Supporting Information for:

Highly Reproducible Label Free Quantitative Proteomic Analysis of RNA Polymerase Complexes

Amber L. Mosley$^{1,2,*}$, Mihaela E. Sardiu$^{1,*}$, Samantha G. Pattenden$^1$, Jerry L. Workman$^1$, Laurence Florens$^1$, and Michael P. Washburn$^{1,3,‡}$

$^1$Stowers Institute for Medical Research, Kansas City, MO 64110, USA.
$^2$Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA
$^3$Department of Pathology and Laboratory Medicine, The University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, KS 66160, USA

* These authors contributed equally to this work.

‡To whom correspondence should be addressed:
Michael P. Washburn
Stowers Institute for Medical Research
1000 E. 50th St.
Kansas City, MO64110
Phone: 816-925-4457
Fax: 816-926-4694
e-mail: mpw@stowers.org

RUNNING TITLE : Quantitative Proteomics of RNA Polymerases
Supplemental Figure 1: Distribution of the first left singular vector (LSV) from singular value decomposition (SVD). We applied singular value decomposition to a matrix (12 purifications x 355 prey proteins) where each matrix element represents the normalized spectral count, i.e. dNSAF, for each prey in every bait. We used the information of the first left singular vector to rank the proteins based on their corresponding coefficient value of the first vector. To determine the proteins that are enriched in the sample, the coefficients of the first vector were plotted in linear scale, as well as in log scale in the inset. The top 31 proteins with the highest coefficient values, which delimited very well from the remaining proteins, were all components of the RNA polymerase complexes.
Supplemental Figure 2: Coefficient of variation (CV) values for the biological versus technical replicates for the Rpb3-TAP and Rpb11-TAP purifications. CV values were calculated by dividing the standard deviation by the average of the dNSAF values for each protein as indicated at the bottom of the graph. We compared the CV values for at least three technical replicates and three biological replicates and determine that the variation for the components of DSIF (Spt4 and Spt5) and TFIIF (Tfg1, Tfg2, and Taf14) are significantly higher across biological replicates than across technical replicates. This indicates that the variation that we observe for these proteins is likely a result of biological dynamics rather than a technical problem with our methods of detection and analysis.

The following supplemental tables are also available as separate files.

Supplemental Table 1: All peptides per protein that were identified in our analyses. Proteins with peptides passing the selection criteria as given in the table below are listed by decreasing sequence coverage (percent of the protein sequence covered by the detected peptides). The accession number for each protein is given in the Locus column. Total sequence coverage per protein is also given in the Sequence Coverage column. Within the peptide sequence the presence of an oxidized methionine (+16 daltons) is represented by M#. CAUTION: This table is over 500 printed pages.

Supplemental Table 2: All dNSAF values for all proteins prior to contaminant extraction. The name of each purification is indicated at the top of each row with the replicate number being represented by P#. The table also contains the accession number (Locus), NCBI description (Description), Yeast open reading frame number (ORF ID), and protein common name (Acronym) for all proteins identified during our analysis. The table is sorted in descending order by the average dNSAF across all purifications.
**Supplemental Table 3: Spectral FDR values for each of the purifications.** The spectral false discovery rate (FDR) is given for each of the purifications as well as the average FDR for all purifications.

**Supplemental Table 4: Final set of RNA Polymerase dNSAF values following contaminant extraction.** The accession number (Locus), protein common name (Acronym), and rank according to the average dNSAF value is given for all proteins passing contaminant extraction as described in the methods section. Zero values indicate that either the protein was not identified in that analysis or that the protein was more enriched in the MOCK purification than in the specific RNA Polymerase purifications.

**Supplemental Table 5: Total number of unique peptides and spectral counts identified in each purification.** The total number of spectral counts identified that were assigned to an RNA Polymerase subunit is also given in order to make an estimate of purification purity.

**Supplemental Table 6: Top 50 most enriched proteins in RNAP dataset as identified by singular value decomposition.** Using SVD, we obtained values for the first left singular vector (LSV) to identify the proteins which were the most highly enriched in the RNAP dataset (which includes the Rpb8-TAP, Rpa190-TAP, Rpb3-TAP, and Ret1-TAP replicates). The top 50 values from the LSV are reported in this table.

**Supplemental Table 7: Total number of spectral counts identified per unique peptide for each of the 31 RNA Polymerase subunits.** The total number of spectral counts (SpC) per unique peptide (UP) is given for all purifications in the manuscript. In this table, the total spectral counts for each unique peptide sequence are summed for all charge and modification states.

**Supplemental Table 8: Proteotypic peptides identified for each of the 31 RNA Polymerase subunits.** Each protein was sorted according to their inclusion in one of the following groups: specific to RNA Polymerase I (group A), specific to RNA Polymerase II (group B), specific to RNA Polymerase III (group C), shared between all three RNA Polymerases (group D), shared between RNA Polymerase I and RNA Polymerase III (group E). For groups A, B, and C; a peptide was considered proteotypic if it was identified by an average of 100 spectra or more in the respective specific enzyme purification. For group D, a peptide was considered to be proteotypic if it was identified by an average of 100 spectra or more in at least 2 of the 4 purifications. For group E, a peptide was considered proteotypic if it was identified by an average of 100 spectra or more in both the Rpa190-TAP and Ret1-TAP replicates.

**Supplemental Table 9: dNSAF and Spectral Count values for the biological and technical replicates for Rpb3-TAP and Rpb11-TAP subunits.** The total number of distributed spectral counts (dS) and the dNSAF value for each technical replicate is given. Each of the technical replicate spectral counts per protein were summed to give the total number of spectra detected per biological replicate.