bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The Modular Adaptive Ribosome

Anupama Yadav[1], Aparna Radhakrishnan[1], Anshuman Panda[2], Amartya Singh[2], Himanshu Sinha[1,3,4,*], Gyan Bhanot[2,5,6*]

[1]Department of Biological Sciences, Tata Institute of Fundamental Research, Mumbai, 400005, India

[2]Department of Physics, Rutgers University, Piscataway, New Jersey 08854, USA

[3]Department of Biotechnology, Indian Institute of Technology Madras, Chennai 600036, India

[4]Initiative for Biological Systems Engineering, Indian Institute of Technology Madras, Chennai 600036, India

[5]Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854, USA

[6]The Cancer Institute of New Jersey, New Brunswick, New Jersey 08903, USA

***Authors for correspondence:** Himanshu Sinha, Department of Biotechnology, Indian Institute of Technology Madras, Chennai 600036, India; Tel: +91-44-22574120, Email: sinha@iitm.ac.in; and Gyan Bhanot, Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854, USA; Tel: +1-848-391-7508, Email: gbhanot@rutgers.edu, gyanbhanot@gmail.com

Short title: The Modular Adaptive Ribosome

Keywords: ribosomal proteins, yeast, mice, humans, ribosomal protein heterogeneity, phenotypic modularity, expression modularity, ribosomal protein nucleotide diversity, adaptation

ABSTRACT

The ribosome is an ancient machine, performing the same function across organisms. Although functionally unitary, recent experiments suggest specialized roles for some ribosomal proteins. Our central thesis is that ribosomal proteins function in a modular fashion

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

to decode genetic information in a context dependent manner. We show through large data analyses that although many ribosomal proteins are essential with consistent effect on growth in different conditions in yeast and similar expression across cell and tissue types in mice and humans, some ribosomal proteins are used in an environment specific manner. The latter set of variable ribosomal proteins further function in a coordinated manner forming modules, which are adapted to different environmental cues in different organisms. We show that these environment specific modules of ribosomal proteins in yeast have differential genetic interactions with other pathways and their 5'UTRs show differential signatures of selection in yeast strains, presumably to facilitate adaptation. Similarly, we show that in higher metazoans such as mice and humans, different modules of ribosomal proteins are expressed in different cell types and tissues. A clear example is nervous tissue that uses a ribosomal protein module distinct from the rest of the tissues in both mice and humans. Our results suggest a novel stratification of ribosomal proteins that could have played a role in adaptation, presumably to optimize translation for adaptation to diverse ecological niches and tissue microenvironments.

INTRODUCTION

A single celled organism displays a range of phenotypes to survive in diverse environments. In complex multicellular organisms, in addition to the external environment, tissue specific cell types display specialized mechanisms to regulate phenotype in local tissue environments. Much of the research in biology has been directed towards understanding the basis of the information flow that gives rise to these diverse phenotypes. This has resulted in the identification of many regulatory processes [1,2] which fine-tune transcriptional expression and modulate the translation of mRNA into proteins [3,4] in response to external and environmental or tissue specific signaling cues. However, in spite of its essential role in this cellular information flow, the ribosome has always been regarded as an inert participant in the information flow that regulates cellular and tissue states.

One may ask, "Why is it that, in spite of tantalizing clues to the contrary, this belief in an invariant, environment independent ribosome has not been significantly challenged?" A possible reason might be that the high degree of conservation of ribosomal proteins across the three domains of life, viz., Archaea, Bacteria and Eukarya, and the slow evolution rates of

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

ribosomal protein sequences [5] seem sufficient evidence for an essential, invariant ribosome, which plays no regulatory role in cells and tissues or in adaptation and speciation.

However, thermophiles have ribosomes that function at extreme temperatures, suggesting that the ribosome has been adapting to the environment for over 3 billion years. There are different types of ribosomes in each domain of life, and additionally, the mitochondrial ribosome distinct in composition from the cytosolic ribosome [6]. It is therefore worth investigating whether the cytosolic ribosome has also evolved ways to optimize its composition in response to environmental cues. Testing the ribosomal composition at a protein level is technically challenging, but significant evidence has been accumulating from the study of ribosomal proteins that argues for a variability in the composition of the ribosome [7-9]. Deletion experiments in yeast show that different ribosomal proteins have a differential effect on replicative lifespan [8]. Transcriptional studies in mice [7] and humans [9] show tissue specific expression of ribosomal proteins. Specific ribosomal proteins are known to be associated with different types of cancers [10,11] and mutations in specific ribosomal proteins result in a class of disorders called ribosomopathies [12,13]. Whereas all these mutations have common effects across development, some cause developmental disorders in specific tissues [12,13], suggesting tissue specificity of the function of at least some ribosomal proteins.

To date, most of these effects have been ascribed to extra-ribosomal functions of ribosomal proteins [14,15]. However, at least some specialized translation by the ribosome is controlled by specific ribosomal proteins [7,16,17], showing that the extra-ribosomal nature of such control is not generic. While preliminary, these studies suggest that ribosomes with variable components may exist to optimize translation, depending on environmental and signaling cues [18-20]. Analysis of the stoichiometry of ribosomal proteins in yeast and embryonic stem cells has demonstrated differential protein composition of ribosomes in different conditions [21], further substantiating the possibility of ribosomal variability at a protein level.

In this paper, we question the unitary nature of the ribosome within an organism across environments by analyzing different properties of ribosomal proteins. Does the observed environment and tissue specific variation in ribosomal proteins represent regulation of the ribosome that is important in evolution and or adaptation? In a changing environment, do all ribosomal proteins have similar properties, or are specific ribosomal proteins used in an

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

92 environment dependent manner? In this paper we begin to address these questions by
93 uncovering evidence from data analyses of yeast deletion [22] and interaction [23] datasets,
94 and the ENCODE [24] and GTEx [25] expression datasets of mice and humans.

95

96 **MATERIALS AND METHODS**

97 **Yeast Data**

98 Growth data and associated microarray files for a genome-wide yeast homozygous deletion
99 collection [22] for all environments were downloaded from
100 http://chemogenomics.stanford.edu/supplements/global/download.html. The normalized gene
101 intensity values from the microarray data were used for analysis. The data of the replicates
102 for each environment were collated using their median. The intensity values were scaled to
103 zero mean and unit variance across all genes and then compared between the rich growth
104 (YPD) and a stress condition, as well as between the pairs of stress conditions (S1 Table). A
105 set of 68 non-essential structural ribosomal proteins with microarray tag intensity greater than
106 2 standard deviations from the background intensity was used for the analysis.

107 Thirty four yeast-specific pathways, known to be involved in growth and stress resistance,
108 described in Wikipathways database [26] were downloaded from
109 http://www.wikipathways.org/index.php/WikiPathways (S2 Table).

110 **Clustering of genetic interactions in yeast**

111 Phenotype data of pair-wise quantitative genetic interactions of 1,712 yeast genes derived
112 from SGA [23] dataset was downloaded from http://drygin.ccbr.utoronto.ca/index.html along
113 with their genetic interaction and *P* values. Genetic interactions of intermediate stringency (*P*
114 < 0.05) were selected for our analysis.

115 In order to cluster the genes into different genetic clusters based on their double-deletion
116 interactions, an adjacency matrix with the interaction scores for the genetic interactions was
117 created from the above data. Pearson correlation among the genes was calculated based on
118 their double deletion interaction scores. Genes with $r^2 > 0.2$ were grouped in the same cluster
119 (S2 Table). The Markov Cluster (MCL) algorithm [27] was applied, using R package

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

function *mcl*, to get highly connected clusters of genes, with an inflation factor of 1.4, as used by the original study Costanzo et al. [23].

Analysis of clusters and pathways in yeast

MCL clusters with 5 or more genes were considered for further analyses (S2 Table). The F-test statistic, BF-test statistic [28], *P* values and variances were computed on the standardized phenotype data derived from the above-standardized Hillenmeyer dataset. Genes in each cluster in each condition were compared to their respective YPD control condition. An identical analysis was carried out for each of the thirty-four pathways downloaded from Wikipathways [26]. Only pathways with 5 or more genes were used (S2 Table). A total of 90 MCL clusters and pathways were considered further and they were associated with several functional categories relevant for growth in yeast, such as known biochemical pathways (TCA, glycolysis), signaling pathways (PKA, MAPK), protein complexes (ribosome, proteasome) and a large number of various other genetic networks (S2 Table).

Hierarchical clustering of cytoplasmic ribosomal proteins in yeast

The genes coding for cytoplasmic ribosomal proteins (n = 68) were considered in 26 environments of the above-standardized Hillenmeyer dataset. A hierarchical clustering of both the environments and the genes were carried out using a Euclidean distance metric using the data, which was standard normalized (mean zero, variance unity) across the ribosomal protein genes.

Enrichment of genetic interactions in Clusters A, B and C

Enrichment of GO categories was carried out in the 121 genes identified to interact with at least 10 out of 65 ribosomal proteins (S3 Table). These 121 interactors interacted either positively or negatively with ribosomal proteins, with a few genes showing both positive and negative interactions (S3 Table). To identify genetic interactors specific to each Cluster, genes interacting with >10% of the ribosomal proteins in each cluster were identified.

Phenotyping single deletions of yeast ribosomal proteins and their deletions with *GCN5* deletion in various environments

Single ribosomal protein and regulatory gene deletions of S288c background (BY4741) were obtained from the haploid genome-wide deletion collection library [29] (GE Healthcare

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Dharmacon Inc.). Additional double-gene deletions were generated using a previously described protocol [30]. SK1 wild type and deletion strains were obtained from Wilkening et al. [31]. Strains were phenotyped in YPD, YPD+Menadione (50µM) and YPD+$CdCl_2$ (10µM). Spot dilutions, ranging from 10^{-3} to 10^{-8} dilutions were incubated at 30ºC and phenotyped at 24, 36 and 48h. The strain and primer list is given in S4 Table.

Sequence analysis of coding and 5'UTR regions in SGRP collection

S. cerevisiae and *S. paradoxus* ribosomal protein and control gene sequences from the SGRP strains [32] were downloaded from http://www.moseslab.csb.utoronto.ca/sgrp/blast_original. Sequence alignments, estimation of the maximum likelihood tree (1,000 permutations) and nucleotide diversity were performed using MEGA 6.06 [33] with default parameters (S5 Table).

For the three gene clusters based on the interaction modules, the nucleotide diversity in the coding and promoter regions of the genes in each cluster was evaluated using Shannon Entropy function as a metric. The sequences of the 5'UTR and coding regions of each gene corresponding to the different strains were aligned separately using the software MUSCLE [34] with default parameters. In the aligned set of sequences for each gene, the mutated sites were identified, and corresponding to each kind of base at the given site, were assigned a value,

$$p_i = \frac{Number\ of\ sequences\ which\ base\ of\ kind\ i\ present\ at\ given\ site}{Total\ number\ of\ sequences\ in\ the\ aligned\ set}$$

Here i = 1, 2, 3, 4 corresponds to A, T, G, C respectively.

The Shannon Entropy at each mutation site was computed using the standard definition:

$$H = -\sum_i p_i \log_2 p_i$$

Variations within a given sequence set can occur in two principal ways: (i) variations in the proportion of bases at any given mutation site, and (ii) variations in the number of sites of mutations. The simplest quantity that accounts for both these is the sum of the H values for all the mutation sites. A higher value for the sum would indicate greater nucleotide diversity in the corresponding 5'UTR or coding region respectively for each gene across the different

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

176 strains. To eliminate any length bias in the comparisons, the sum of the *H* values was
177 normalized by the length of the aligned sequences for each 5'UTR or coding region (S5
178 Table).

179 **Prediction of transcription factor binding sites**

180 The database YEASTRACT [35] was used to download known transcription factor (TF)
181 binding sites of the various ribosomal proteins and to predict potential binding sites on UTR
182 of ribosomal proteins of different SGRP strains. Of the 700 TFs reported in YEASTRACT,
183 216 have been shown to experimentally bind to the promoter region of at least one ribosomal
184 protein present in the three Clusters identified in our study (S6 Table). These 216 TFs were
185 enriched for various signaling pathways and chromatin remodeling complexes (S6 Table),
186 substantiating the enrichment of chromatin remodelers among the positive genetic interactors
187 of ribosomal proteins. Of these 216, the TFs binding exclusively to ribosomal proteins in
188 Cluster A, B and C were identified.

189 **Analysis of human and mouse ENCODE and GTEx data**

190 Tissue specific count (transcripts per million or TPM) data for human and mouse were
191 downloaded from ENCODE (https://www.encodeproject.org/), and mapped to Entrez genes
192 using annotation packages *org.Hs.eg.db* and *org.Mm.eg.db* in R. The genes that were not
193 expressed in any replicate were discarded. Replicates in which an unusually high (as
194 determined from *sfigx_h* and *sfigx_m*) fraction of Entrez genes were not expressed were
195 discarded as well. In the remaining replicates, to reduce relative systematic error among
196 replicates, the median for each gene was normalized to unity in each replicate by dividing the
197 count for the gene by the median count in each replicate array. These median adjusted TPM
198 values were log transformed to obtain the final expression *X* of each gene in each replicate as
199 follows:

200 $$X = \log_2\left(1 + 1023 \times median\,adjusted\,TPM\,value\,of\,the\,gene\,in\,the\,replicate\right)$$

201 This normalization ensured that genes that were not expressed at all were mapped to X = 0,
202 and the median of all genes in a replicate was mapped to $X = 10$. The mean and standard
203 deviation (sd) of expression levels over replicates was computed for each gene-tissue pair to
204 generate a distribution of $x = \log_{10}\left(sd/mean\right)$ for human and mouse data respectively (Fig

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

S1). Based on these distributions cutoffs x_h and x_m were established and the gene-tissue pair was excluded from further analysis if x was greater $x_h \sim$ -0.4 and $x_m \sim$ -0.6 for human and mouse data respectively. For the gene-tissue pairs that passed this check, the expression of a gene in a tissue was defined as the mean over replicates. If a gene had to be excluded in many tissues, then that gene was excluded altogether. Sixty-six ribosomal proteins in 110 tissues in humans (S7 Table), and 42 ribosomal proteins in 18 tissues in mice (S8 Table), passed this filter. The R package *pvclust* was used to perform bootstrapping [36] of hierarchical clustering of ribosomal proteins and tissues in mouse and humans. Similar filtering was performed for GTEx data. GTEx data consists of RNAseq data of 54 tissues from 544 donors amounting to a total of 8,555 samples. Gene-tissue pair with $x_g > 0.1$ were excluded from the data. Seventy-nine ribosomal proteins in 54 tissues passed this filter (S9 Table).

RESULTS

Phenotypic variability of ribosomal proteins in yeast

In all organisms, the ribosome is a ribonucleoprotein complex composed of two subunits each with an RNA core and large number of ribosomal proteins. In eukaryotes, the 60S large subunit consists of 46 proteins, and the 40S small subunit consists of 33 proteins [37]. In yeast that has undergone whole genome duplication [38], most of the ribosomal proteins (paralogs) are duplicated, as a result of which it contains 137 ribosomal proteins, of which 107 are non-essential [22].

Deletion collection in yeast allows testing of phenotypic effect of deletions of non-essential genes in yeast in diverse environments. In order to identify genes which show maximum phenotypic variability across environments, we reanalyzed deletion phenotypes for 4,769 single gene deletions grown in 293 diverse environments using a previously published dataset [22] (S1 Table) for all genes, including ribosomal proteins. The surprising observation was that across all yeast genes, deletions of ribosomal proteins had the highest differential effect on growth in different environments i.e., no effect in some environments and strong effect in others. Among the 191 genes with variance $\sigma^2 > 0.8$ across the 293 environments, components of the ribosome were significantly enriched (21/191, $P < 0.01$, S10 Table, Fig S2). Fourteen out of these 21 genes belonged to the large ribosomal subunit. These 21 genes

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

contain only one paralog of the ribosomal proteins, either A or B, indicating a possible but small differential role of paralogs in responding to environmental heterogeneity. Only ribosomal protein *RPL34* was an exception to this where both paralogs - *RPL34A* and *RPL34B* showed high phenotypic variance.

To test whether this phenotypic variability was a non-specific cellular effect or was specific to the ribosome, we compared phenotype variability in growth for deletions of genes in 90 different pathways and protein complexes across 293 stress conditions versus growth in rich media (YPD). These 90 pathways and complexes were defined using both a biased (Wikipathways) [26] and an unbiased (SGA clustering) [23] approaches (see Methods) and included signaling pathways such as the MAPK and Ras/PKA pathway, protein complexes such as the proteasome and ribosome, cellular processes such as chromatin remodeling complexes and vesicular transport machinery, etc. (S2 Table). Differences in variance of a pathway or a complex between stress and YPD indicate variable roles of its constituents in different conditions. High correlation of constituents of a pathway between stress and YPD would indicate that independent of the essentiality of the pathway; different constituents have similar functions in both conditions. Moreover, a higher variance in YPD compared to stress would indicate that the constituents of the pathway show a more diverse response in YPD but show similar phenotype in stress (Fig 1A). Such a co-ordination of stress specific genes has previously been observed [39] in multiple stresses where the whole pathway is essential to respond to the stress. On the other hand, higher variance of the pathway in stress compared to YPD would indicate that different components of the pathway have differential roles in stress and therefore function in a different manner than in YPD (Fig 1B). Deletions of constituent proteins in 13 pathways showed a significant difference in variance in 3 or more stress environments compared to YPD ($P < 0.01$ by Brown Forsythe test, S2 Table, Fig S2), with the higher variance in YPD in most cases, showing that the pathway was essential in stress. Additionally, constituent of these pathways showed high correlation of phenotype across YPD and stress indicating that the functional hierarchy of the genes was conserved (functional homogeneity), but the phenotypic contribution of the module increased during stress, reducing phenotypic variance. In contrast, for the cytoplasmic ribosomal proteins, there was a significantly higher variance in 28 stress environments compared to YPD (Fig S2, S2 Table), suggesting that ribosomal proteins are differentially used in the stress condition. Unlike other pathways, poor correlation was observed between phenotype of ribosomal proteins across YPD and stress, indicating overall functional heterogeneity in stress and YPD

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(Fig 1B). These results independently show that among different pathways, deletions of genes in the ribosomal pathway has the greatest effect on growth in stress versus rich media, suggesting a unique property of the ribosomal genes that they are the most variable proteins in the cell when comparing diverse environments. A heatmap of growth for single deletions of the 68 ribosomal proteins with consistent replicate data in 25 stress conditions and YPD (Fig 1C) reinforces the above results and shows that a number of these ribosomal proteins have high phenotypic variability, i.e. that they are required for growth in some environments but expendable in others.

We next asked whether these ribosomal proteins have different phenotypic profiles across environments i.e., whether they work independently, or whether they form modules, whose constituents show coordinated regulation across different environments. A clustering analysis of the Pearson correlation of these ribosomal proteins across environments showed functional modularity (Fig 2A) in the form of three distinct clusters (S11 Table). Ribosomal proteins in Cluster 1 were both highly correlated, enriched in large subunit proteins and had high phenotypic diversity across environments. On the other hand, proteins in Cluster 2, although highly correlated and important for growth across most environments, were enriched in small subunit and pre-ribosomal components (important for ribosomal assembly), which explains the constitutive growth defect when these proteins were deleted (S10 Table). Proteins in Cluster 3, however, showed low correlation amongst themselves and were important in different environments. The conclusion that emerges from this analysis is that subsets of ribosomal proteins in Cluster 1 act together in diverse environments, whereas proteins in Cluster 2 act together in most environments. Proteins in Cluster 3 on the other hand, seem to play specialized roles in specific environments.

These results strongly suggest that the yeast ribosomal proteins do not function in a uniform manner when the environment is varied. While this environmental variability of deletions of ribosomal proteins has previously been observed, we have identified a novel underlying modularity among these ribosomal proteins, potentially to optimize yeast growth in different environments. *This is the main finding of the present work that distinguishes it from previous studies.* In summary, whereas a core set of ribosomal proteins are important in all environments, different combinations of a subset of variable ribosomal proteins are functional in different environments to optimize growth.

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

To test whether the deletion phenotype of ribosomal proteins is conserved among yeast strains, we compared growth of ribosomal protein deletions in a soil isolate, SK1, in an oxidative stress (Fig S3). While different ribosomal protein deletions show diverse phenotypic defects, indicating differential use of ribosomal proteins in SK1, the identity of the variable ribosomal proteins was different among the two strains, SK1 and S288c. This indicates that different strains functionally employ ribosomal proteins in different ways, potentially as a result of having adapted in different ways to strain specific selection pressures.

Interactions among ribosomal proteins and genes in cellular pathways

If this observed phenotypic modularity of ribosomal proteins is indeed real, then it should be reflected in their genetic interactions with both upstream and downstream pathways. This crosstalk between ribosomal proteins and genes in other pathways was captured by studying the genome-wide gene-gene interaction SGA dataset [23], to identify positive and negative genetic interactors of the ribosomal proteins. A positive or negative interaction is one where the double deletion is respectively better or worse for growth than the sum of the single deletions. A positive interaction indicates that the genes are in the same pathway, while a negative interaction indicates compensatory pathways [26].

The interactions of a diverse set of genes with 65 ribosomal proteins, 57 of which overlapped with the ribosomal proteins in the deletion phenotype analysis, was analyzed using the SGA dataset [23]. While multiple genes showed double deletion genetic interactions with the ribosomal proteins, our aim was to identify genetic interactors that are common to the ribosome and not to a single ribosomal protein. Hence, only those genes that showed a significant interaction ($P < 0.05$) with at least 10 of the 65 ribosomal proteins were considered for further analysis. This identified a total of 121 genes, 23 ribosomal and 98 non-ribosomal, which had a significantly positive or negative interaction with at least 10 ribosomal proteins (modified F-test, $P < 0.05$, S3 Table). Although the 65 ribosomal proteins had many interactions with other genes, only 12 ribosomal proteins interacted positively among themselves (i.e., were in the same pathway). The rest interacted either positively or negatively with non-ribosomal proteins. The ribosomal proteins form a separate cluster in the yeast gene-gene interaction dataset [23] due to an enrichment of interactions within the complex. Our results show that this enrichment is the result of only 12 ribosomal proteins

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(enriched in small ribosomal subunit); the other ribosomal proteins show high interactions with different cellular pathways.

If the ribosomal proteins indeed function in a coordinated and modular manner to regulate growth across environments, then this modularity should also be evident in their genetic interactions. A clustering analysis of the pairwise correlations of the positive or negative genetic interactions of the 65 ribosomal proteins (Fig 2B) identified three clusters, which had a highly significant overlap with the previous clustering based on phenotypic profiling (90% overlap between Clusters 1 and A and 89% between clusters 2 and B, Fisher Exact test, P = 0.006, S4 Table). Ribosomal proteins in Cluster A interacted mainly with genes involved in mRNA processing, whereas those in Cluster B interacted with other ribosomal proteins. Ribosomal proteins in Cluster C interacted with genes involved in diverse pathways (S10 Table). This strong overlap of corresponding clusters identified independently through phenotype association (Clusters 1, 2, 3) and double deletion analysis (Clusters A, B, C) further reinforces our claim of the modularity of ribosomal proteins wherein subsets of ribosomal proteins function in a coordinated manner and interact in diverse ways amongst themselves and with other non-ribosomal pathways.

We observe that despite high sequence similarity, sometimes differing in only a few bases, ribosomal proteins and their paralogs show diverse genetic interactions. Only 4 out of 15 paralog ribosomal proteins show similar genetic interactions i.e., fall in the same cluster. These are *RPL6A* and *RPL6B*, and *RPL9A* and *RPL9B* in Cluster A and *RPS0A* and *RPS0B*, and *RPS29A* and *RPS29B* in Cluster B (S11 Table). All the remaining paralogs fall into separate clusters or are in Cluster C, indicating that they have differential genetic interactions. Thus, in spite of sequence similarity amongst paralogs, the regulation of these modules in yeast seems to have evolved since the duplication event to create novel functions for these paralogs.

Positive non-ribosomal interactors of the ribosomal proteins were enriched in chromatin regulators and remodelers (S10 Table). Under the prevailing unitary ribosome hypothesis, this association is believed to result from a general control of cellular proliferation by epigenetic regulators [40]. To test this, we experimentally studied the observed positive interaction between the chromatin histone deacetylase *GCN5* and the ribosomal proteins *RPL11B* (Cluster A), *RPL6B* (Cluster B), *RPL38* (Cluster C) and *RPL26B* (Cluster C). *GCN5* regulates transcription of these ribosomal proteins, presumably to control cellular

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

proliferation [40]. If this were true, then deletion of *gcn5*Δ with deletion of any of these ribosomal proteins should have the same growth phenotype as single deletion of *gcn5*Δ. Since ribosomal proteins showed diverse phenotypes in oxidative stresses (Fig 3), we tested the premise of these genetic interactions by phenotyping double deletions of *GCN5* and ribosomal proteins in YPD, and oxidative stresses menadione and $CdCl_2$. While *gcn5*Δ *rpl38*Δ behaved the same as *gcn5*Δ in menadione, indicating that *GCN5* controls cellular proliferation using *RPL38,* double deletion *gcn5*Δ *rps6b*Δ resulted in a 20-fold reduced growth compared to either of the single deletions, indicating parallel or independent roles of both the genes in growth. Furthermore, deletion of the remaining ribosomal proteins (*rpl11b*Δ and *rpl26b*Δ) rescued the growth phenotype of *gcn5*Δ (Fig 3). This observed antagonistic effect suggests that a more likely scenario is that these ribosomal proteins have a direct functional effect on pathways that are lost as a result of *gcn5*Δ deletion. This varied environmental dependence of genetic interactions of ribosomal protein and *GCN5* indicates that some epigenetic regulators and signaling pathways that interact with the ribosome employ the flexibility of the ribosome to mediate phenotypic choice in diverse environments.

Differential evolution of ribosomal proteins

If different subsets of ribosomal proteins are under selection for adaptation to different environmental conditions, there should be a signature of this effect in yeast strains adapted to different environments. The yeast SGRP population [32] consists of 38 *S. cerevisiae* strains isolated from diverse ecological and geographical niches. To detect potential footprints of evolutionary selection, we compared the nucleotide diversity of ribosomal proteins and a control set of housekeeping genes in the SGRP population in their coding and 5'UTR sequences (S5 Table). While the coding sequence nucleotide diversity was similar for both sets, the diversity in the 5'UTR regions within ribosomal proteins was twice that of the control genes ($P < 0.005$, Fig S4A, S5 Table). Furthermore, in the YEASTRACT database [35], this variability altered the predicted transcription factor binding motifs on ribosomal proteins in diverse strains (S7 Table). Thus, while ribosomal proteins among the SGRP strains have similar coding sequences, their promoter regions have been significantly altered, presumably to adapt to different ecological niches.

Coding and 5'UTR regions of ribosomal proteins in the three clusters in Fig 1C were compared using normalized Shannon entropy (S5 Table). All three clusters showed a significant difference in entropy between 5'UTR and coding regions (Fig S4B). Cluster A

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

showed significantly high entropy (diversity) for both 5'UTR and coding regions of ribosomal proteins compared to Clusters B and C (Fig S4B). This differential variability shows that the ribosomal proteins in the clusters are evolving at different rates across the SGRP population.

Using the YEASTRACT database, we find that whereas most transcription factors (TFs) bind to ribosomal proteins in all three clusters, some are cluster specific. Transcription Factors regulating ribosomal proteins in Cluster A are enriched in the histone deacetylase complex while those that bind to ribosomal proteins in Cluster C are enriched in the HIR (Histone Regulatory) complex (Fig S5, S6 Table). This could be a possible explanation for the different rates of evolution of the 5'UTR regions of the proteins in these three clusters.

Modular ribosomal proteins in higher eukaryotes

Our above results establish modularity in both phenotype and genetic interactions of ribosomal proteins in yeast. It can be argued that this is merely a unique feature of the yeast ribosome, presumably because of the whole genome duplication event, which might have allowed differential adaptation of duplicated ribosomal proteins. To understand whether the ribosomal modularity observed in yeast extends to higher eukaryotes, we investigated expression levels of ribosomal proteins in mice and humans, which have a single copy of most ribosomal proteins.

In complex eukaryotes, the analog of adaptation of unicellular organisms like yeast to different environments is adaptation to different cellular and tissue microenvironments. We therefore expect that if our thesis of the modularity of ribosomal proteins is valid beyond single celled eukaryotes, ribosomal proteins should be differentially used across cell types and tissues in mice and humans. In multicellular systems like humans and mice, ribosomal proteins are present in a single copy, whose deletion results in both cellular and organismal lethality. Our hypothesis for complex eukaryotes would then be that ribosomal proteins are expressed at significantly different levels in different tissue microenvironments in mice and humans. We tested this hypothesis by comparing the expression levels of ribosomal proteins in diverse cell types and tissues in human and mouse samples using RNASeq data for mRNA transcript levels from the ENCODE and the GTEx projects.

A total of 66 ribosomal proteins with consistent transcript expression levels across replicates in the ENCODE data were identified in 110 cell types and tissues in humans (see Methods).

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

As previously observed [41], we found that while the majority of ribosomal protein transcripts are highly expressed across diverse tissues and cell types, a few showed low expression levels throughout. It has been shown that ribosomal proteins can show differential expression levels based on the proliferation or turnover rate of the cell type. To normalize such global differences, each ribosomal protein within each tissue was assigned a rank based on its expression level (rank 1 for lowest expression and rank 66 for highest expression, S7 Table). Despite being a part of the ribosome, it is known that not all ribosomal proteins are equally expressed in a given tissue. In complex eukaryotes, just as in yeast, some of the ribosomal proteins are involved in ribosomal assembly. However, since this is true for all tissues and cell types, their rank normalized expression levels should be consistent across tissues.

Hierarchical clustering of all ribosomal proteins expression ranks across all tissues resulted in a single highly correlated cluster (Fig S6). However, our results from yeast show that while some ribosomal proteins are essential and behave similarly across environments, others show high variability. To identify these highly variable ribosomal proteins, the 66 ribosomal proteins were filtered based on their ranks. In each tissue, the expression level ranks of the proteins were stratified into 4 classes: class I (rank 1-17), II (17-34), III (35-51) and IV (51-66) (see Methods). This showed that across tissues, 46 out of 66 ribosomal proteins were classified into the same or adjacent classes, while the remaining 20 were classified into 3 classes for 11 or more tissues types per class (S7 Table). These 20 were termed as variable ribosomal proteins and analyzed further. Note that had the rank assignments merely amplified small differences in expression levels (noise) for a given protein across tissues, such stratification would not have been observed. Instead, we would have seen a random assignment of ranks across tissues, which is not what was observed.

The 20 variable ribosomal proteins spanned mostly classes II, III and IV i.e., their transcripts were both highly expressed and highly variable across tissues, thereby eliminating technical noise as the cause of the observed variability in ranks (S7 Table). Hierarchical clustering of the ranks of these 20 proteins across cell types and tissues showed that these proteins assort into distinct groups (Fig 4A). An identical clustering can also be observed in a heat map of Pearson rank correlations (Fig 4B, correlation $P < 0.05$). In the hierarchical clustering, distinct sets of ribosomal proteins were associated with two discrete clusters of epithelial cells, a cluster of the nervous tissue (tissue from different sections of the brain and spinal

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

cord) and a cluster of human cell lines (Fig 4A). We note that cell lines cluster separately, indicating that similar to modification of their signaling pathways [42], expression patterns of ribosomal proteins are also rewired in these cell lines compared to other human cells and tissues. To investigate this tissue specific ribosomal modularity further, we separately analyzed the 20 ribosomal proteins in the clusters associated with epithelial and nervous tissues. This again showed that the nervous tissue cluster is quite distinct in its use of variable ribosomal proteins compared to epithelial cells (Fig 5A, 5B). Even though the nervous tissues are known to have a reduced expression of ribosomal proteins compared to other more proliferating tissues, our results show that the ribosomal protein module in the nervous tissue is distinct from that in the epithelial cells. These results show that, analogous to yeast, ribosomes in humans show expression modularity across tissues.

While the different cell types separated into different clusters based on the expression of variable ribosomal proteins in ENCODE data, they did not show a significant tissue bias except for nervous tissue. For example, the two clusters of epithelial cells did not segregate on the basis of organ of origin, suggesting that modularity of ribosomal proteins plays a role at the resolution of cell types instead of whole organs. To validate our result of modularity of ribosomal proteins in humans and to understand whether this modularity was a property of cell types or is also visible in bulk tissue, we compared the expression patterns of ribosomal proteins using the GTEx dataset. The GTEx data consists of RNAseq analysis of 54 different tissues (S9 Table). We extracted expression values of ribosomal proteins from these samples and performed the same analysis as for the ENCODE data (see Methods). Seventy nine ribosomal proteins passed our filtering criteria and were classified into ranks ranging from 1 for the least expression and 79 for the highest expression in each tissue. These were further stratified into 4 classes: class I (rank 1-20), II (21-40), III (41-60) and IV (61-79). We found that ribosomal proteins from the GTEx data showed less variability in classes across tissues compared to the ENCODE data (S9 Table). Consequently, ribosomal proteins that fell into two or more classes with at least 10 tissues per class were identified as variable ribosomal proteins. A total of 18 ribosomal proteins were identified to be variable, of which 7 were the same as those identified in the ENCODE data, showing a significant overlap (Fisher's Exact test, $P < 0.1$) between variable ribosomal proteins identified using ENCODE and GTEx data (Fig 6A). The variable ribosomal proteins in the GTEx data separated into two modules (Pearson correlation, $P < 0.01$, Fig 6A). Similar to the ENCODE data, the nervous tissues (brain and spinal cord) formed a separate cluster, validating our previous observation that a

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

different module of ribosomal proteins is used in the nervous system compared to other tissues (Fig 6B). However, bulk tissues did not cluster separately on the basis of variable ribosomal proteins (Fig 6B). The lack of modularity of ribosomal proteins at the level of tissues indicates that cell specific differences in ribosomal protein expression levels are lost when dealing with data from bulk tissue, because cell specific identity is lost in the GTEx data.

A similar analysis was carried out for expression levels of ribosomal proteins in various tissues in mice from the ENCODE data. We note that the signal to noise ratio in the mouse data in ENCODE was significantly higher than in the human data. Consequently, only 42 ribosomal proteins in 18 different tissues passed our filtering criteria and were stratified into four classes (S8 Table for details of classes). As in the human data, while the majority of mice ribosomal proteins showed high expression and invariant classification, 14 of these 42 ribosomal proteins were as variable across tissues (spanning 3 classes in more than 2 tissues). As in ENCODE and GTEx data from humans, mouse brain tissues also formed a unique cluster, indicating that nervous tissue in general uses a distinct module of ribosomal proteins compared to other tissues (Fig S7A, 7B).

We identified *RPL38* as a variable protein in ENCODE and GTEx data in humans as well as ENCODE data in mice (Fig 5, 6, S7B). *RPL38* is the most extensively studied ribosomal protein associated with ribosomal heterogeneity. This heterogeneity is due to its specialized translation of only the *hox* mRNA, without affecting translation of other mRNA [19]. Identification of *RPL38* as a variable ribosomal protein in our study serves as an independent validation of our analyses. While different sets of ribosomal proteins were found to be variable in mice and humans, 6 out of 18 ribosomal proteins variable in mice were also variable in humans (4 in ENCODE data and 5 in GTEx data, Fig S7B). Furthermore, 4 of these 6 conserved variable ribosomal proteins fell in one cluster in mouse indicating a partial conservation of variability of ribosomal proteins across the two species (Fig S7B).

DISCUSSION

Our study provides several arguments and multiple evidences for the existence of modularity of ribosomal proteins across eukaryota, presumably to facilitate optimized translation efficiency in different environments. We show that, at least in yeast, we see evidence that the

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

5'UTRs of ribosomal proteins that form the modules seem to be under selection pressure, which suggests that they play a role in evolutionary adaptation. We interpret our results as evidence for a hitherto unrecognized ribosomal code, wherein specific ribosomal proteins are used in an environment specific manner in yeast and in cell and tissue specific ways in mice and humans. *The existence of such a dynamic modularity of ribosomal proteins is the main finding of this paper.* The mechanisms that regulate these modules remain to be elucidated and are outside the scope of this paper.

Our study also uncovered some general, conserved properties of ribosomal proteins. We find that a subset of variable ribosomal proteins contribute to the plasticity of the ribosome by functioning independently or in concert across different environments by forming *modules*, defined as sets of proteins functioning in a coordinated manner. This modularity indicates that they have been optimized over the course of evolution based on the need for functional adaptation. We note that only a subset of ribosomal proteins vary among cell types and tissues, with the core ribosome remaining unaffected. Hence cell type specific structural changes resulting from such variation may be difficult to detect.

Our findings would argue that, in spite of high sequence conservation [43], the inability of human ribosomal genes to substitute for yeast ribosomal genes [44] is probably because of species specific functioning of ribosomal modules. This, along with differential expression variability of ribosomal proteins in mice and humans, indicates that each species optimizes the composition of its ribosome to adapt to species specific selection pressures, not by substantially altering the sequence of the ribosomal proteins but by regulating their expression in an environment dependent manner using mechanisms yet to be discovered.

Our results from two independent expression datasets (ENCODE and GTEx) show that the nervous tissues use a unique ribosomal code compared to the rest of the tissues in both mice and humans. While an overall reduced expression of ribosomal proteins in the brain has been observed previously, it has been attributed to the reduced proliferation of nervous cell types. Here, we show that along with a reduced expression, a unique composition of ribosomal proteins is utilized by nervous tissue. These may play a role in the fundamental physiological differences observed between the brain and the rest of the body.

In mice and humans, recently evolved paralogs *RPL27L*, *RPL22L1*, *RPL7L1*, and *RPL39L*, showed poor but highly tissue specific expression compared to core ribosomal proteins. This

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

552 suggests that there is an ongoing process of adaptation driving modular ribosomes, with
553 recent paralog proteins still evolving in response to selection pressures on them and on other,
554 more ancient ribosomal proteins and pathways.

555 Our results show that ribosome modularity is a dynamic, evolving process which seems to be
556 involved in the evolution of species specific ribosomal proteins, the diversification of their
557 sequences and functions and the creation of novel, species specific ribosomal proteins [45,46]
558 leading to diverse phenotypic adaptations.

559

560 **REFERENCES**

561 1. Keene JD. RNA regulons: coordination of post-transcriptional events. Nat Rev Genet.
562 2007;8: 533–543. doi:10.1038/nrg2111

563 2. Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from
564 molecular and genomics approaches. Nat Rev Mol Cell Biol. 2009;10: 741–754.
565 doi:10.1038/nrm2777

566 3. Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-
567 transcriptional gene regulation. FEBS Lett. 2008;582: 1977–1986.
568 doi:10.1016/j.febslet.2008.03.004

569 4. Araujo PR, Yoon K, Ko D, Smith AD, Qiao M, Suresh U, et al. Before it gets started:
570 regulating translation at the 5' UTR. Comp Funct Genomics. 2012;2012: 475731.
571 doi:10.1155/2012/475731

572 5. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. Why highly expressed
573 proteins evolve slowly. Proc Natl Acad Sci U S A. 2005;102: 14338–14343.
574 doi:10.1073/pnas.0504070102

575 6. Ramakrishnan V. The ribosome emerges from a black box. Cell. 2014;159: 979–984.
576 doi:10.1016/j.cell.2014.10.052

577 7. Kondrashov N, Pusic A, Stumpf CR, Shimizu K, Hsieh AC, Xue S, et al. Ribosome-
578 mediated specificity in *hox* mRNA translation and vertebrate tissue patterning. Cell.
579 2011;145: 383–397. doi:10.1016/j.cell.2011.03.028

580 8. Mccormick MA, Delaney JR, Tsuchiya M, Tsuchiyama S, Shemorry A, Sim S, et al. A
581 comprehensive analysis of replicative lifespan in 4,698 single-gene deletion strains
582 uncovers conserved mechanisms of aging. Cell Metab. 2015.
583 doi:10.1016/j.cmet.2015.09.008

584 9. Holt CE, Schuman EM. The Central Dogma Decentralized: New perspectives on RNA
585 function and local translation in neurons. Neuron. 2013;80: 648–657.
586 doi:10.1016/j.neuron.2013.10.036

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

587 10. las Heras-Rubio de A, Perucho L, Paciucci R, Vilardell J, LLeonart ME. Ribosomal
588 proteins as novel players in tumorigenesis. Cancer Metastasis Rev. 2013;33: 115–41.
589 doi:10.1007/s10555-013-9460-6

590 11. Diesch J, Hannan RD, Sanij E. Perturbations at the ribosomal genes loci are at the
591 centre of cellular dysfunction and human disease. Cell Biosci. 2014;4: 1–11.
592 doi:10.1186/2045-3701-4-43

593 12. McCann KL, Baserga SJ. Genetics. Mysterious ribosomopathies. Science. 2013;341:
594 849–850. doi:10.1126/science.1244156

595 13. Teng T, Thomas G, Mercer CA. Growth control and ribosomopathies. Curr Opin
596 Genet Dev. 2013;23: 63–71. doi:10.1016/j.gde.2013.02.001

597 14. Lindström MS. Emerging functions of ribosomal proteins in gene-specific
598 transcription and translation. Biochem Biophys Res Commun. 2009;379: 167–170.
599 doi:10.1016/j.bbrc.2008.12.083

600 15. Warner JR, McIntosh KB. How common are extraribosomal functions of ribosomal
601 proteins? Mol Cell. 2009;34: 3–11. doi:10.1016/j.molcel.2009.03.006

602 16. Xue S, Barna M. Specialized ribosomes: a new frontier in gene regulation and
603 organismal biology. Nat Rev Mol Cell Biol. 2012;13: 355–369. doi:10.1038/nrm3359

604 17. Lee AS-Y, Burdeinick-Kerr R, Whelan SPJ. A ribosome-specialized translation
605 initiation pathway is required for cap-dependent translation of vesicular stomatitis
606 virus mRNAs. Proc Natl Acad Sci USA. 2013;110: 324–329.
607 doi:10.1073/pnas.1216454109

608 18. Sauert M, Temmel H, Moll I. Heterogeneity of the translational machinery: Variations
609 on a common theme. Biochimie. 2015;114: 39–47. doi:10.1016/j.biochi.2014.12.011

610 19. Shi Z, Barna M. Translating the genome in time and space: Specialized ribosomes,
611 RNA regulons, and RNA-binding proteins. Annu Rev Cell Dev Biol. 2015;31: 31–54.
612 doi:10.1146/annurev-cellbio-100814-125346

613 20. Dinman JD. Pathways to specialized ribosomes: The Brussels Lecture. J Mol Biol.
614 2016;428: 2186-94. doi:10.1016/j.jmb.2015.12.021

615 21. Slavov N, Semrau S, Airoldi E, Budnik B, van Oudenaarden A. Differential
616 stoichiometry among core ribosomal proteins. Cell Rep. 2015;13: 865–873.
617 doi:10.1016/j.celrep.2015.09.056

618 22. Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, et al. The
619 chemical genomic portrait of yeast: Uncovering a phenotype for all genes. Science.
620 2008;320: 362–365. doi:10.1126/science.1150021

621 23. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, et al. The
622 genetic landscape of a cell. Science. 2010;327: 425–431. doi:10.1126/science.1180823

623 24. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the
624 human genome. Nature. 2012;489: 57–74. doi:10.1038/nature11247

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

625 25. GTEx Consortium, Getz G, Kellis M, Dermitzakis ET. Human genomics. The
626 Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in
627 humans. Science. 2015;348: 648–660. doi:10.1126/science.1262110

628 26. Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, et al.
629 WikiPathways: building research communities on biological pathways. Nucl Acids
630 Res. 2011;40: D1301–D1307. doi:10.1093/nar/gkr1074

631 27. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale
632 detection of protein families. Nucl Acids Res. 2002;30: 1575–1584.

633 28. Lee C-R, Anderson JT, Mitchell-Olds T. Unifying genetic canalization, genetic
634 constraint, and genotype-by-environment interaction: QTL by genomic background by
635 environment interaction of flowering time in *Boechera stricta*. PLoS Genet. 2014;10:
636 e1004727. doi:10.1371/journal.pgen.1004727.s015

637 29. Winzeler E, Shoemaker D, Astromoff A, Liang H, Anderson K, Andre B, et al.
638 Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel
639 analysis. Science. 1999;285: 901–906. doi:10.1126/science.285.5429.901

640 30. Goldstein A, Pan X, McCusker J. Heterologous *URA3MX* cassettes for gene
641 replacement in *Saccharomyces cerevisiae*. Yeast. 1999;15: 507–511.
642 doi:10.1002/(SICI)1097-0061(199904)15:6<507::AID-YEA369>3.0.CO;2-P

643 31. Wilkening S, Lin G, Fritsch ES, Tekkedil MM, Anders S, Kuehn R, et al. An
644 evaluation of high-throughput approaches to QTL mapping in *Saccharomyces*
645 *cerevisiae*. Genetics. 2014;196: 853–865. doi:10.1534/genetics.113.160291

646 32. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, et al. Population
647 genomics of domestic and wild yeasts. Nature. 2009;458: 337–341.
648 doi:10.1038/nature07743

649 33. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular
650 Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol. 2013;30: 2725–2729.
651 doi:10.1093/molbev/mst197

652 34. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high
653 throughput. Nucl Acids Res. 2004;32: 1792–1797. doi:10.1093/nar/gkh340

654 35. Teixeira MC, Mira NP, Sá-Correia I. A genome-wide perspective on the response and
655 tolerance to food-relevant stresses in *Saccharomyces cerevisiae*. Curr Opin Biotechnol.
656 2011;22: 150–156. doi:10.1016/j.copbio.2010.10.011

657 36. Suzuki R, Shimodaira H. Pvclust: an R package for assessing the uncertainty in
658 hierarchical clustering. Bioinformatics. 2006;22: 1540–1542.
659 doi:10.1093/bioinformatics/btl117

660 37. Jenner L, Melnikov S, de Loubresse NG, Ben-Shem A, Iskakova M, Urzhumtsev A, et
661 al. Crystal structure of the 80S yeast ribosome. Curr Opin Struct Biol. 2012;22: 759–
662 767. doi:10.1016/j.sbi.2012.07.013

663 38. Komili S, Farny NG, Roth FP, Silver PA. Functional specificity among ribosomal

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

proteins regulates gene expression. Cell. 2007;131: 557–571. doi:10.1016/j.cell.2007.08.037

39. Gasch AP, Huang M, Metzner S, Botstein D, Elledge SJ, Brown PO. Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. Mol Biol Cell. 2001;12: 2987–3003. doi:10.1091/mbc.12.10.2987

40. Kurdistani SK, Grunstein M. Histone acetylation and deacetylation in yeast. Nat Rev Mol Cell Biol. 2003;4: 276–284. doi:10.1038/nrm1075

41. Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA. Selection for short introns in highly expressed genes. Nat Genet. 2002;31: 415–418. doi:10.1038/ng940

42. Pan C, Kumar C, Bohl S, Klingmueller U, Mann M. Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions. Mol Cell Proteomics. 2009;8: 443–450. doi:10.1074/mcp.M800258-MCP200

43. Yoshihama M, Uechi T, Asakawa S, Kawasaki K, Kato S, Higa S, et al. The human ribosomal protein genes: sequencing and comparative analysis of 73 genes. Genome Res. 2002;12: 379–390. doi:10.1101/gr.214202

44. Kachroo AH, Laurent JM, Yellman CM, Meyer AG, Wilke CO, Marcotte EM. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. Science. 2015;348: 921–925. doi:10.1126/science.aaa0769

45. Vinckenbosch N, Dupanloup I, Kaessmann H. Evolutionary fate of retroposed gene copies in the human genome. Proc Natl Acad Sci U S A. 2006;103: 3220–3225. doi:10.1073/pnas.0511307103

46. Wong QW-L, Li J, Ng SR, Lim SG, Yang H, Vardy LA. RPL39L is an example of a recently evolved ribosomal protein paralog that shows highly specific tissue expression patterns and is upregulated in ESCs and HCC tumors. RNA Biol. 2014;11: 33–41. doi:10.4161/rna.27427

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

FIGURE LEGENDS

Fig 1: Phenotypic variability of yeast ribosomal proteins

(A) Growth of deletions of DNA repair cluster genes (black dots) in rich medium YPD (x-axis) versus a DNA damaging agent, cisplatin (y-axis). (B) Growth of deletions of ribosomal proteins (black dots) in rich medium, YPD (x-axis) versus an oxidizing stress, Cadmium chloride ($CdCl_2$) (y-axis). (C) Hierarchical clustering heat map of normalized growth of yeast strains for 68 single deletions of ribosomal proteins in 26 environments. The red arrow indicates YPD (rich medium).

Fig 2: Differential modules of ribosomal proteins in yeast

Hierarchical clustering of Pearson correlations among ribosomal proteins for (B) normalized growth from single deletion of ribosomal proteins in 26 environments, (C) double deletion interactions with 121 genes. Ribosomal proteins within each cluster are significantly correlated with $P < 0.05$ for each pair.

Fig 3: Genetic interactions of ribosomal proteins with *GCN5*

10-fold spot dilutions series (starting with 10^8 cells/ml) of wild type, single and double deletions of *GCN5* with *RPS6B*, *RPL38*, *RPS26B* and *RPL11B* phenotyped in YPD (rich medium) and an oxidative stress, Cadmium chloride ($CdCl_2$, 10μM).

Fig 4: Clustering of human cell types and tissues based on rank order expression of variable ribosomal proteins in ENCODE data

(A) A hierarchical clustered tree of 110 human cell types and tissues (1,000 bootstraps) based on rank order expression of 20 ribosomal proteins in ENCODE data. (B) Pearson correlation ($P < 0.05$) heatmap based on rank order expression of 20 variable ribosomal proteins results in distinct clusters and sub-clusters.

Fig 5: Modules within the 20 variable ribosomal proteins in human tissues

(A) The modular organization derived from hierarchical clustering of ribosomal proteins rank order expression in 110 cell types and tissues (1,000 bootstraps). (B) Heatmap showing different modules of 20 variable ribosomal proteins active in epithelial cells and the nervous tissue.

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Fig 6: Clustering of human tissues and variable ribosomal proteins in GTEx data

(A) Pearson correlation ($P < 0.01$) heatmap of 18 variable ribosomal proteins based on their rank order expression in 54 different tissues in GTEx data. Ribosomal proteins marked in red are also variable in ENCODE data. (B) Pearson correlation ($P < 0.01$) heatmap of 54 different tissues based on rank order expression of 18 variable ribosomal proteins. Nervous tissue (brain and spinal cord) form a distinct cluster and are marked red.

SUPPLEMENTARY INFORMATION

Fig S1: Distribution of signal to noise in human and mouse data from ENCODE.

ENCODE data was normalized by median subtraction per array and then log transformed (see Methods). The mean and standard deviation (sd) over replicates was computed to obtain $x = \log_{10}\left(sd/mean\right)$. The distribution of x for human (A) and mouse (B) ENCODE data and (C) human GTEx data was used to determine the cutoffs x_h, x_m and x_g for reliability of the ENCODE data for human, mouse and GTEx data for humans, respectively. Gene-tissue pairs for which $x_h > -0.4$, $x_m > -0.6$ and $x_g > 0.1$ were excluded from the analysis.

Fig S2: Phenotypic variation of yeast ribosomal proteins.

(A) Distribution of variance of normalized growth of all non-essential genes in yeast (4,769) in 293 different environments. The genes are on the x-axis, arranged in increasing order of the variance of normalized growth in 293 environments (y-axis) due to their deletion. The 191 genes to the right of the red line have variance greater than 0.8. (B) Stacked histogram showing the number of cases when the variance of growth from deletion of genes in various pathways was greater or smaller in stress compared to YPD. Gene deletions in the cytoplasmic ribosome and mitochondrial tRNA synthesis pathways had the highest variance in stress compared to YPD. However in mitochondrial tRNA synthesis pathway genes, the variance is greater or smaller in stress compared to YPD for equal numbers of genes. Only in the ribosomal pathway is the variance in stress conditions greater than YPD for all genes. (C) This figure shows the same data as in B but with the genes stratified into clusters based on their double deletion interactions (see Methods). The ribosomal cluster has the highest variance in stress compared to YPD.

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Fig S3: Phenotype of ribosomal protein deletions in SK1 strain

Ten-fold spot dilutions series (starting with 10^8 cells/ml) of wild type and ribosomal protein deletion strains of SK1 background phenotyped in rich medium YPD and an oxidative stress, Cadmium chloride ($CdCl_2$ 500 μM)

Fig S4: Nucleotide diversity of ribosomal proteins across the SGRP strains:

(A) Nucleotide diversity of coding and promoter sequences of ribosomal proteins and a control set of genes using Tukey's multiple comparisons' test ($P < 0.05$). The bars with the same letter code do not differ significantly.

(B) Normalized Shannon Entropy of coding region and 5'UTRs of Cluster A, B and C from Fig 2B. Bars with the same letter code do not differ significantly (Tukey's multiple comparisons' test, $P < 0.05$). The figure shows that: (i) The 5'UTR regions of the ribosomal protein sequences are most variable compared to their coding region as well as the 5'UTR and coding regions of the control set of genes; (ii) The 5'UTRs of all the ribosomal proteins in the three clusters are significantly more variable ($P < 0.01$) than their coding regions; (iii). Proteins in Cluster A have significantly more variable coding regions than clusters B and C ($P < 0.05$ and $P < 0.01$ respectively); (iv). Proteins in Cluster A have significantly more variable 5'UTR than Clusters B and C ($P < 0.05$ and $P < 0.01$ respectively).

Fig S5: Networks of transcription factors that bind uniquely to ribosomal proteins in Cluster A, B and C (see Fig 2B).

These network clusters were identified using the STRING database (http://string-db.org). The thickness of blue lines connecting two transcription factors indicates the strength of experimental evidence for their interaction. Gene enrichment ($P < 0.001$) of transcription factors in Cluster A is for Histone Deacetylase Complex and in Cluster C for the HIR Complex.

Fig S6: Pearson correlation heatmap ($P < 0.05$) of 110 human cell types and tissues based on rank order expression of 66 ribosomal proteins in ENCODE data.

Fig S7: Variable ribosomal proteins across tissues in mouse

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

779 (A) Hierarchical clustering of 18 tissues in mice based on expression rank orders of 14
780 variable ribosomal proteins (1,000 bootstraps).

781 (B) Hierarchical clustering of the 14 variable ribosomal proteins based on their expression
782 rank orders in 18 tissues in mice (1,000 bootstraps). The ribosomal proteins form three
783 distinct clusters indicated in different colors. The red arrows indicate ribosomal proteins that
784 are also variable in human ENCODE and GTEx data.

785

786 **S1 Table**: Yeast deletion collection phenotyped in 293 environments.

787 **S2 Table**: Variance analysis of various pathways and genetic

788 **S3 Table**: Double deletion genetic interactions of the 65 ribosomal proteins with 121
789 ribosomal and non-ribosomal proteins.

790 **S4 Table**: Strains used in this study and Primer sequences

791 **S5 Table**: Nucleotide diversity and Shannon Entropy comparisons of ribosomal proteins in
792 Clusters A, B and C.

793 **S6 Table**: Transcription Factors binding to ribosomal proteins in Clusters A, B and C
794 downloaded from YEASTRACT

795 **S7 Table**: Expression ranks and classes of 66 ribosomal proteins in 110 human cell types and
796 tissues downloaded from ENCODE

797 **S8 Table**: Expression ranks and classes of 42 ribosomal proteins in 18 mouse tissues
798 downloaded from ENCODE

799 **S9 Table**: Expression ranks and classes of 79 ribosomal proteins in 54 human tissues
800 downloaded from GTEx portal.

801 **S10 Table**: Gene Ontology Enrichment for various groups and clusters

802 **S11 Table**: ribosomal proteins in Clusters 1, 2, 3 and Clusters A, B, C

803

Figure 1

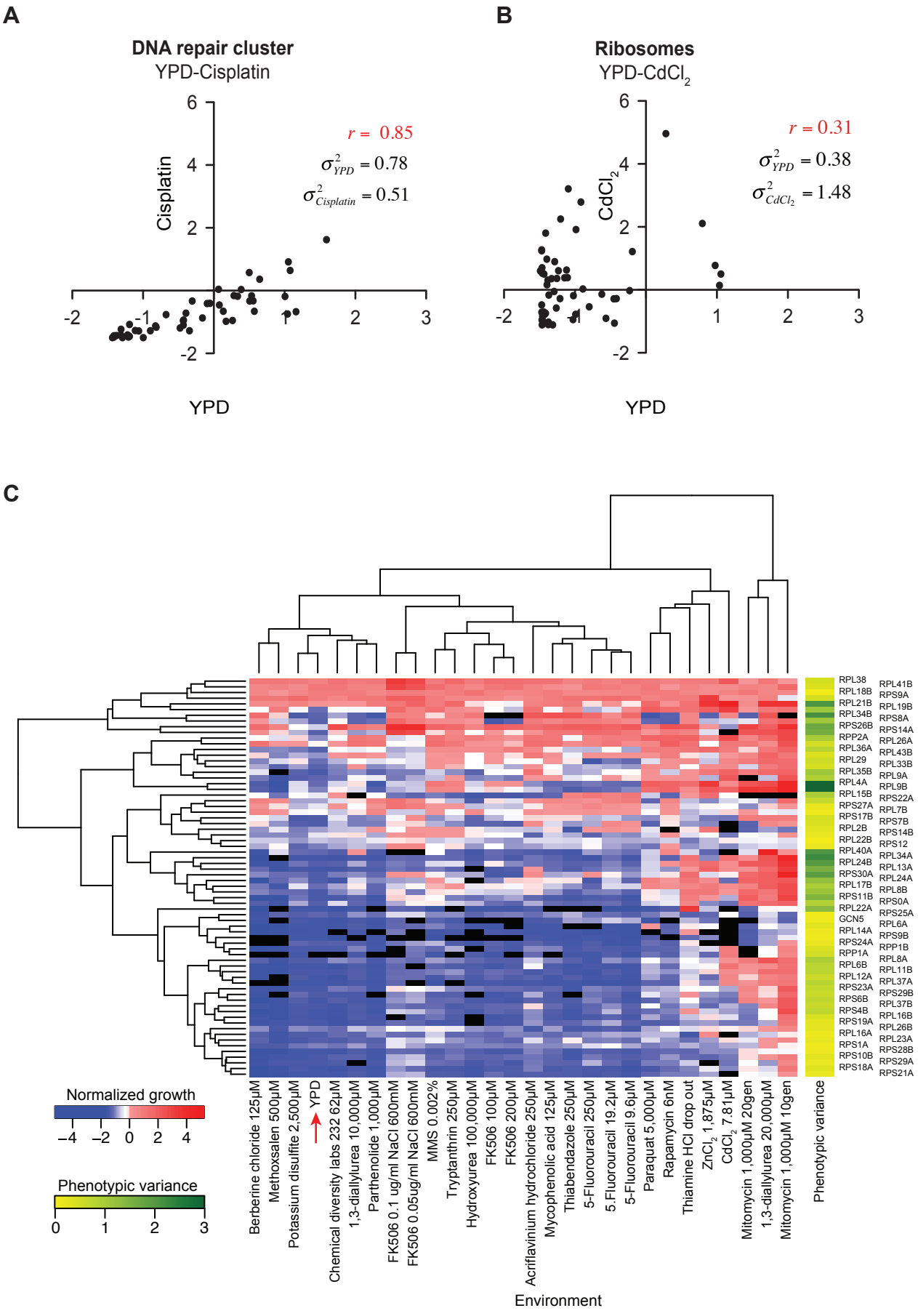

A
DNA repair cluster
YPD-Cisplatin
r = 0.85
$\sigma^2_{YPD} = 0.78$
$\sigma^2_{Cisplatin} = 0.51$
Cisplatin
YPD
B
Ribosomes
YPD-CdCl2
r = 0.31
$\sigma^2_{YPD} = 0.38$
$\sigma^2_{CdCl_2} = 1.48$
CdCl2
YPD
C
Normalized growth
Phenotypic variance
Berberine chloride 125µM
Methoxsalen 500µM
Potassium disulfite 2,500µM
YPD
Chemical diversity labs 232 62µM
1,3-diallylurea 10,000µM
Parthenolide 1,000µM
FK506 0.1 ug/ml NaCl 600mM
FK506 0.05ug/ml NaCl 600mM
MMS 0.002%
Tryptanthrin 250µM
Hydroxyurea 100,000µM
FK506 100µM
FK506 200µM
Acriflavinium hydrochloride 250µM
Mycophenolic acid 125µM
Thiabendazole 250µM
5-Fluorouracil 250µM
5.Fluorouracil 19.2µM
5-Fluorouracil 9.6µM
Paraquat 5,000µM
Rapamycin 6nM
Thiamine HCl drop out
ZnCl2 1,875µM
CdCl2 7.81µM
Mitomycin 1,000µM 20gen
1,3-diallylurea 20,000µM
Mitomycin 1,000µM 10gen
Phenotypic variance
Environment

Figure 2

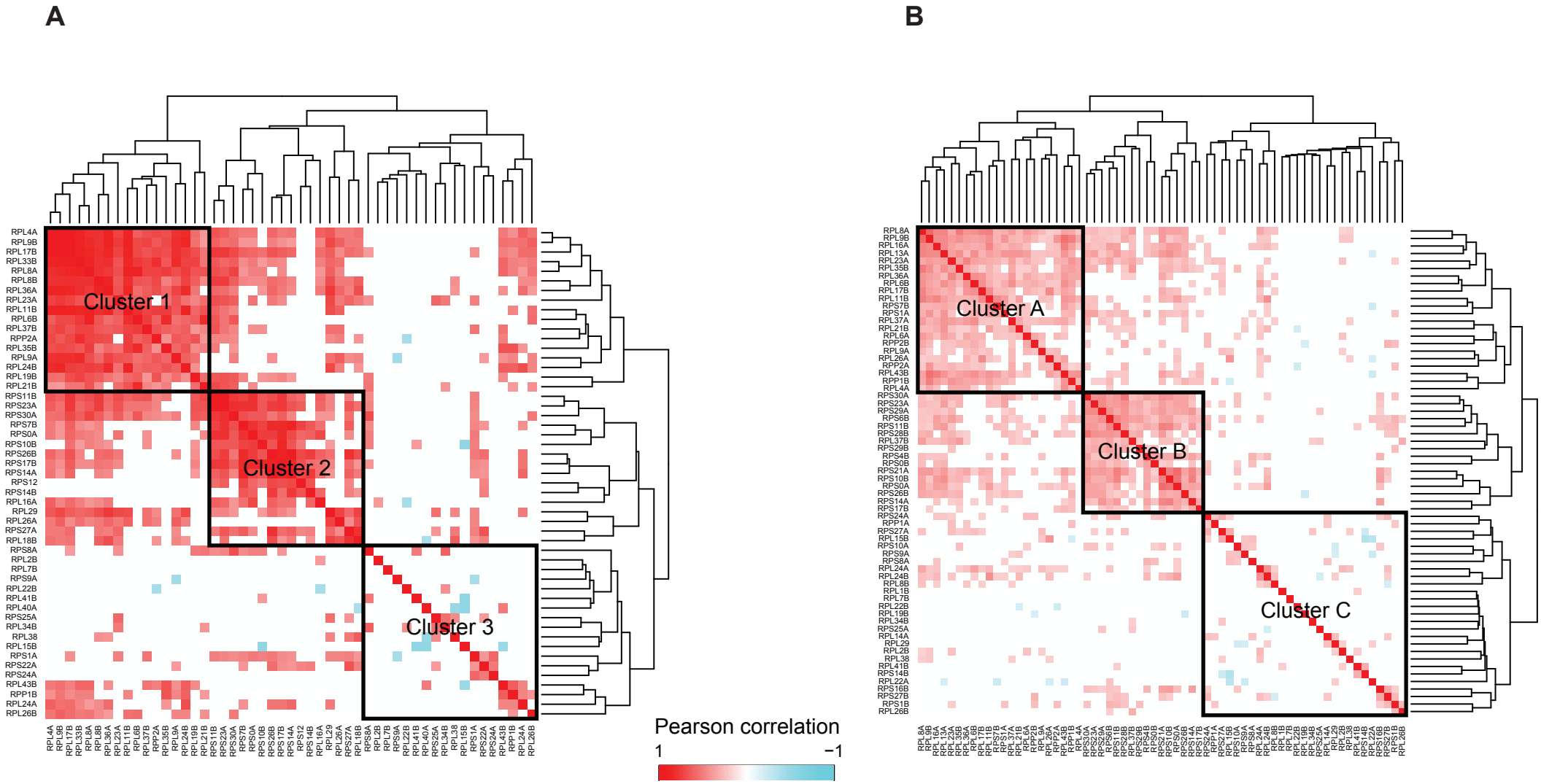
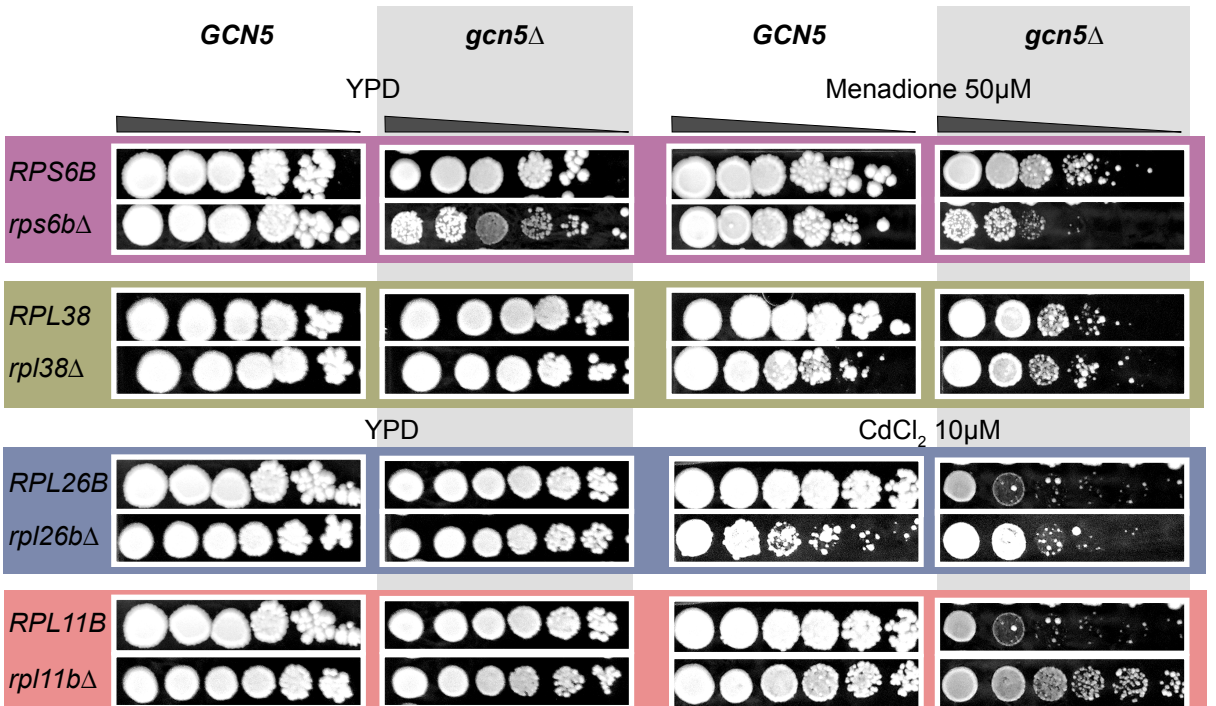
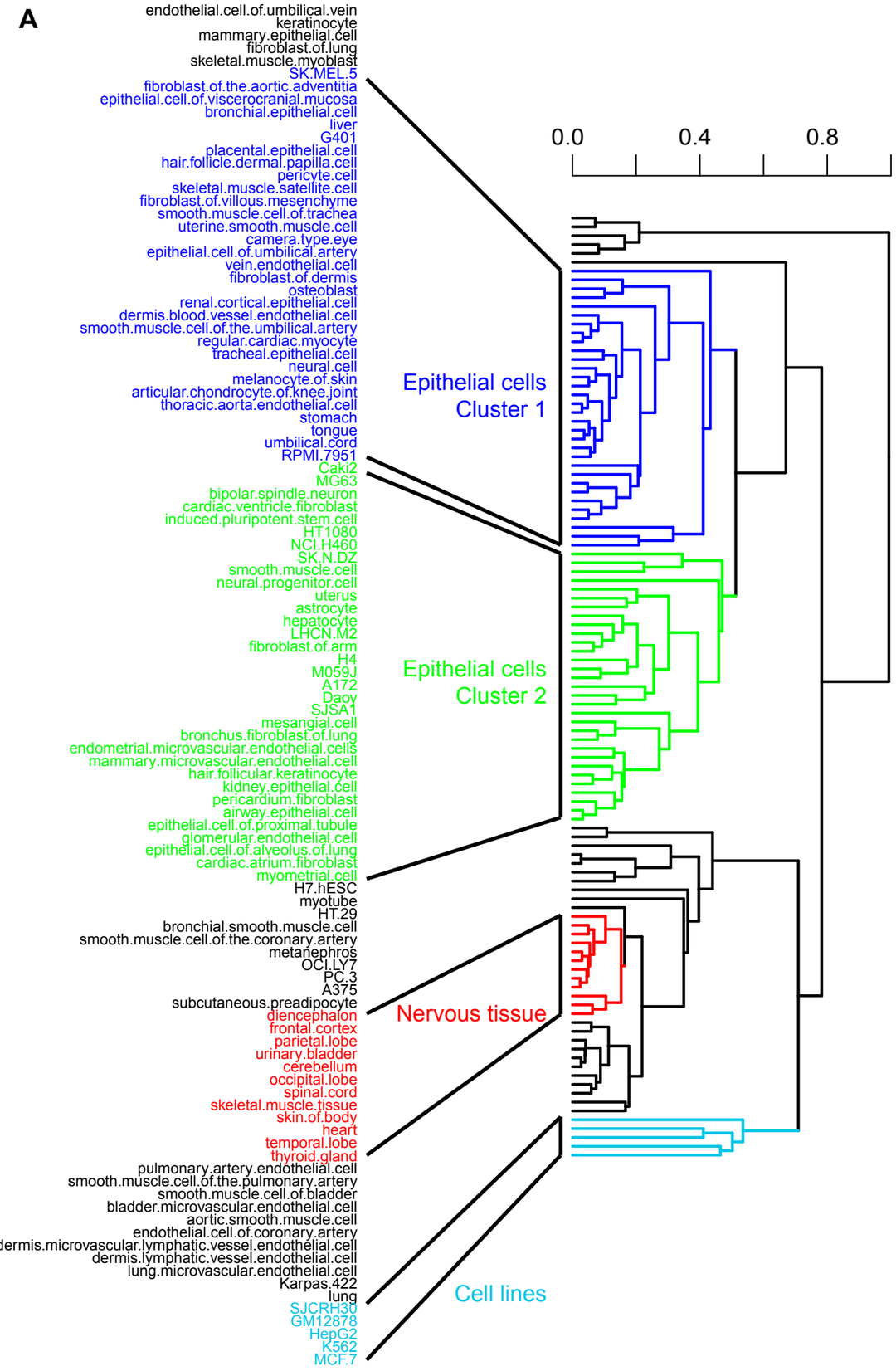

A
Cluster 1
Cluster 2
Cluster 3
Pearson correlation
1
−1
B
Cluster A
Cluster B
Cluster C

Figure 3


GCN5
gcn5Δ
GCN5
gcn5Δ
YPD
Menadione 50μM
RPS6B
rps6bΔ
RPL38
rpl38Δ
YPD
CdCl2 10μM
RPL26B
rpl26bΔ
RPL11B
rpl11bΔ

bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

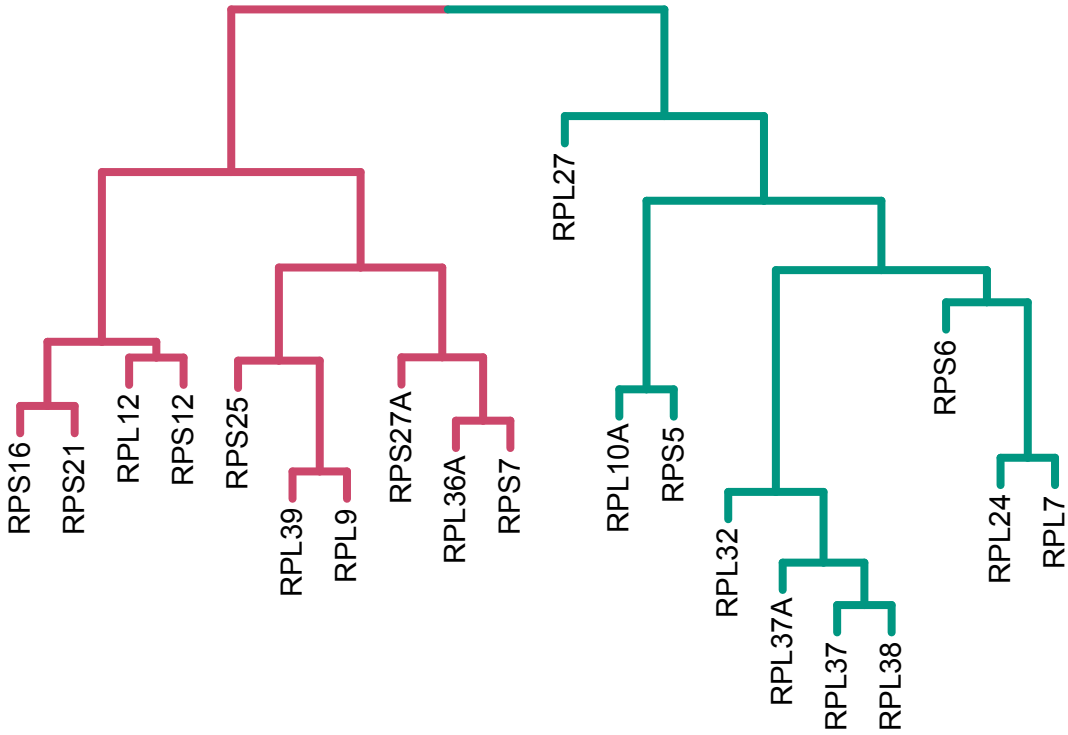
Figure 4



bioRxiv preprint doi: https://doi.org/10.1101/068361; this version posted August 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 5

A



B

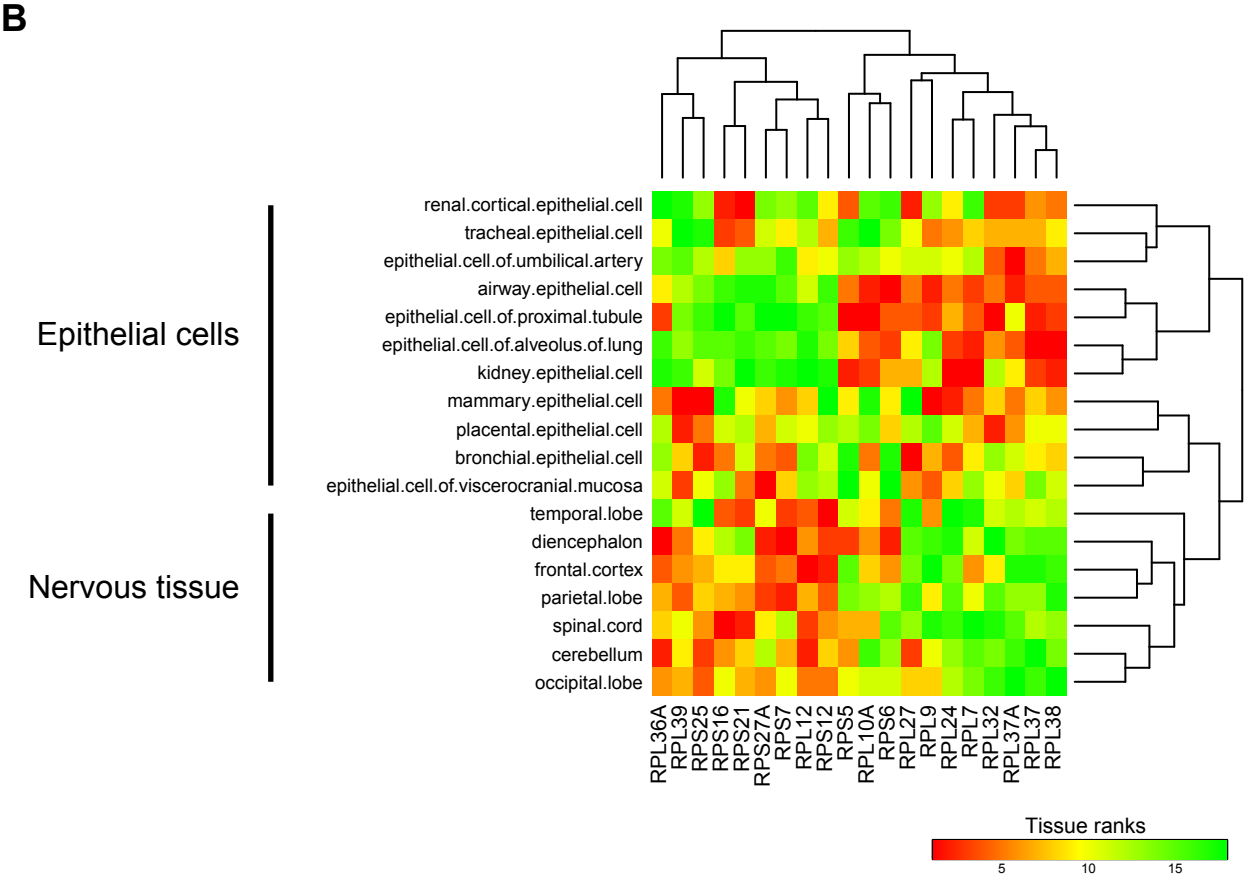
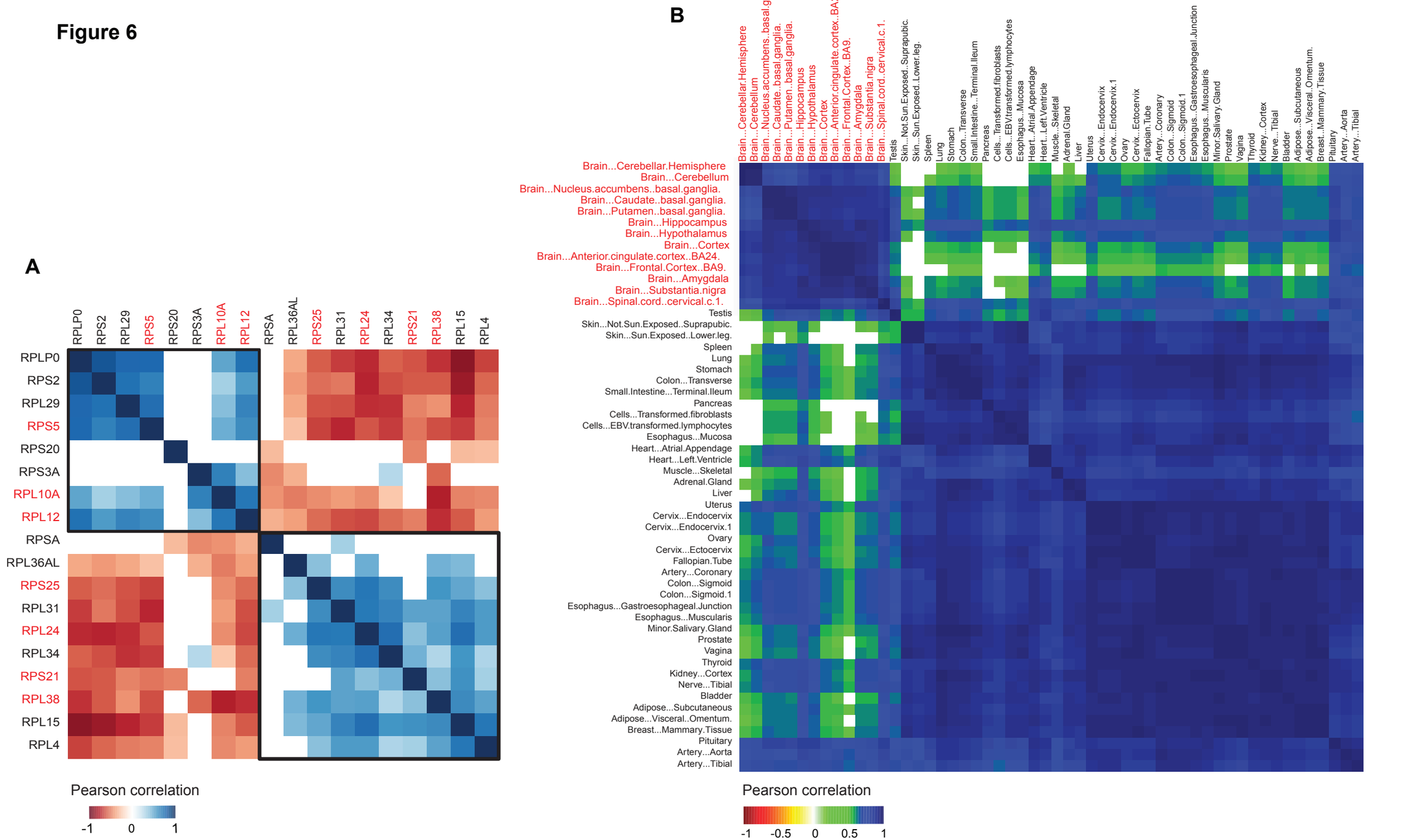


Figure 6


A
RPLP0
RPS2
RPL29
RPS5
RPS20
RPS3A
RPL10A
RPL12
RPSA
RPL36AL
RPS25
RPL31
RPL24
RPL34
RPS21
RPL38
RPL15
RPL4
Pearson correlation
-1
0
1
B
Brain...Cerebellar.Hemisphere
Brain...Cerebellum
Brain...Nucleus.accumbens..basal.ganglia.
Brain...Caudate..basal.ganglia.
Brain...Putamen..basal.ganglia.
Brain...Hippocampus
Brain...Hypothalamus
Brain...Cortex
Brain...Anterior.cingulate.cortex..BA24.
Brain...Frontal.Cortex..BA9.
Brain...Amygdala
Brain...Substantia.nigra
Brain...Spinal.cord..cervical.c.1.
Testis
Skin...Not.Sun.Exposed..Suprapubic.
Skin...Sun.Exposed..Lower.leg.
Spleen
Lung
Stomach
Colon...Transverse
Small.Intestine...Terminal.Ileum
Pancreas
Cells...Transformed.fibroblasts
Cells...EBV.transformed.lymphocytes
Esophagus...Mucosa
Heart...Atrial.Appendage
Heart...Left.Ventricle
Muscle...Skeletal
Adrenal.Gland
Liver
Uterus
Cervix...Endocervix
Cervix...Endocervix.1
Ovary
Cervix...Ectocervix
Fallopian.Tube
Artery...Coronary
Colon...Sigmoid
Colon...Sigmoid.1
Esophagus...Gastroesophageal.Junction
Esophagus...Muscularis
Minor.Salivary.Gland
Prostate
Vagina
Thyroid
Kidney...Cortex
Nerve...Tibial
Bladder
Adipose...Subcutaneous
Adipose...Visceral..Omentum.
Breast...Mammary.Tissue
Pituitary
Artery...Aorta
Artery...Tibial
Pearson correlation
-1
-0.5
0
0.5
1