FLAG-GATAD1 (*Homo sapiens*)

Method:
Immunoprecipitation followed by mass spectrometry

Caption:
HepG2 whole cell lysate was immunoprecipitated using the primary antibody (Sigma; F1804). The IP fraction was loaded on a 12% Bio-Rad TGX gel and separated with the Bio-Rad Tetra Cell system. The whole lane was excised and sent to the University of Alabama at Birmingham Cancer Center Mass Spectrometry/Proteomics Shared Facility.

Analysis of whole lane gel from HepG2: The sample was analyzed on a LTQ XL Linear Ion Trap Mass Spectrometer by LC-ESI-MS/MS. Peptides were identified using SEQUEST tandem mass spectral analysis with probability based matching at $p < 0.05$. SEQUEST results were reported with ProteinProphet protXML Viewer (TPP v4.4 JETSTREAM) and filtered for a minimum probability of 0.9. All protein hits that met these criteria were reported, including common contaminants. Fold enrichment for each protein reported was determined using a custom script based on the FC-B score calculation from the reference Mellacheruvu et al., 2013. The CRAPome: a contaminant repository for affinity purification mass spectrometry data. *Nat. Methods.* 10(8):730-736. Doi:10.1038/nmeth.2557. The target protein, GATAD1, was identified as the 18th enriched protein and the 2nd transcription factor based on IP-Mass Spectrometry.

Submitted by:
Collin White & Mark Mackiewicz

Lab:
Richard Myers, HAIB

Grant:
U54HG006998

Download:

Fold Enrichment (FC-B) Calculation

Mellacheruvu et al., 2013. The CRAPome: a contaminant repository for affinity purification mass spectrometry data. *Nat. Methods.* 10(8):730-736. Doi:10.1038/nmeth.2557.

Identification of target protein(s) from an antibody immunoprecipitation of whole cell lysates is challenging as there are numerous background contaminant peptides that are co-purified. These peptides are often not direct interactors with the target protein and do not contain epitopes that are specifically recognized by the antibody. Fold enrichment calculations such as those used by the **C**ontaminant **R**epository for **A**ffinity **P**urification Mass Spectrometry Data (CRAPome) can help in distinguishing true targets from background noise.

Using ~150 unique affinity purification mass spectrometry datasets generated by our lab, we grouped together (~6 datasets per group on average) similar experiments based on the cell line for which the mass spectrometry was performed, the organism the primary antibody was developed, and the molecular size of the protein targeted by the antibody. For example, one grouping was made based on 5 TFs (BATF, TSC22D3, MSC, KLF7, and CREM) with a size range from 14-39 kDa, for which the primary antibody was developed in mouse and the immunoprecipitation performed in GM12878 cells. Another grouping was made for 8 TFs (ZNF777, NFATC3, RC3H2, ZBTB11, CTCF, WDHD1, POLG, and KDM4B) with a size range of 85-142 kDa, using a rabbit antibody for IPs performed in K562 cells. Fold enrichment calculations are made for all identified peptides within a group. The script adopts a bootstrapping sampling procedure in which one immunoprecipitation dataset from a group serves as the experiment, and the remaining immunoprecipitations from the group serve as negative controls. Sampling continues until all immunoprecipitations within a group serve as both experiment and control.

	Baits (Antibodies)		
	j_1	j_2	j_3
Prey (Proteins) i_1	SC_{i1j1}	SC_{i1j2}	SC_{i1j3}
i_2	SC_{i2j1}	SC_{i2j2}	SC_{i2j3}
i_3	SC_{i3j1}	SC_{i3j2}	SC_{i3j3}

Where i represents protein i (the TF you are interested in)
Where j represents bait j (the antibody for your TF target)
Where SC_{ij} represents Spectral Count for protein i with bait j

Fold Change is calculated as the ratio of the normalized spectral count of the identified protein (prey) with its bait; and the average of the three highest normalized spectral counts for the identified protein across all negative control baits. A background factor, α, is included when C_i =0 (i.e., the protein is not observed in the negative controls) to prevent divisibility by zero. The fold change is calculated for each protein identified by mass spectrometry with its given bait. The TF target for the antibody should be highly enriched (ranked in the top 20 proteins by fold change) compared to other protein/contaminants identified with that antibody.

$$FC_{i,j} = (T_{i,j} + \alpha)/(C_i + \alpha)$$

$FC_{i,j}$ = Fold change for protein i with bait j
$T_{i,j}$ = Normalized spectral count of protein i with bait j
C_i = Average of 3 highest normalized spectral counts of protein i across the negative controls
α = Background factor = avg(N_x); where N_x is the normalization factor for bait x (where x is not j)

$$T_{i,j} = SC_{i,j}/N_j$$

$SC_{i,j}$ = Spectral count for protein i with bait j
N_j = Normalization factor for bait j

$$N_j = \Sigma SC_{xj}$$

SC_{xj} = Spectral count for protein x with bait j
Where protein x represents all other proteins including protein i

$$C_i = [\Sigma(SC_{i,x}/N_x)]/n$$

$SC_{i,x}$ = Spectral count for protein i with bait x (where bait x is not bait j)
n = Number of baits

$$N_x = \Sigma SC_{x,x}$$

N_x = Normalization factor for bait x (where bait x is not bait j)
$SC_{x,x}$ = Spectral count for protein x with bait x