Fold Enrichment (FC-B) Calculation

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.

<u>Baits (Antibodies)</u>

	j_1	j_2	j_3
i_1	$SC_{i_1 j_1}$	$SC_{i_1 j_2}$	$SC_{i_1 j_3}$
i_2	$SC_{i_2 j_1}$	$SC_{i_2 j_2}$	$SC_{i_2 j_3}$
i_3	$SC_{i_3 j_1}$	$SC_{i_3 j_2}$	$SC_{i_3 j_3}$

Prey (Proteins)

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.

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

$FC_{i,j}$ = 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

$$N_j = ΣSC_{xj}$$

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

$$C_i = [Σ(SC_{i,x}/N_x)]/n$$

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

$$N_x = ΣSC_{x,x}$$

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