1 Summary of ChIP-seq data analyzed

Data set	n_{double} (n_{peaks}, f)	Losses	Increases	Decreases	Unchanged
FOXA1-HNF4A	21832 (30428,0.72)	7730 (9606,0.8)	1378 (2567,0.54)	2161 (2707,0.8)	10563 (15548,0.68)
FOXA1-CEBPA	14064 (30428,0.46)	1442 (2358,0.61)	502 (1706,0.29)	1090 (1814,0.6)	11030 (24550,0.45)
HNF4A-CEBPA	16235 (39955,0.41)	1939 (4219,0.46)	733 (2321,0.32)	1000 (1941,0.52)	12563 (31474,0.4)
(EE) FIS-CRP	293 (1545,0.19)	140 (935,0.15)	—	—	153 (610,0.25)
(ME) FIS-CRP	487 (594,0.82)	36 (70,0.51)	—	—	451 (524,0.86)
(EE) CRP-FIS	460 (1551,0.3)	176 (1050,0.17)	—	—	284 (501,0.57)
ERα-FOXA1	3269 (7369,0.44)	2738 (5689,0.48)	—	21 (66,0.32)	510 (1614,0.32)
CDX2-HNF4A	18398 (30963,0.59)	14291 (25664,0.56)	—	—	4107 (5299,0.78)
RTG3-GCN4	651 (1065,0.61)	562 (954,0.59)	—	—	89 (111,0.8)
GCN4-RTG3	722 (2407,0.3)	379 (1486,0.26)	—	—	343 (921,0.37)

Table 1: Summary of ChIP-seq data from data sets analyzed in Figure 2. The second column describes the total number of doubly bound regions in each data set. In parentheses, the total number of primary TF peaks, and the fraction of primary TF peak regions that are doubly bound are shown ($f = n_{double}/n_{peaks}$). The remaining columns list the number of primary TF peaks that were lost, increased in rank, decreased in rank, or unchanged in rank, from genomic regions bound by both primary and partner TFs. The numbers in parentheses are the number of primary TF peaks that were lost, increased in rank, decreased in rank, or unchanged in rank, from all genomic regions bound by the primary TF.

2 Log-normal distributions typically fit peak intensity distributions better than Gamma or Gaussian distributions

The log-likelihood obtained from fitting a distribution to a set of peak intensity pairs is computed as

$$\log(P(\mathbf{D}|\mathbf{\Theta}_X, \mathbf{\Theta}_Y)) = \sum_{i=1}^{N} \log\left(f(x_i; \mathbf{\Theta}_X) f(y_i; \mathbf{\Theta}_Y)\right)$$

where f is either the cooperative (f_1) or non-cooperative (f_0) distribution, $\mathbf{D} = \{(x_i, y_i)\}_{i=1}^{N}$ are peak intensity pairs, while $\mathbf{\Theta_X}$ and $\mathbf{\Theta_Y}$ are parameters of f. A larger log-likelihood value indicates a better fit to data. We chose f to be Log-normal, Gamma and Gaussian distributions while fitting peak intensity pairs from cooperatively and non-cooperatively bound regions –

$$f(x; m, \sigma) = \frac{e^{-(\ln(x)/m)^2/2(\sigma)^2}}{x\sigma\sqrt{2\pi}}$$

$$f(x; \mu, \gamma, \beta) = \frac{(\frac{x}{\beta})^{\gamma-1} \exp\left(-\frac{x}{\beta}\right)}{\beta\Gamma(\gamma)} \qquad x \geq 0; \gamma, \beta > 0,$$

$$f(x; \mu, \sigma) = \frac{\exp\left(-(x-\mu)^2/2\sigma^2\right)}{\sigma\sqrt{2\pi}}.$$

For a set of peak intensity pairs $\{(x_i, y_i)\}_{i=1}^{N}$, we computed $\mathbf{\Theta}_X$ and $\mathbf{\Theta}_Y$ for each distribution using maximum likelihood estimators for each of the three distributions, using the `fit` routines provided in the statistics library of the Python package SciPy [1]. The maximum likelihood estimate of $\mathbf{\Theta}_X$ was calculated using only primary TF peak intensities ,i.e., $\{x_i\}_{i=1}^{N}$, while the maximum likelihood estimate of $\mathbf{\Theta}_Y$ was calculated using only partner TF peak intensities $\{y_i\}_{i=1}^{N}$. The log-likelihood values calculated for each of the three distributions across all ChIP-seq data sets is shown in Table 2.

Dataset	Primary TF					
	Cooperative			Non-cooperative		
	Log-normal	Gamma	Gaussian	Log-normal	Gamma	Gaussian
FOXA1-HNF4A	**−37367**	−37694	−39881	**−148661**	−150528	−164422
FOXA1-CEBPA	**−7413**	−7497	−8032	**−125906**	−125971	−134171
HNF4A-CEBPA	**−10360**	−10419	−10968	−145841	**−144745**	−149395
(EE) FIS-CRP	−82	**−81**	−88	**−1756**	−4064	−2063
(EE) CRP-FIS	−567	**−565**	−576	−1145	**−1144**	−1150
(ME) FIS-CRP	**−455**	−457	−518	−620	**−619**	−645
ERα-FOXA1	**−9573**	−9587	−11633	−2309	**−2087**	−2441
CDX2-HNF4A	**−39995**	−52278	−45658	**−107992**	−117804	−155132
GCN4-RTG3	**−931**	−933	−995	**−1132**	−1145	−1313
RTG3-GCN4	**−1283**	−1289	−1454	−240	**−239**	−263

Dataset	Partner TF					
	Cooperative			Non-cooperative		
	Log-normal	Gamma	Gaussian	Log-normal	Gamma	Gaussian
FOXA1-HNF4A	−46926	**−46736**	−47168	**−162580**	−162688	−170887
FOXA1-CEBPA	−9556	**−9546**	−9793	**−123278**	−123870	−132830
HNF4A-CEBPA	**−11709**	−11710	−12085	**−128582**	−129452	−139457
(EE) FIS-CRP	−124	**−114**	−128	−1640	**−1636**	−1661
(EE) CRP-FIS	**−656**	−1329	−801	**−1199**	−2699	−1402
(ME) FIS-CRP	−546	**−542**	−642	**−621**	−1127	−713
ERα-FOXA1	−11377	**−11220**	−12652	−2102	**−2069**	−2421
CDX2-HNF4A	**−57440**	−57973	−65126	**−141792**	−261112	−192010
GCN4-RTG3	**−735**	−739	−843	**−797**	−801	−901
RTG3-GCN4	**−1831**	−1880	−2233	−291	**−289**	−313

Table 2: Log-likelihood values obtained from fitting log-normal, Gaussian and Gamma distributions to cooperative and non-cooperative peak intensities of the data sets shown in Table 1. The maximum log-likelihood values are indicated in bold. Across most data sets, the log-normal distribution typically provides the best fit to peak intensity distributions.

3 Mutual information between peak intensity pairs from cooperatively and non-cooperatively bound regions

Given a probability disribution $f(x, y)$ over the set of peak intensity pairs $\{(x_i, y_i)\}_{i=1}^{N}$, mutual information (MI) is calculated as

$$MI = \sum_{i=1}^{n}\sum_{i=1}^{n} f_{XY}(x_i, y_i) \log_2 \left(\frac{f_{XY}(x_i, y_i)}{f_X(x_i) f_Y(y_i)} \right),$$

where $f_X(y)$ and $f_Y(y)$ are the marginal distributions of $f_{XY}(x, y)$, and n is the number of $\{(x_i, y_i)\}$ pairs. MI is a non-negative quantity, measured in units of bits, whose value is zero if X and Y are statistically independent i.e. if $f_{XY}(x, y) = f_X(x) f_Y(y)$. From the knockout data available for each of the TF pairs, we separated the peak intensity pairs $\{(x_i, y_i\}_{i=1}^{N}$ into a set of cooperatively bound peak intensity pairs $A_c = \{(x_j, y_j)\}$, and a set of non-cooperatively bound peak intensity pairs $A_{nc} = \{(x_k, y_k)\}$. We separately computed the MI of peak intensity pairs in A_c (setting $f = f_1$ in equation (3)) and A_{nc} (setting $f = f_0$ in equation (3)) in each ChIP-seq knockout data set we analyzed (Table 3).

When $\{x_i\}$ and $\{y_i\}$ are real numbers, which is true of peak intensities, MI becomes tricky to estimate through direct use of the definition; such MI estimates can be biased, or go to infinity in the case of certain distributions [2]. We estimated MI using the LNC algorithm implemented in [2] that circumvents these issues. The drawback of the LNC algorithm was that it gave non-negative values of MI only when a sufficient number of peaks were present. We were thus unable to reliably estimate MI in some of the ChIP-seq data sets we analyzed.

FOXA1-HNF4A	FOXA1-CEBPA	HNF4A-CEBPA	(EE) CRP-FIS	(EE) FIS-CRP
0.02, 0.04	–,0.03	0.01, 0.02	–, –	–, –

(ME) CRP-FIS	ERα-FOXA1	CDX2-HNF4A	GCN4-RTG3	RTG3-GCN4
–,–	–,–	0.29,0.09	–,0.05	0.03,0.12

Table 3: Each entry is a pair of mutual information values (in bits) between primary and partner peak intensity distributions, computed from peak intensities of cooperatively and non-cooperatively bound regions, respectively. Instances where the number of peak intensity pairs is too low for mutual information to be reliably estimated are shown as "–". The mutual information of most peak intensity pairs was very low. This makes it possible to approximate the joint density of peak intensity pairs as a product of marginal distributions.

4 Calculation of receiver operating characteristic (ROC) curves

For each ChIP-seq data set, we ran the EM algorithm to find parameters of the CPI-EM mixture model from the Q function described in equation (2) in the main text. The parameter $\mathbf{\Theta} = (\pi_0, \boldsymbol{\theta}_0^X, \boldsymbol{\theta}_1^X, \boldsymbol{\theta}_0^Y, \boldsymbol{\theta}_1^Y)$ was substituted into equation (3) in the main text to compute the probability $P(L_i = 1|x_i, y_i)$ of the i^{th} location being cooperatively bound. The ROC curve was then calculated by picking thresholds α on the probabilities $\{P(L_i = 1|x_i, y_i)\}_{i=1}^N$ that corresponded to false positive rates between 0.1 and 1 in steps of 0.1. The true positive rate at each of these thresholds was then computed. The area under the ROC curve was then calculated using the trapezoidal integration rule available in the Python `numpy` library. This procedure was repeated for each of the three variants of the CPI-EM algorithm. Similarly, the ROC curve of the peak distance algorithm was computed by choosing thresholds on the peak distance that corresopnded to false positive rates between 0.1 and 1 in steps of 0.1. After the true positive rate at each threshold was calculated, the area under the ROC was computed with the trapezoidal integration rule.

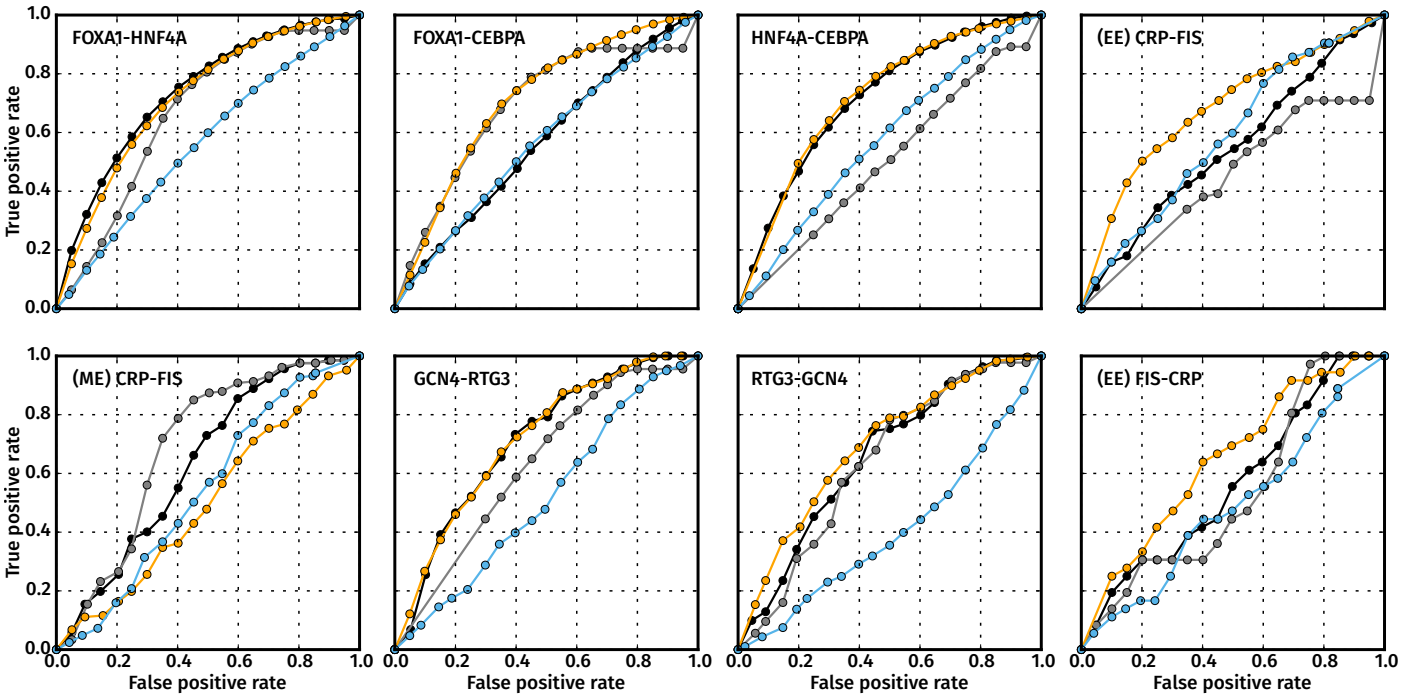

Figure 1: ROC curves of runs of the CPI-EM variants (log-normal in orange, Gamma in black, and Gaussian in gray) and peak distance (sky blue) algorithms on data sets in Figure 4 of the main text, and in Table 1 above.

5 Area under ROC of a chance detector is 0.5

The chance detector is based purely on using tosses from a biased coin to detect cooperative interactions. Let the probability of the coin showing heads be α. Out of a set of N peak intensity pairs $\{(x_i, y_i)\}$, suppose there are N_c and N_{nc} cooperatively and non-cooperatively bound pairs, respectively. The number of false positives, resulting from N tosses of the coin, would be $N_{nc}\alpha$, while the number of true positives would be $N_c\alpha$. This means that both the FPR and TPR of the chance detector would be α. Thus, as α is varied between 0 and 1, the ROC of the chance detector will be the straight line $FPR = TPR = \alpha$, which encloses an area of 0.5.

6 Q function used in Expectation Maximization Algorithm

The Q function is defined as

$$Q(\mathbf{\Theta}, \mathbf{\Theta}') = \sum_{\mathbf{L}\in S} \log\left(P\left(\mathbf{D}, \mathbf{L}|\mathbf{\Theta}\right)\right) P(\mathbf{L}|\mathbf{D}, \mathbf{\Theta}') = \sum_{\mathbf{L}\in S} \log\left(P\left(\mathbf{L}|\mathbf{D}, \mathbf{\Theta}\right) P(\mathbf{D}|\mathbf{\Theta})\right) P(\mathbf{L}|\mathbf{D}, \mathbf{\Theta}') \tag{1}$$

Since a peak intensity pair is either cooperative or non-cooperative, we can write $P(X_i = x_i, Y_i = y_i) = \pi_0 f_0(x_i, y_i; \boldsymbol{\theta}_0) + \pi_1 f_1(x_i, y_i; \boldsymbol{\theta}_1)$, where $\pi_0 + \pi_1 = 1$. Since we consider (X_i, Y_i) and (X_j, Y_j) $(i \neq j)$ to be statistically independent,

$$\log P(\mathbf{D}|\mathbf{\Theta}) = \sum_{i=1}^{N} \log\left(\pi_0 f_0(x_i; \boldsymbol{\theta}_0^X) f_0(y_i; \boldsymbol{\theta}_0^Y) + \pi_1 f_1(x_i; \boldsymbol{\theta}_1^X) f(y_i; \boldsymbol{\theta}_1^Y)\right)$$

$P(\mathbf{L}|\mathbf{D},\mathbf{\Theta})$ in equation (1) is –

$$\begin{aligned} P(\mathbf{L}|\mathbf{D},\mathbf{\Theta}) &= \prod_{i=1}^{N} P(L_i = l_i|\mathbf{D},\mathbf{\Theta}) \\ &= \prod_{i=1}^{N} \frac{\pi_{l_i} f_{l_i}(x_i, y_i; \boldsymbol{\theta}_{l_i})}{\pi_0 f_0(x_i, y_i; \boldsymbol{\theta}_0) + \pi_1 f_1(x_i, y_i; \boldsymbol{\theta}_1)}, \end{aligned}$$

where l_i is 0 or 1. Substituting the above two expansions into the expression for Q in equation (1), it can be shown that Q simplifies to the form shown below, where it is a sum over only N terms (page 4 in [3])

$$\begin{aligned} Q(\mathbf{\Theta},\mathbf{\Theta'}) &= \sum_{i=1}^{N}\sum_{l=1}^{2} \log(\pi_l) P(L_i = l|x_i, y_i, \mathbf{\Theta'}) + \sum_{i=1}^{N}\sum_{l=1}^{2} P\big(L_i = l|x_i, y_i, \mathbf{\Theta}\big) \log\Big(f_l(x_i; \boldsymbol{\theta}_l^{X'}) f_l^Y(y_i; \boldsymbol{\theta}_l^{Y'})\Big) \\ &= Q_1(\mathbf{\Theta'}) + Q_2(\mathbf{\Theta},\mathbf{\Theta'}), \end{aligned} \tag{2}$$

where, $\mathbf{\Theta'} = (\pi_0', \boldsymbol{\theta}_0^{X'}, \boldsymbol{\theta}_0^{Y'}, \boldsymbol{\theta}_1^{X'}, \boldsymbol{\theta}_1^{Y'})$. Note that the first term is independent of $\mathbf{\Theta'}$, so it can be maximized independently of the second term. The choice of π_l that maximizes the first term (page 5 in [3]) is –

$$\pi_l = \frac{1}{N}\sum_{i=1}^{N} P(L_i = l|x_i, y_i, \mathbf{\Theta'}) \text{ for } l = 0, 1.$$

The $k-th$ EM iteration involves choosing a value $\mathbf{\Theta} = \mathbf{\Theta}^{(k+1)}$ that maximizes $Q(\mathbf{\Theta}, \mathbf{\Theta}^{(k)})$, where $\mathbf{\Theta}^{(k)}$ is kept fixed at the value obtained in the previous EM iteration that maximizes $Q(\mathbf{\Theta}, \mathbf{\Theta}^{(k-1)})$. EM involves two steps, an E-step and an M-step, which are both needed to maximize $Q(\mathbf{\Theta}, \mathbf{\Theta}^{(k)})$. The E- and M- steps evaluated at the $i-th$ iteration in our algorithm are [4] –

- E-step : Compute

$$P(L_i = l|x_i, y_i, \mathbf{\Theta}^{(k)}) = \frac{{\pi_l}^{(k)} f_l(x_i, y_i; \mathbf{\Theta}^{(k)})}{{\pi_0}^{(k)} f_0(x_i, y_i; \mathbf{\Theta}^{(k)} + {\pi_1}^{(k)} f_1(x_i, y_i; \mathbf{\Theta}^{(k)}))} \text{ for } l = 0, 1;\ i = 1, 2, \ldots, N.$$

- M-step (1) : Compute

$$\pi_l^{(k+1)} = \sum_{i=1}^{N} P(L_i = l|x_i, y_i, \mathbf{\Theta}^{(k)}) \text{ for } l = 0, 1$$

  This step maximizes Q_1 in equation (2).

- M-step (2) : Use Powell's gradient search method, as implemented in the Scipy optimization toolbox [5] to find $\mathbf{\Theta}^{(k+1)}$ that maximizes $Q_2(\mathbf{\Theta}, \mathbf{\Theta}^{(k)})$ in equation (2), with $\mathbf{\Theta}^{(k)} = (\pi_0^k, \boldsymbol{\theta}_0^{X,(k)}, \boldsymbol{\theta}_0^{Y,(k)}, \boldsymbol{\theta}_1^{X,(k)}, \boldsymbol{\theta}_1^{Y,(k)})$ kept constant –

$$Q_2(\mathbf{\Theta}, \mathbf{\Theta}^{(k)}) = \sum_{i=1}^{N}\sum_{l=1}^{2} P\big(L_i = l|x_i, y_i, \mathbf{\Theta}\big) \log\Big(f_l(x_i; \boldsymbol{\theta}_l^{X,(k)}) f_l^Y(y_i; \boldsymbol{\theta}_l^{Y,(k)})\Big)$$

References

[1] Eric Jones, Travis Oliphant, and Pearu Peterson. {Scipy}: open source scientific tools for {Python}. 2014.

[2] Shuyang Gao, Greg Ver Steeg, and Aram Galstyan. Efficient estimation of mutual information for strongly dependent variables. In *AISTATS*, 2015.

[3] Jeff A. Bilmes and others A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models *International Computer Science Institute*, 510(4):126, 1998.

[4] Ioannis Kosmidis and Dimitris Karlis. Model-based clustering using copulas with applications. *Statistics and Computing*, 1–21, 2015.

[5] MJD Powell. Direct search algorithms for optimization calculations. *Acta numerica*, pages 287–336, 1998.