Project title:

The natural diversity of the yeast proteome reveals chromosome-wide dosage compensation in aneuploids

Key words:

Yeast, large-scale, high-throughput, diversity, aneuploidy, proteomics, wild isolates

Project description

Aneuploidy, an imbalance in chromosome copy numbers, causes genetic disorders, and drives cancer progression, drug tolerance, and antimicrobial resistance. While aneuploidy can confer stress resistance, it is not well understood how cells overcome the fitness burden caused by aberrant chromosomal copy numbers. Studies using both systematically generated and natural aneuploid yeasts triggered an intense debate about the role of dosage compensation, concluding that aneuploidy is transmitted to the transcriptome and proteome without significant buffering at the chromosome-wide level, and is, at least in lab strains, associated with significant fitness costs. Conversely, systematic sequencing and phenotyping of large collections of natural isolates revealed that aneuploidy is frequent and has few – if any – fitness costs in nature. To address these discrepant findings, we developed a platform that yields highly precise proteomic measurements across large numbers of genetically diverse samples and applied it to natural isolates collected as part of the 1011 genomes project (Peter, J. *et al*, 2018). For 613 of the isolates, we were able to match the proteomes to their corresponding transcriptomes and genomes, subsequently quantifying the effect of aneuploidy on gene expression by comparing 95 aneuploid with 518 euploid strains. We find, as in previous studies, that aneuploid gene dosage is not buffered chromosome-wide at the transcriptome level. Importantly, in the proteome, we detect an attenuation of aneuploidy by about 25% below the aneuploid gene dosage in natural yeast isolates. Furthermore, this chromosome-wide dosage compensation is associated with the ubiquitin-proteasome system (UPS), which is expressed at higher levels and has increased activity across natural aneuploid strains. Thus, through systematic exploration of the species-wide diversity of the yeast proteome, we shed light on a long-standing debate about the biology of aneuploids, revealing that aneuploidy tolerance is mediated through chromosome-wide dosage compensation at the proteome level.

Data Processing

Protein-wise fasta files were created by inferring single nucleotide polymorphisms for each strain based on the reference genome of the S288c strain. In case of heterozygosity, one of the possible alleles was randomly inferred (Peter, J. *et al*, 2018; Legras, J.-L. *et al.* ,2018). For non-reference genes, a single representative sequence per protein was available based on the genomes. The proteome for the reference strain S288c was obtained from Uniprot (UP000002311, accessed 10/02/2020). Sequences of strains present in the original strain collections (Peter, J. *et al*, 2018; Legras, J.-L. *et al.* ,2018) and subject to intellectual property restrictions were excluded from our study, leading to the inclusion of 1023 strains in the processing.

In order to reduce the processing time and limit the search space to relevant peptides, the protein-wise fasta files were processed to select peptides well shared across the strain collection. The protein sequences were thus trypsin-digested *in-silico* whilst disregarding missed cleavages. Non-proteotypic peptides were excluded and only peptides shared by 80% of the strains were selected for further analysis. This list of peptides was used to filter the experimental library. Raw mass spectrometry files were processed using the filtered spectral library with the DIA-NN software (Version 1.7.12, Demichev, V.,*et al.* 2020). Default parameters of the software were used except for the following: Mass accuracy: 20, Mass accuracy MS1: 12. As the peptides selected were not necessarily present ubiquitously in all the strains, an additional step was required to remove false-positive peptide assignments (entries where a peptide is detected in a strain where it should be absent). This filter led to the exclusion of ~1% of the entries.

Samples with insufficient MS2 signal quality (~5.7×10^7), and entries with Q.Value (> 0.01), PG.Q.Value (> 0.01) were removed. Outlier samples were detected based on both the total ion chromatograms (TIC) and the number of identified precursors per sample (Z-Score > 2.5 SD) and were excluded from further analysis. Precursors normalized values as inferred by DIA-NN that were well detected across the samples (in at least 80% of the strains) and with CV < 0.3 in the quality control samples were retained. Subsequently, batch correction was carried out at the precursor level by bringing median precursor quantities of each batch to the same value. Proteins were then quantified using the maxLFQ (Cox, J. *et al*, 2014) function implemented in the DIA-NN R package, resulting in a dataset containing 1576 proteins for 796 strains. Missing values (<4% of all values) were imputed using K-Nearest Neighbors (KNN) imputation (Troyanskaya, O. *et al.,* 2001).

Sample preparation and LC-MS

The samples for proteomics were prepared in 96-well plates as previously described (Messner, C. B. *et al.* 2021), with up to four plates processed in parallel. For yeast lysis, 200 µL of lysis buffer (100 mM ammonium bicarbonate, 7 M urea) and ~100 mg glass beads were added to each well, followed by 5 min bead beating at 1500 rpm (Spex Geno/Grinder). For reduction and alkylation, 20 µL of 55 mM DL-dithiothreitol (1 h incubation at 30 °C) and 20 µL of 120 mM iodoacetamide (incubated for 30 min in the dark at ambient temperature). Subsequently, 1 mL of 100 mM ammonium bicarbonate was added per well, centrifuged (3220 x g, 3 min) and 230 µL of this mixture were transferred to plates pre-filled with 0.9 µg trypsin per well. The samples were incubated for 17 h at 37 °C and the digestion was subsequently stopped by adding 24 µL of 10% formic acid. The mixtures were cleaned-up using C18 96-well plates, with 1 min centrifugations between the steps at the described speeds. The plates were conditioned with methanol (200 µL, centrifuged at 50 x g), washed twice with 50% acetonitrile (ACN, 200 µL, centrifuged at 50 x g) and equilibrated three times with 3% ACN/0.1% formic acid (200 µL, centrifuged at 50, 80 and 100 x g, respectively). Then, 200 µL of digested sample was loaded (centrifuged at 100 x g) and washed three times with 3% ACN/0.1% formic acid (200 µL, centrifuged at 100 x g). After the last washing step, the plates were centrifuged at 180 x g. Subsequently, peptides were eluted in three steps, twice with 120 µL and once with 130 µL of 50% ACN (180 x g), and collected in a plate (1.1 mL, square well, V-bottom). The collected material was completely dried on a vacuum concentrator and redissolved in 40 µL of 3% ACN/0.1% formic acid before transfer to a 96-well plate. The final peptide concentration was estimated by absorption measurements at 280 nm with a Lunatic photometer (Unchained Labs, 2 µL of sample). All pipetting steps were performed with a liquid handling robot (Biomek NXP) and samples were shaken on a thermomixer (Eppendorf ThermomixerC) after each step.

For the collection of natural isolates, liquid chromatography was performed on a nanoAcquity UPLC system (Waters) coupled to a Sciex TripleTOF 6600. Peptides (2 µg) were separated on a Waters HSS T3 column (150 mm x 300 µm, 1.8 µm particles) ramping in 19 minutes from 3% B to 40% B (Solvent A: 1% acetonitrile/0.1% formic acid; solvent B: acetonitrile/0.1% formic acid) with a non-linear gradient (SI Table 9). The flow rate was set to 5 µL/min. The SWATH acquisition method consisted of an MS1 scan from m/z 400 to m/z 1250 (50 ms accumulation time) and 40 MS2 scans (35 ms accumulation time) with variable precursor isolation width covering the mass range from m/z 400 to m/z 1250

The library used was generated from “gas-phase fractionation” runs using scanning SWATH and small precursor isolation windows 5 µg of yeast digests were injected and run on a nanoAcquity UPLC (Waters) coupled to a SCIEX TripleTOF 6600 with a DuoSpray Turbo V source. Peptides were separated on a Waters HSS T3 column (150 mm × 300 µm, 1.8 µm particles) with a column temperature of 35 °C and a flow rate of 5 µL/min. A 55-min linear gradient ramping from 3% ACN/0.1% formic acid to 40% ACN/0.1% formic acid was applied. The ion source gas 1 (nebulizer gas), ion source gas 2 (heater gas) and curtain gas were set to 15 psi, 20 psi and 25 psi, respectively. The source temperature was set to 75 °C and the ion spray voltage to 5,500 V. In total, 12 injections were run with the following m/z mass ranges: 400–450, 445–500, 495–550, 545–600, 595–650, 645–700, 695–750, 745–800, 795–850, 845–900, 895–1,000 and 995–1,200. The precursor isolation window was set to m/z 1 except for the mass ranges m/z 895–1,000 and m/z 995–1,200, where the precursor windows were set to m/z 2 and m/z 3, respectively. The cycle time was 3 s, consisting of high- and low-energy scans, and data were acquired in ‘high-resolution’ mode. The spectral libraries were generated using library-free analysis with DIA–NN directly from these Scanning SWATH acquisitions. For this DIA–NN analysis, MS2 and MS1 mass accuracies were set to 25 and 20 ppm, respectively, and scan window size was set to 6. This library was later filtered in R to reduce the search space by retaining peptides shared across 80% of the strains, as described in the Data processing section.

File naming

The file names include the following information:
injectionnr_wellnumber_platenumber_strainname_abs280

For example, the file 27_25_R1_25_BFG_0.51 was:
-the injection nr 30 (within the plate R1);
-the strain was in well number 25 (counted row-wise);
-the sample was in the plate R1;
-repeated well number / injection source (e.g. qc);
-the sample is from strain BFG;
-the measured absorption at 280 nm is 0.51 (was used for adjusting the injection volume)

References

Peter, J. *et al.* Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature* **556**, 339–344 (2018).

Legras, J.-L. *et al.* Adaptation of S. cerevisiae to Fermented Food Environments Reveals Remarkable Genome Plasticity and the Footprints of Domestication. Mol. Biol. Evol. 35, 1712–1727 (2018).

Messner, C. B. *et al.* Ultra-fast proteomics with Scanning SWATH. Nat. Biotechnol. 39, 846–854 (2021).

Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat. Methods 17, 41–44 (2020).

Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol. Cell. Proteomics 13, 2513–2526 (2014).

Troyanskaya, O. et al. Missing value estimation methods for DNA microarrays. Bioinformatics 17, 520–525 (2001).