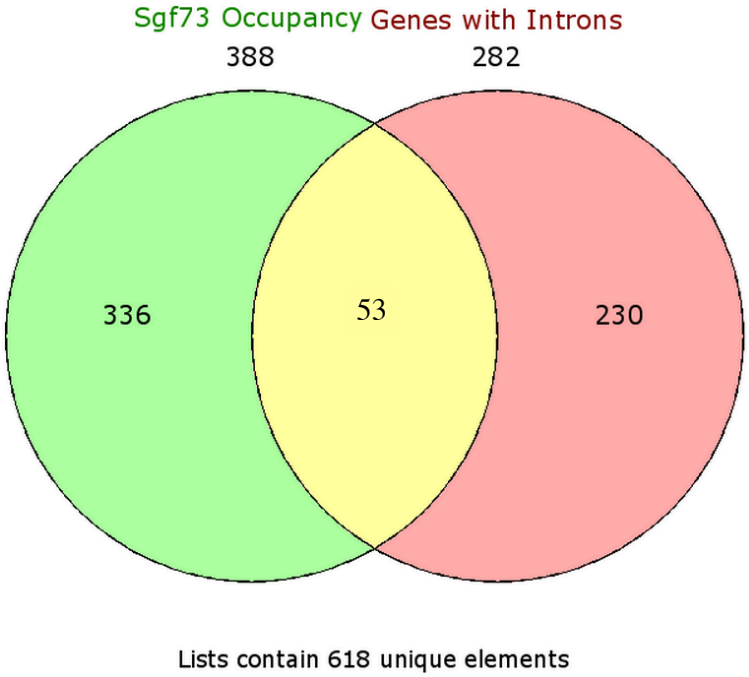
We find that there is an enrichment of ICGs in the list of genes with Sgf73 occupancy (Figure A1 A, B), with 53 intron containing genes having Sgf73 occupancy (p (>=53)= 9.75 e-15 by hypogeometric test). Additionally the majority of Sgf73 occupied RPGs have introns (Figure A2), 45 out of 57, corresponding to 79%. Since deletion of some RPGs extends replicative lifespan (RLS) in yeast we evaluated Sgf73 occupancy on intron containing RPGs. There is not a further increase in intron containing RPGs occupied by Sgf73 that are connected to RLS as 16 out of 19 RLS-linked RPGs have introns. This corresponds to ~ 84%, which is very similar to the 79% of Sgf73 occupied RPGs that have introns (Figure A2 orange boxes). These findings however are not far from what we would expect given that most RPGs have introns. In addition to ribosomal protein genes containing introns there was Sgf73 occupancy at an additional eight intron containing genes (Table A2).

From our previous studies in chapter 3 we have examined the expression of 31 Sgf73 occupied RPGs by qPCR, 19 of which when deleted cause RLS extension. Of those investigated RPGs 24 have introns. All 24 intron containing genes trended towards having reduced RNA expression in *sgf73Δ* mutants with 13 having significantly reduced expression (Figure A3; n=6 $p<0.05$).

Table A-1: Intron containing ribosomal protein genes

ID	Gene Name	ID	Gene Name
3127777	40S ribosomal protein S0-A	3124765	60S ribosomal protein L16-A
3127102	40S ribosomal protein S0-B	3127521	60S ribosomal protein L16-B
3126016	40S ribosomal protein S10-A	3123769	60S ribosomal protein L17-A
3129530	40S ribosomal protein S10-B	3125042	60S ribosomal protein L17-B
3123233, 3123233	40S ribosomal protein S11	3124027, 3124027	60S ribosomal protein L18
3125525	40S ribosomal protein S13	3125206, 3125206	60S ribosomal protein L19
3129220	40S ribosomal protein S14-A	3123909, 3123909	60S ribosomal protein L2
3122707	40S ribosomal protein S14-B	3123308, 3123308	60S ribosomal protein L20
3124908, 3124908	40S ribosomal protein S16	3123757	60S ribosomal protein L21-A
3122853	40S ribosomal protein S17-A	3128683	60S ribosomal protein L21-B
3127737	40S ribosomal protein S17-B	3126472	60S ribosomal protein L22-A
3129721, 3129721	40S ribosomal protein S18	3125062	60S ribosomal protein L22-B
3126041	40S ribosomal protein S19-A	3126543, 3126543	60S ribosomal protein L23
3125102	40S ribosomal protein S19-B	3125177	60S ribosomal protein L24-A
3122937	40S ribosomal protein S21-A	3122744	60S ribosomal protein L24-B
3128402	40S ribosomal protein S21-B	3128374	60S ribosomal protein L25
3123054	40S ribosomal protein S22-B	3124067	60S ribosomal protein L26-A
3126010, 3126010	40S ribosomal protein S23	3123368	60S ribosomal protein L26-B
3126837, 3126837	40S ribosomal protein S24	3128151	60S ribosomal protein L27-A
3123972	40S ribosomal protein S25-A	3123955	60S ribosomal protein L27-B
3127696	40S ribosomal protein S25-B	3128121	60S ribosomal protein L28
3128673	40S ribosomal protein S26-A	3126658	60S ribosomal protein L29
3127032	40S ribosomal protein S26-B	3128613	60S ribosomal protein L30
3129194	40S ribosomal protein S27-A	3128650	60S ribosomal protein L31-A
3127745	40S ribosomal protein S27-B	3127013	60S ribosomal protein L31-B
3124772	40S ribosomal protein S29-A	3125689	60S ribosomal protein L32
3127473	40S ribosomal protein S29-B	3124590	60S ribosomal protein L33-A
3125207, 3125207	40S ribosomal protein S30	3127270	60S ribosomal protein L33-B
3125664, 3125664	40S ribosomal protein S4	3123714	60S ribosomal protein L34-A
3126492, 3126492	40S ribosomal protein S6	3127408	60S ribosomal protein L34-B
3129538	40S ribosomal protein S7-A	3128509, 3128509	60S ribosomal protein L35
3125390	40S ribosomal protein S7-B	3129610	60S ribosomal protein L36-A
3126357, 3126357	40S ribosomal protein S8	3123662	60S ribosomal protein L36-B
3124369	40S ribosomal protein S9-A	3126298	60S ribosomal protein L37-A
3124324	40S ribosomal protein S9-B	3129351	60S ribosomal protein L37-B
3124945	60S acidic ribosomal protein P1-beta	3122660	60S ribosomal protein L39
3128205	60S ribosomal protein L13-A	3122967, 3122967	60S ribosomal protein L40; Ubiquitin
3127712	60S ribosomal protein L13-B	3126157, 3126157	60S ribosomal protein L42
3126771	60S ribosomal protein L14-A	3128355, 3128355	60S ribosomal protein L43
3129071	60S ribosomal protein L14-B	3123103	60S ribosomal protein L6-A
3126998	60S ribosomal protein L7-A	3123372	60S ribosomal protein L6-B
3126533	60S ribosomal protein L7-B		

A)



B)

YBL027W	YDL191W	YGL103W	YKL006W	YMR116C	YPL249C-A
YBL087C	YDR447C	YGL178W	YKL180W	YMR142C	YPR043W
YBR048W	YDR471W	YGR214W	YLR048W	YNL069C	YMR242C
YBR082C	YDR500C	YIL052C	YLR061W	YNL096C	
YBR084C-A	YER074W	YIL123W	YLR344W	YNL162W	
YBR101C	YER102W	YIL148W	YLR448W	YNL302C	
YBR189W	YER117W	YJL130C	YML025C	YOR096W	
YCR031C	YFR031C-A	YJL136C	YML026C	YOR312C	
YDL075W	YGL030W	YJL189W	YML034W	YPL081W	
YDL083C	YGL076C	YJR094W-A	YML073C	YPL143W	

Figure A-1: Significant overlap of Sgf73 occupied genes and genes with introns
A) Venn diagram overlapping all Sgf73 occupied genes and all genes in the genome that have an intron, overlap is significant by hypogeometric test p(>=53) = 9.75 e-15. B) List of genes with Sgf73 occupancy that contain introns

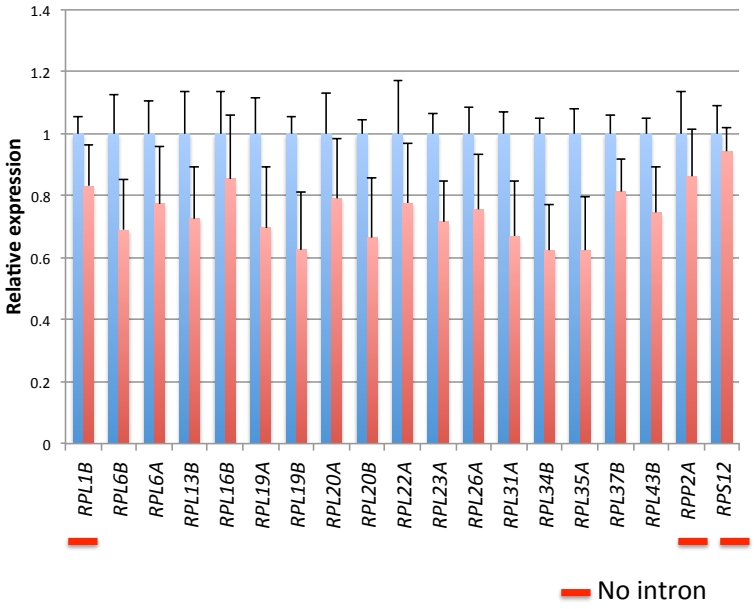
RPL13B	RPL26A	RPL39	RPS14A
RPL14A	RPL27B	RPL40A	RPS16B
RPL16B	RPL28	RPL42A	RPS17B
RPL17A	RPL2A	RPL43A	RPS18B
RPL19A	RPL30	RPL43B	RPS19B
RPL19B	RPL31A	RPL6A	RPS21B
RPL20A	RPL33A	RPL6B	RPS24A
RPL20B	RPL34B	RPL7A	RPS7A
RPL22A	RPL35A	RPS0A	RPS7B
RPL23A	RPL36B	RPS0B	RPS8B
RPL23B	RPL37B	RPS11B	RPS9A
			RPS9B

Figure A-2: Chart of all intron containing RP genes with Sgf73 occupancy
Chart lists all RP genes that are present in the Sgf73 occupancy data and that also have introns. Orange boxes highlight intron containing RP genes that are RLS-linked though conferring extended RLS upon deletion.

Table A-2: Non ribosomal intron containing genes with Sgf73 occupancy
Description from GO database

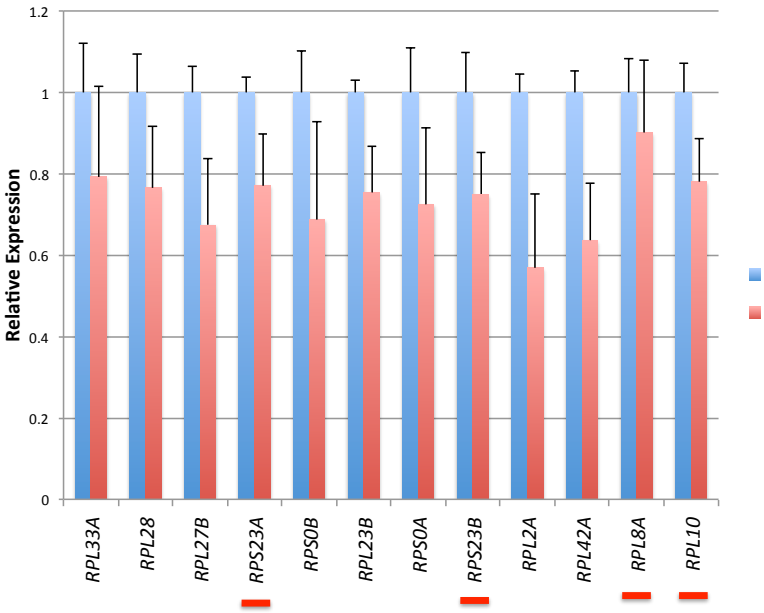
Gene Name	Gene Description
ASC1	Abundant protein with effects on translational efficiency and cell size, has two WD (WD-40) repeats.
FES1	Hsp70 (Ssa1p) nucleotide exchange factor, cytosolic homolog of Sil1p, which is the nucleotide exchange factor for BiP (Kar2p) in the endoplasmic reticulum
MPT5	Protein required for high temperature growth, recovery from alpha-factor arrest, and normal lifespan of yeast cells.
SIM1	(putative) involved in control of DNA replication
SRC1	Protein of unknown function, expression is cell cycle-regulated
UBC4	Ubiquitin-conjugating enzyme that with Ubc5p is responsible for most of ubiquitin-dependent protein degradation of short-lived and abnormal proteins.
YJL130C	Bifunctional carbamoylphosphate synthetase (CPSase)-aspartate transcarbamylase (ATCase), catalyzes the first two enzymatic steps in the de novo biosynthesis of pyrimidines; both activities are subject to feedback inhibition by UTP
YML025C	Mitochondrial ribosomal protein of the large subunit, belongs to the L4 family of prokaryotic ribosomal proteins.

A)



	P value
RPL1B	0.11682
RPL6B	0.03401
RPL6A	0.11218
RPL13B	0.06782
RPL16B	0.39965
RPL19A	0.05881
RPL19B	0.01071
RPL20A	0.19207
RPL20B	0.02295
RPL22A	0.20124
RPL23A	0.01373
RPL26A	0.08684
RPL31A	0.0198
RPL34B	0.00306
RPL35A	0.00934
RPL37B	0.04842
RPL43B	0.03324
RPP2A	0.33276
RPS12	0.49153

B)



	pvalue
RPL33A	0.24593
RPL28	0.07111
RPL27B	0.01436
RPS23A	0.02732
RPS0B	0.08616
RPL23B	0.01109
RPS0A	0.07656
RPS23B	0.01870
RPL2A	0.00321
RPL42A	0.00318
RPL8A	0.45990
RPL10	0.02785

Figure A-3: Expression analysis of RP genes containing and lacking introns
A) RNA expression analysis of RP genes that are linked to RLS. Intron status is denoted by a red line below the gene name to indicate the gene does not have an intron. B) RNA expression analysis of RP genes that are not linked RLS. Genes without introns are marked with a red line as in A.

References:

1. Grate, L. & Ares, M., Jr. Searching yeast intron data at Ares lab Web site. *Methods Enzymol* **350**, 380-92 (2002).

2. Munding, E.M., Shiue, L., Katzman, S., Donohue, J.P. & Ares, M., Jr. Competition between pre-mRNAs for the splicing machinery drives global regulation of splicing. *Mol Cell* **51**, 338-48 (2013).

3. Parenteau, J., Durand, M., Morin, G., Gagnon, J., Lucier, J.F., Wellinger, R.J., Chabot, B. & Elela, S.A. Introns within ribosomal protein genes regulate the production and function of yeast ribosomes. *Cell* **147**, 320-31 (2011).

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

5. Huang da, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).

6. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).

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