Molecular Cell, Volume *27*

Supplemental Data

Systematic Analysis of the Protein Interaction Network for the Human Transcription Machinery Reveals the Identity of the 7SK Capping Enzyme

Célia Jeronimo, Diane Forget, Annie Bouchard, Qintong Li, Gordon Chua, Christian Poitras, Cynthia Thérien, Dominique Bergeron, Sylvie Bourassa, Jack Greenblatt, Benoit Chabot, Guy G. Poirier, Timothy R. Hughes, Mathieu Blanchette, David H. Price, and Benoit Coulombe

Supplemental Discussion

Computational Validation of Protein Interactions

Our purification procedure was specifically designed to preserve the integrity of the purified complexes as they exist in live human cells. Affinity purification of tagged proteins theoretically allows the isolation of all protein complexes containing the tagged polypeptide. This method does not, however, allow for the direct determination of the abundance of the purified complexes. In addition, the high sensitivity of mass spectrometry requires the development of methods that discriminate between specific and spurious interactions (Patil and Nakamura, 2005; Krogan et al., 2006; Gavin et al., 2006). This was accomplished through the development of an algorithm that selects high confidence interactions by assigning Interaction Reliability (IR) scores to each protein interaction (a schematic representation of our computational procedure is provided in Supplemental Figure S2). The sensitivity and specificity of the algorithm was evaluated using literature-based classification of protein interactions. We selected as high confidence interactions those for which the IR score exceeded a threshold (IR score

above 0.6729) predicted to miss as false-negatives only 17% of a set of literature-supported interactions while incorrectly retaining only 17% of a set of interactions without literature support as false-positives. The selected interactions were used to build a protein interaction map (see Figure 2 of the article).

Supplemental Experimental Procedures

Interaction Reliability Score

Let I_0 be the set of 2008 putative interactions detected by mass spectrometry and for which the Mascot (MS) score in the induced eluate was at least five times larger than in the equivalent non-induced eluate (Supplemental Table S1). Each interaction in I_0 was assigned an Interaction Reliability (IR) score, reflecting the likelihood that the interaction is correct, calculated as follows. Let m(i,j) be the MS score for interaction between proteins P_i and P_j. Let t(i,j) be the number of triangles in which interaction (P_i,P_j) is involved: $t(i,j) = | \{k: (i,k) \in I_0 \text{ and } (k,j) \in I_0\} |$. Interactions involved in a number of triangles are more likely to be correct than those involved in none. A set $P \subset I_0$ of 135 interactions strongly supported by the literature was extracted and used as representatives of true interactions (Supplemental Table S1). A set $N \subset I_0$ of 53 interactions that were judged to have low biological support were used as representatives of likely false-positives (Supplemental Table S1). The posterior probabilities $Pr[m(i,j) \mid (i,j) \in P]$ and $Pr[m(i,j) \mid (i,j) \in N]$, were then empirically estimated by fitting a gamma distribution to each subset. Similarly, posterior probabilities $Pr[t(i,j) \mid (i,j) \in P]$ and $Pr[t(i,j) \mid (i,j) \in N]$ were estimated using empirical frequencies. Finally, the reliability score for the predicted interaction (i,j) is given by the Naïve Bayes classifier:

$$\text{Reliability}(i,j) = \Pr[\,(i,j) \in P \mid m(i,j), t(i,j)\,]$$

$$= \frac{\Pr[\,m(i,j) \mid (i,j)\in P\,] * \Pr[\,t(i,j) \mid (i,j)\in P\,] * \Pr[\,(i,j)\in P\,]}{\Sigma_{x\in\{P,N\}} \Pr[\,m(i,j) \mid (i,j)\in x\,] * \Pr[\,t(i,j) \mid (i,j)\in x\,] * \Pr[\,(i,j)\in x\,]}$$

where we set prior probabilities $\Pr[\,(i,j)\in P\,] = \Pr[\,(i,j)\in N\,] = 0.5$.

Expression of TAP-Tagged Proteins and Affinity Purification of Protein Complexes

Induction for 24-72 hours with 3-6 μM ponasterone A (Invitrogen) was used to express the TAP-tagged proteins at near physiological levels. Whole cell extracts prepared from induced and non-induced stable EcR-293 cell lines were subjected to purification by the TAP procedure as previously described (Jeronimo et al., 2004). Nuclease-treated (RNase A and/or DNase) whole cell extracts were used in some experiments to ensure that the purified interaction partners are associated with the tagged complex through protein-protein interactions (Robert et al., 2002).

Identification of Proteins by Mass Spectrometry

The TAP eluates were run on SDS gels, stained with silver or Sypro Ruby (Bio-Rad) and gel slices were excised and digested with trypsin as previously described (Jeronimo et al., 2004). The resulting tryptic peptides were purified and identified by tandem mass spectrometry (LC-MS/MS) with microcapillary reversed-phase high-pressure liquid chromatography coupled to an LCQ DecaXP (ThermoFinnigan), LTQ or LTQ-Orbitrap (ThermoElectron) quadrupole ion trap mass spectrometer with a nanospray interface.

Gel Filtration Chromatography

Affinity purified protein complexes were concentrated by dialysis in buffer F containing 10 mM Hepes pH 7.9, 100 mM NaCl, 0.1 mM EDTA, 5% glycerol and 0.5

mM DTT. An aliquot (50 µL) of the concentrated eluate was fractionated on a Superose 6 PC 3.2/30 column (2.4 ml) previously equilibrated in buffer F using the ÄKTA FPLC system (GE Healthcare). The column was run in buffer F at a flow rate of 0.04 ml/minutes and 50-µL fractions were collected. Aliquots of each five fractions were pooled, concentrated and analyzed by Western blot.

Antibodies

Primary antibodies anti-RPAP1, anti-FLJ21908 and anti-XAB1 were obtained from the CIM Antibody Core at Arizona State University (Tempe, Arizona). The rabbit anti-BCDIN3 antibody was generated using the BCDIN3 peptide VPPHQEAASGELRGGTERGPGRC and obtained from Chemicon (Millipore). The TAP-specific anti-Calmodulin Binding Peptide (CBP) antibody (clone C16T) was purchased from Upstate. Other primary antibodies were anti-SART3/Tip110 (gift from J.J. He, Indiana University School of Medicine, Indianapolis, Indiana) and anti-CDK9 (C-20, Santa Cruz). Anti-Cyclin T1 and anti-HEXIM1 antibodies have been previously characterized (Michels et al., 2003). The horseradish peroxidase-conjugated secondary antibodies were purchased from GE Healthcare.

DNA Microarray Analysis of Tet-Promoter Mutants

Mutant and isogenic wild type cultures were grown in parallel in SC medium with 10 µg/ml doxycycline for a total of 22 hours. Total RNA was prepared by hot acidic phenol:chloroform extraction followed by ethanol precipitation. Microarrays were composed of 70-mer oligonucleotides, each specific to a different yeast gene (Operon Technologies), which were spotted onto poly-L-lysine-coated slides. Each array was normalized using grid-by-grid lowess smoothing. All measurements were taken in fluor-

reversed pairs (*e.g.*, each time a mutant was analyzed, it was hybridized to two arrays, combined by averaging the normalized log(ratio)).

RNA Blot

Total RNAs from HeLa cells, HEK 293 cells treated with siRNA or BCDIN3-TAP eluate were extracted using mirVana miRNA isolation kit (Ambion). From an aliquot of the BCDIN3 eluate containing about 10 μg of protein, 3 μg of RNA was recovered. Probes for 7SK, U6 and U2 were 5'-end labeled using T4 PNK (NEB), following manufacturer's instructions. RNAs were separated on a 6% denaturing gel, and transferred to a Nytran N membrane (Whatman). The membrane was blocked with ULTRAhyb (Ambion), and probed with labeled 7SK, U6, or U2 oligos, following the manufacturer's instruction. RNA probes used for 7SK, U6 and U2 were as follows (numbers in parenthesis indicate the regions to which the probes hybridize):

7SK (121-160): GGGGAUGGUCGUCCUCUUCGACCGAGCGCGCAGCUUCGGG

U6 (60-99): GGAACGCUUCACGAAUUUGCGUGUCAUCCUUGCGCAGGGG

U2 (97-136): CGGAGCAAGCUCCUAUUCCAUCUCCCUGCUCCAAAAAUCC

Immunodepletion

Protein A-sepharose beads (GE Healthcare) were saturated with BSA (Roche) and incubated with preimmune or anti-BCDIN3 serum for 12 hours at 4°C. The antibody-bound beads were then divided into three 50 μl-aliquots and each aliquot was used for one round of depletion. A total of 100 μl of HEK 293 whole cell extracts containing approximately 0.5 mg of protein was add to the first aliquot of antibody-bound beads and rocked for 2 hours at 4°C. After a brief spin, the supernatant was transferred to the next aliquot of beads and the process repeated. After three rounds of depletion, the

supernatants were analyzed by Western blotting and assayed in the *in vitro* methyltransferase assay.

Methyltransferase Assay

The RNA methyltransferase assay was performed in 20 µl of total reaction mixture, containing 20 mM Tris-HCl (pH 8.0), 0.5 mM DTT, 2 mM EDTA, 50 mM KCl, 5% Glycerol, 100 units of RNAguard RNase inhibitor (GE Healthcare), 5 µCi of ^{3}H-AdoMet (TRK865, GE Healthcare) (63.0 Ci/mmol), 250 ng of purified BCDIN3-His and 200 ng of recombinant 7SK snRNA prepared as described previously (Li et al., 2005). When using HEK 293 whole cell extracts as a source of enzyme, the final volume of the reaction was increased to 50 µl. When mentioned, 100 µM of S-adenosyl-L-homocysteine (AdoHcy) (Sigma) and 0.5 mg/ml of RNase A (USB) were used. After a 1-hour incubation at 30°C, the reaction was terminated by adding 200 µl of a stop solution (0.1 M NaOAc (pH 5.2), 0.5% SDS, 2 mM EDTA, 100 µg/ml tRNA). The reaction was extracted with phenol/chloroform and the RNA was ethanol precipitated and analyzed in a 6% urea-polyacrylamide gel. The gel was fixed (45% methanol, 10% acetic acid), treated with Amplify (NAMP100, GE Healthcare) and exposed to Hyperfilm-MP film (GE Healthcare) at -80°C.

Dephosphorylation and Decapping Assay

Recombinant 7SK snRNA was incubated for 1 hour at 37°C with Calf Intestinal Alkaline Phosphatase (CIAP) (Invitrogen), following manufacturer's instructions and supplemented by 100 units of RNAguard RNase inhibitor (GE Healthcare). The volume of the reaction was brought to 200 µl with nuclease-free H_2O containing 20 µg of

glycogen (Roche), phenol:chloroform extracted and the RNA ethanol precipitated, air-dried and resuspended in H_2O.

Five pmoles of *in vitro* capped 7SK snRNA was digested with Tobacco Acid (TA) pyrophosphate (2.5, 7.5 and 15 units) at 37°C for 1 hour according to the manufacturer's instructions (Epicentre Biotechnologies), supplemented by 100 units of RNAguard RNase inhibitor (GE Healthcare). The reaction was stopped by adding 200 µl of a stop solution, phenol/chloroform extracted and ethanol precipitated as mentioned above. The RNA was analyzed in a 6% urea-polyacrylamide gel and ethidium bromide stained. The gel was then fixed (45% methanol, 10% acetic acid), treated with Amplify (NAMP100, GE Healthcare) and exposed to Hyperfilm-MP film (GE Healthcare) at -80°C.

Supplemental References

Gavin,A.C., Aloy,P., Grandi,P., Krause,R., Boesche,M., Marzioch,M., Rau,C., Jensen,L.J., Bastuck,S., Dumpelfeld,B., Edelmann,A., Heurtier,M.A., Hoffman,V., Hoefert,C., Klein,K., Hudak,M., Michon,A.M., Schelder,M., Schirle,M., Remor,M., Rudi,T., Hooper,S., Bauer,A., Bouwmeester,T., Casari,G., Drewes,G., Neubauer,G., Rick,J.M., Kuster,B., Bork,P., Russell,R.B., and Superti-Furga,G. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631-636.

Jeronimo,C., Langelier,M.F., Zeghouf,M., Cojocaru,M., Bergeron,D., Baali,D., Forget,D., Mnaimneh,S., Davierwala,A.P., Pootoolal,J., Chandy,M., Canadien,V., Beattie,B.K., Richards,D.P., Workman,J.L., Hughes,T.R., Greenblatt,J., and Coulombe,B. (2004). RPAP1, a novel human RNA polymerase II-associated protein affinity purified with recombinant wild-type and mutated polymerase subunits. Mol Cell Biol *24*, 7043-7058.

Krogan,N.J., Cagney,G., Yu,H., Zhong,G., Guo,X., Ignatchenko,A., Li,J., Pu,S., Datta,N., Tikuisis,A.P., Punna,T., Peregrin-Alvarez,J.M., Shales,M., Zhang,X., Davey,M., Robinson,M.D., Paccanaro,A., Bray,J.E., Sheung,A., Beattie,B., Richards,D.P., Canadien,V., Lalev,A., Mena,F., Wong,P., Starostine,A., Canete,M.M., Vlasblom,J., Wu,S., Orsi,C., Collins,S.R., Chandran,S., Haw,R., Rilstone,J.J., Gandi,K., Thompson,N.J., Musso,G., St Onge,P., Ghanny,S., Lam,M.H., Butland,G., Altaf-Ul,A.M., Kanaya,S., Shilatifard,A., O'Shea,E., Weissman,J.S., Ingles,C.J., Hughes,T.R., Parkinson,J., Gerstein,M., Wodak,S.J., Emili,A., and Greenblatt,J.F. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637-643.

Li,Q., Price,J.P., Byers,S.A., Cheng,D., Peng,J., and Price,D.H. (2005). Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. J. Biol Chem. *280*, 28819-28826.

Michels,A.A., Nguyen,V.T., Fraldi,A., Labas,V., Edwards,M., Bonnet,F., Lania,L., and Bensaude,O. (2003). MAQ1 and 7SK RNA interact with CDK9/cyclin T complexes in a transcription-dependent manner. Mol Cell Biol *23*, 4859-4869.

Patil,A. and Nakamura,H. (2005). Filtering high-throughput protein-protein interaction data using a combination of genomic features. BMC. Bioinformatics. *6*, 100.

Robert,F., Blanchette,M., Maes,O., Chabot,B., and Coulombe,B. (2002). A human RNA polymerase II-containing complex associated with factors necessary for spliceosome assembly. J. Biol. Chem. *277*, 9302-9306.

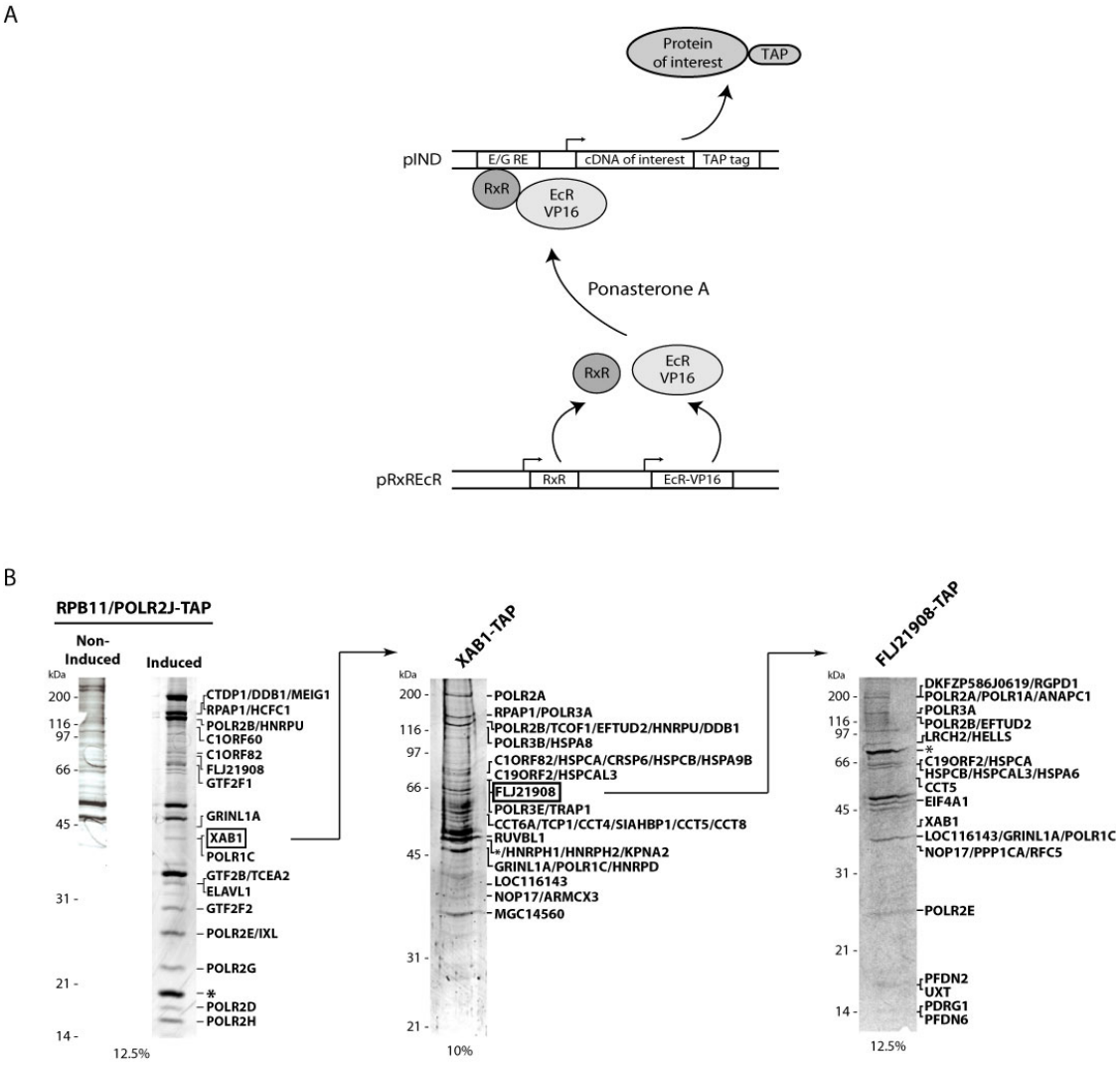

Figure S1. A Method for the Purification of Protein Complexes in Human Cells

(A) Schematic representation of the ecdysone-inducible system used to express, at physiological levels, the tagged protein upon addition of the inducer Ponasterone A. Expression was maintained near physiological levels through the use of the ecdysone-inducible system, which allows, by varying the dose of the inducer and the time of induction, calibration of the amount of the tagged protein made within the cells, as determined by Western blotting ((Jeronimo et al., 2004) and data not shown).

(B) Examples of SDS gels showing the composition of affinity purified protein complexes in reciprocal tagging experiments. Most bands that are not labeled contain proteins that have been either identified in a control eluate prepared from a non-induced cell extract or represent interactions that had IR scores below our selected threshold. The tagged protein is indicated by an asterisk. The proteins targeted in our reciprocal tagging experiments are boxed.

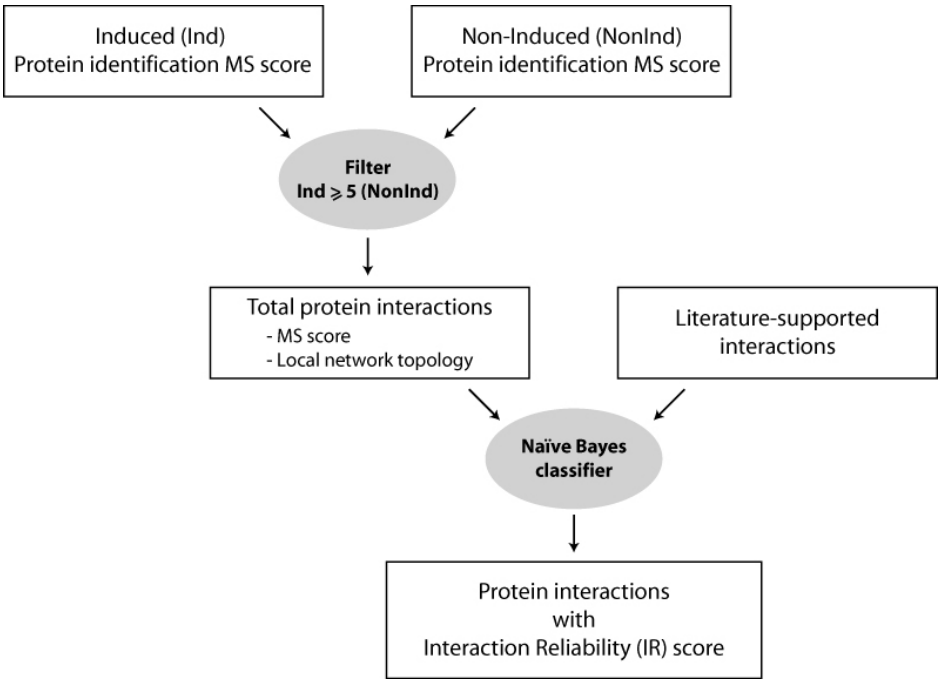

Figure S2. Computational Validation of Protein Interactions

For each purification, the non-specific interactions were filtered out by selecting proteins with a MS score in the induced (Ind) eluate at least 5 fold higher than in the equivalent non-induced (NonInd) eluate (Ind ≥ 5 (NonInd)), thereby eliminating proteins that bind non-specifically to our columns and very abundant cellular proteins that may remain as contaminants after the affinity purification steps. Each putative protein interaction selected at this stage was then assigned an IR score based on two criteria: (*i*) the MS score and (*ii*) the number of triangles, as defined by the local topology of the network (see Supplemental Experimental Procedures and Supplemental Table S1). The IR score was computed using a Naïve Bayes classifier, trained on a set of 135 positive interactions validated from the literature and a set of 53 negative interactions that were judged likely to be false-positives (based on the literature, but independently of the results of our experiments) (Supplemental Table S1).