Supplementary Information for:

Inferring chromatin-bound protein complexes from genome-wide binding assays

Eugenia G. Giannopoulou[1,2] and Olivier Elemento[1,2]

[1]HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, Cornell University, 1305 York Avenue, New York, NY 10021, USA

[2]Department of Physiology and Biophysics, Weill Cornell Medical College, Cornell University, 1300 York Avenue, New York, NY, 10021, USA

Overview of Supplementary Information:

Supplementary Notes

1. NMF after filtering out single TF-bound CREs.
2. Large protein complexes discovered at low rank factorization.

Supplementary Tables

1. List of ENCODE ChIP-seq datasets for three cell types.
2. Summary of CREs discovered in three human cell lines.
3. Additional data for GeneMANIA physical interactions.
4. The predicted complexes with the highest contribution in gene expression.
5. Physical interactions between the TFs of the low NMF complexes of Suppl. Figure 15.

Supplementary Figures

1. Quantitative measures estimated for multiple NMF runs using different values of k.
2. Consensus matrices for the H1 ESC dataset.
3. Consensus matrices for the GM12878 dataset.
4. Consensus matrices for the K562 dataset.
5. Histograms of the number of TFs bound at the CREs in each of the three human cell lines.
6. Filtering of the RC matrix.
7. Protein complexes revealed after filtering of the RC matrix.
8. Regulatory motifs analysis.
9. The contribution of complexes to gene expression in the linear and nonlinear regression models.
10. Shuffling of the RC matrix.
11. NMF runs with multiple algorithms for each k.
12. The exponential function in the BIS score formula is affected by d_0.
13. Linear regression coefficients of the complexes after changing the d_0 parameter.

14. Importance measures for the complexes-predictors in random forests regression.
15. Protein complexes revealed at low rank factorization.
16. Physical interaction networks for the low NMF protein complexes.

Supplementary Notes

1. NMF after filtering out single TF-bound CREs

We repeated the NMF analysis described in the main text (using the same NMF parameters) after filtering out the CREs that are bound only by one TF, thus keeping only the CREs bound by at least 2 TFs (**Supplementary Fig. 6**), and sought to investigate the complexes discovered after this filtering. The heatmaps displaying the coefficients matrices of the best fit (after performing 100 NMF runs) for the filtered H1 ESC, GM12878 and K562 datasets are shown in **Supplementary Figure 7**. We observed that 82% (14/17), 80% (16/20) and 88% (15/17) of the complexes discovered after filtering for the H1 ESC, GM12878 and K562 datasets are highly similar to the ones coming from the non-filtered datasets (Figure 3a). We compared the coefficients of each complex from the filtered datasets to the coefficients of every complex from the non-filtered datasets, identifying best reciprocal matches using the Pearson correlation. We observed that 82% (14/17), 80% (16/20) and 88% (15/17) of the complexes discovered when not applying any filtering were also present (Pearson > 0.92) in the filtered datasets for the H1 ESC, GM12878 and K562 datasets, respectively (**Supplementary Figure 7**). Therefore, keeping the CREs that are bound by only 1 TF does not strongly impact the detection of TF protein complexes. Since we envision that TFs may either bind to chromatin alone in specific contexts (our analysis allows a TF to participate in multiple contexts) or with other co-factors not assayed by ChIP-seq, we did not filter out in CREs bound by only 1 TF in the present analysis.

2. Large protein complexes discovered at low rank NMF

The heatmaps in **Supplementary Figure 15** show the coefficient matrices of NMF runs discovering 6, 4, and 4 for H1 ESC, GM12878, and K562 respectively. Similarly with the factorization at higher levels, we observed that complexes associated with transcription initiation or promoter marks (e.g., *H3K4me2, H3K4me3, H3K79me2*) are preferentially located within promoters (H-5, G-2, K-1 in **Supplementary Fig. 15b**), compared to complexes associated with enhancer marks (e.g., *EP300, H3K4me1*), which have higher presence in distal and intergenic regions (G-4, K-2 in **Supplementary Fig. 15b**).

Motifs analysis was performed with FIRE, to search for over-represented DNA motifs within the complex-specific CREs in order to identify sequence-specific TFs likely to target the complexes to the chromatin. We found the *REST* motifs in complexes where *REST* is present with a high coefficient (i.e., H-3, G-3, K-3), as well as the *CTCF* motif in complexes with binding of *CTCF-RAD21* (i.e., H-4, G-1, K-4), suggesting that *REST* and *CTCF* directly bind the DNA and recruit the rest of their partners. A sequence resembling the *SPIB* motif of the *ETS* family was found in the CREs of the K-2 predicted complex, which indeed contains several *ETS* family proteins (i.e., *ELF1*, *ETS1*, *GABPA*). Moreover, the AP-1 and AML-1A motifs (*RUNX* family) were enriched in the CREs of G4, as expected by the presence of *ATF2* (part of the AP-1 complex) and *RUNX3* in that complex (**Supplementary Fig. 15c**).

Additionally, we sought to search for known physical interactions from GeneMANIA (http://genemania.org/data/) within the factors of each complex. Not surprisingly, most complexes had real physical interactions between their members (**Supplementary Table 5**). Importantly, we reproduced same size complexes with the predicted ones by randomly combining genes from the pool of the 64 TFs studied in this work (**Supplementary Table 1**), and found less physical interactions occurring, suggesting that there are more physical interactions between the TFs of each of the NMF predicted complexes than could be expected by chance (**Supplementary Table 5**). The physical interaction networks of the corresponding complexes are shown in **Supplementary Fig. 16.**

Supplementary Table 1. List of ENCODE ChIP-seq datasets for three cell types. Letters correspond to the lab producing the data: B for Bernstein - Broad Institute, I for Iyer – UT Austin, M for Myers - Hudson Alpha. All datasets used comply to the ENCODE Consortium Data Release Policy Summary at the time of submission.

	CELL TYPES					
Experiment Name	**H1 ESC**		**GM12878**		**K562**	
ATF2	✓	M	✓	M		
ATF3	✓	M	✓	M	✓	M
BATF			✓	M		
BCLAF1			✓	M	✓	M
BCL11A	✓	M	✓	M		
BCL3			✓	M	✓	M
CHD1	✓	B			✓	B
CTCF	✓	M, B	✓	B	✓	M, B
E2F6					✓	M
EBF1			✓	M		
EGR1	✓	M	✓	M	✓	M
ELF1			✓	M	✓	M
EP300	✓	M	✓	M	✓	B
ETS1			✓	M	✓	M
EZH2					✓	B
FOSL1	✓	M			✓	M
GABP	✓	M	✓	M	✓	M
GATA2					✓	M
HDAC1					✓	B
HDAC2	✓	M			✓	M, B
HEY1					✓	M
H2AFZ	✓	B	✓	B	✓	B
IRF4			✓	M		
JUND	✓	M				
KDM5A (JARID1A)	✓	B				
KDM5B (PLU1)					✓	B
MAX					✓	M
MEF2A			✓	M	✓	M
MEF2C			✓	M		
MTA3			✓	M		
MYC	✓	I				
NANOG	✓	M				
NFATC1			✓	M		
PAX5			✓	M		
PBX3			✓	M		
PHF8					✓	B
PML			✓	M	✓	M

Supplemental Table 1. (continued)

	Cell Types					
Experiment Name	**H1 ESC**		**GM12878**		**K562**	
POU2F2			✓	M		
POU5F1	✓	M				
RAD21	✓	M	✓	M	✓	M
RBBP5	✓	B			✓	B
REST (NRSF)	✓	M	✓	M	✓	M
RUNX3			✓	M		
RXRA	✓	M	✓	M		
SAP30					✓	B
SIN3A	✓	M			✓	M
SIX5	✓	M	✓	M	✓	M
SPI1 (PU1)			✓	M	✓	M
SP1	✓	M	✓	M	✓	M
SP2	✓	M			✓	M
SP4	✓	M				
SRF	✓	M	✓	M	✓	M
STAT5A			✓	M		
TAF1	✓	M	✓	M	✓	M
TAF7	✓	M			✓	M
TCF12	✓	M	✓	M		
TCF3			✓	M		
TEAD4					✓	M
THAP1					✓	M
USF1	✓	M	✓	M	✓	M
YY1	✓	M	✓	M	✓	M
ZBTB33			✓	M	✓	M
ZBTB7A					✓	M
ZEB1			✓	M		
H3K27ac	✓	B	✓	B	✓	B
H3K27me3	✓	B	✓	B	✓	B
H3K36me3	✓	B	✓	B	✓	B
H3K4me1	✓	B	✓	B	✓	B
H3K4me2	✓	B	✓	B	✓	B
H3K4me3	✓	B	✓	B	✓	B
H3K79me2	✓	B	✓	B	✓	B
H3K9ac	✓	B	✓	B	✓	B
H3K9me1					✓	B
H3K9me3	✓	B	✓	B	✓	B
H4K20me1	✓	B	✓	B	✓	B

Supplementary Table 2. Summary of CREs discovered in three human cell lines. Based on a large collection of ChIP-seq datasets more than 100.000 regulatory elements are identified in each cell type, covering 7-8% of the human genome.

Cell Types	Number of regulatory elements	Number of nucleotides	Percentage of the genome (%)	ChIP-seq	
				TFs	HMs
H1 ESC	107,328	208,692,978	6.7	30	11
GM12878	100,798	255,653,490	8.25	38	11
K562	109,492	245,269,528	7.9	40	12

Supplementary Table 3. Additional data for GeneMANIA physical interactions. The table provides additional information for the physical interactions shown in Table 1: the sources of interactions (e.g., BioGRID, publications) available in http://genemania.org/data/, and the GeneMANIA weight scores between two genes.

Cell Type (#complexes)	Predicted Complex ID (#interactions)	Physical interaction	GeneMANIA scores[1]	Source of interaction (Interaction network files)
H1 ESC (17)	H-14 (2)	*JUND-ATF3*	1.10E-01 9.60E-02 1.30E-01 7.30E-02 4.80E-02 8.20E-02 1.00E-01	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt Physical_interactions.Newman-Keating-2003.txt
		JUND-FOSL1	9.20E-02 5.00E-02 9.90E-02	Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.Newman-Keating-2003.txt
	H-15 (1)	*EP300-TCF12*	8.20E-03 9.50E-02 7.40E-03 2.40E-02 5.30E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt
GM12878 (20)	G-4 (3)	*EP300-TCF12*	8.20E-03 9.50E-02 7.40E-03 2.40E-02 5.30E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt
		TCF3-EP300	4.90E-03 7.40E-03 3.50E-03 7.60E-03 2.80E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt
		TCF3-TCF12	2.30E-02 4.60E-02 1.60E-02	Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt
	G-10 (1)	*IRF4-BATF*	1.80E-01 1.30E-01 5.40E-01	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.Ravasi-Hayashizaki-2010.txt
K562 (17)	K-3 (9)	*BCL3-EP300*	8.40E-03 1.20E-02 5.60E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt
		BCL3-HDAC1	6.60E-03 6.50E-02 1.10E-02 6.40E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt
		EP300-HDAC1	1.00E-03 3.90E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-OPHID.txt
		EP300-SP1	2.20E-03 3.40E-03 2.40E-03 1.70E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt
		EZH2-HDAC1	4.60E-03 5.90E-03 7.70E-03 1.20E-02 5.70E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt
		HDAC1-CHD1	1.20E-02 1.50E-02 2.60E-02 3.90E-02 1.40E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt
		HDAC1-SAP30	5.40E-03 9.80E-03 1.10E-02 2.50E-02 6.20E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt
		HDAC1-SP1	1.70E-03 3.00E-03 3.60E-03 5.90E-02 1.90E-01 5.60E-03 2.00E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-MINT.txt Physical_interactions.IREF-MPPI.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt
		TAF7-SP1	8.80E-02 2.80E-02 1.10E-02	Physical_interactions.IREF-DIP.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt
	K-9 (1)	*PML-GATA2*	2.40E-02 3.20E-02 3.10E-02 1.00E-01 1.90E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt
	K-15 (1)	*SRF-SIX5*	2.80E-01	Physical_interactions.IREF-OPHID.txt

[1] Note that these scores do not contribute to p-value calculation.

Supplementary Table 4. The predicted complexes with the highest contribution in gene expression. The complexes shown in **Figure 3**, with the most significant estimated coefficients measured by linear regression (denoted with *** in **Fig. 3d**) and with more that ~40% increase in mean squared error (MSE) are shown in the table. For each complex, the most contributing factors and histone modifications (coefficient > 0.3) are shown.

Cell Types	Predicted Complex ID	TFs		HMs	
H1 ESC	H-1	*RAD21*		-	
	H-4	*TAF1*		-	
	H-8	*CHD1* *MYC*		*H3K79me2* *H3K9me3* *H3K9ac*	*H3K36me3* *H4K20me1* *H3K27ac*
	H-9	-		*H3K27me3*	
GM12878	G-1	*CTCF*		-	
	G-12	-		*H3K79me2* *H3K9ac* *H3K27ac*	
	G-17	*REST*		*H3K27me3*	
K562	K-2	-		*H3K79me2*	
	K-3	*BCL3* *CHD1* *EP300* *EZH2* *HDAC1*	*SAP30* *SP1* *SP2* *TAF7* *ZBTB33*	*H3K9me3* *H3K27me3* *H3K36me3*	*H3K9me1* *H4K20me1*
	K-7	*CTCF*		-	
	K-9	*GATA2* *FOSL1* *MEF2A*	*PML* *TEAD4*	-	
	K-11	-		*H3K9ac* *H3K4me2* *H3K4me3* *H3K27ac*	
	K-14	*HEY1* *KDB5B* *PHF8*	*RBBP5* *TAF1*	-	

Supplementary Table 5. Physical interactions between the TFs of the low NMF complexes of Suppl. Figure 15. The sources of interactions (e.g., BioGRID, iRef) and GeneMANIA scores between two genes are shown. p-values were calculated after estimating the interactions that occur overall in 10,000 sets of complexes with random genes. The GeneMANIA scores do not contribute to p-value calculation.

Cell Type (#complexes)	Predicted Complex ID (#interactions)	Physical interaction	GeneMANIA scores	Source of interaction (Interaction network files)	p-value
H1 ESC (6)	H-1 (22)	*ATF2-FOSL1*	7.50E-02 4.60E-02 8.90E-02	Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.Newman-Keating-2003.txt	0.063
		ATF2-JUND	5.20E-02 2.70E-02 7.20E-02	Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.Newman-Keating-2003.txt	
		TAF7-SP1	8.80E-02 2.80E-02 1.10E-02	Physical_interactions.IREF-DIP.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		GABPA-SP1	3.50E-02 4.20E-02 3.90E-02 2.20E-01 7.60E-02 2.50E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-MINT.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		EGR1-SP1	1.90E-02 1.40E-01 2.70E-02 1.30E-02	Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-MINT.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		SRF-RXRA	6.80E-03 1.20E-02 5.80E-03	Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		SRF-SP1	7.00E-03 8.70E-03 4.70E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		SRF-SIX5	2.80E-01	Physical_interactions.IREF-OPHID.txt	
		JUND-FOSL1	9.20E-02 5.00E-02 9.90E-02	Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.Newman-Keating-2003.txt	
		SP1-RXRA	5.40E-03 7.00E-03 5.20E-03 1.70E-01 9.00E-03 4.00E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-MPPI.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		SP1-HDAC2	2.60E-03 4.40E-03 5.90E-03 9.40E-03 2.90E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		EP300-ATF2	9.90E-03	Physical_interactions.IREF-OPHID.txt	
		EP300-GABPA	1.80E-02 1.30E-02	Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		EP300-TCF12	8.20E-03 9.50E-02 7.40E-03 2.40E-02 5.30E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		EP300-EGR1	1.00E-02 1.20E-01 6.20E-01 1.30E-02 9.00E-03 1.90E-02 7.00E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-CORUM.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		EP300-JUND	1.00E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt	
		EP300-SP1	2.20E-03 3.40E-03 2.40E-03 1.70E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		EP300-YY1	4.30E-03 2.00E-01 8.50E-03 7.40E-03 1.20E-02 4.30E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-DIP.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		YY1-ATF2	2.10E-02 6.10E-02 3.50E-02 3.00E-02 2.10E-02 5.20E-01	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt Physical_interactions.Ravasi-Hayashizaki-2010.txt	
		YY1-TAF7	1.50E-01 5.90E-02 2.70E-02	Physical_interactions.IREF-DIP.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		YY1-SP1	7.00E-03 1.30E-01 1.40E-01 1.30E-02 1.60E-02 1.80E-02 8.10E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-DIP.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		YY1-HDAC2	5.20E-03 9.80E-02 1.10E-02 1.80E-02 2.00E-02 7.30E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	

Supplemental Table 5. (continued)

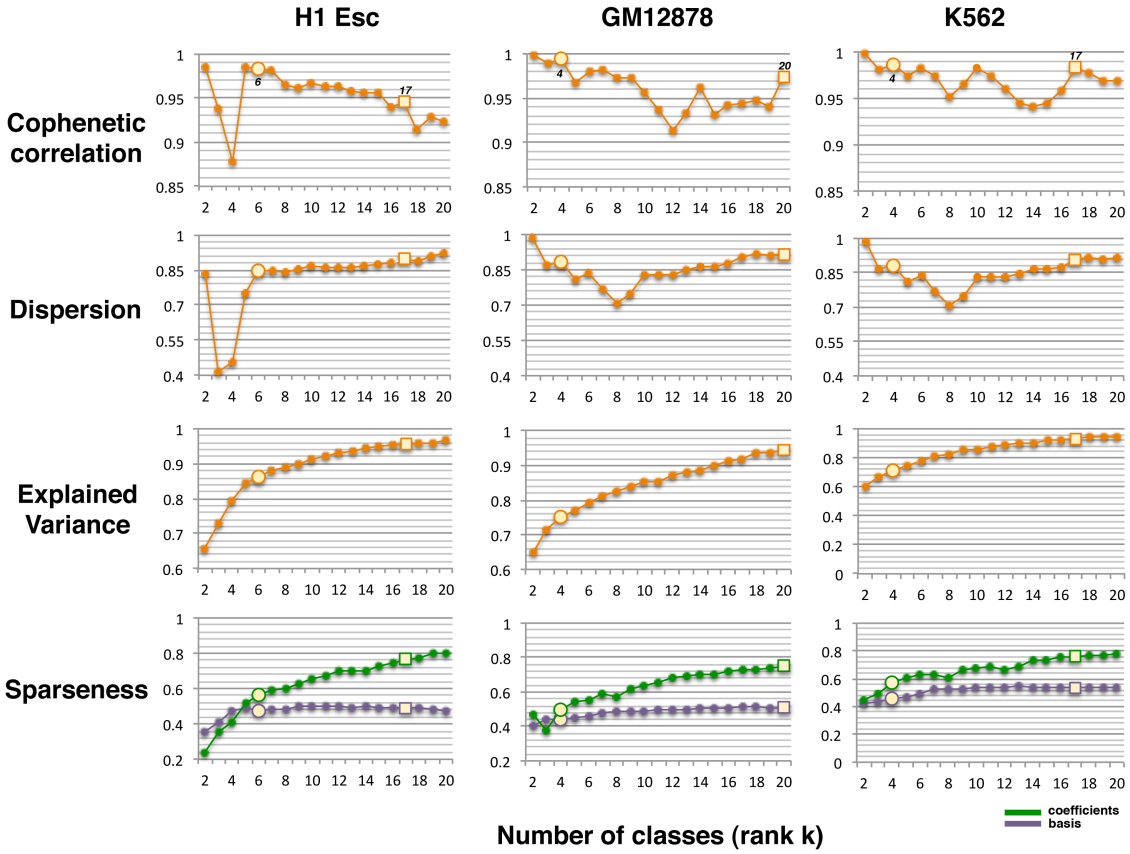
Cell Type (#complexes)	Predicted Complex ID (#interactions)	Physical interaction	GeneMANIA scores	Source of interaction (Interaction network files)	p-value
H1 ESC (6)	H-4 (1)	*CTCF-RAD21*	1.30E-01 7.10E-01 6.20E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-MINT.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
	H-5 (2)	*TAF1-TAF7*	5.20E-02 1.30E-01 6.20E-02 3.70E-02 4.50E-02 2.70E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-DIP.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		TAF1-SIN3A	8.50E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt	
GM12878 (4)	G-1 (1)	*CTCF-RAD21*	1.30E-01 7.10E-01 6.20E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-MINT.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	0.045
	G-3 (12)	*ETS1-PAX5*	2.40E-01 5.40E-02 1.10E-01 5.30E-02	Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		ETS1-SP1	1.20E-02 9.30E-03	Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		EGR1-SP1	1.90E-02 1.40E-01 2.70E-02 1.30E-02	Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-MINT.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		REST-SP1	1.90E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt	
		SRF-SIX5	2.80E-01	Physical_interactions.IREF-OPHID.txt	
		SRF-SP1	7.00E-03 8.70E-03 4.70E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		SRF-ELF1	9.30E-02	Physical_interactions.IREF-OPHID.txt	
		ELF1-SP1	4.30E-02 4.90E-02 3.30E-02 6.90E-02 6.90E-02 2.50E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-INTACT.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		TCF3-TCF12	2.30E-02 4.60E-02 1.60E-02	Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		TCF3-SRF	9.50E-03 7.70E-03	Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		GABPA-SP1	3.50E-02 4.20E-02 3.90E-02 2.20E-01 7.60E-02 2.50E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-MINT.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		YY1-SP1	7.00E-03 1.30E-01 1.40E-01 1.30E-02 1.60E-02 1.80E-02 8.10E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-DIP.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
	G-4 (18)	*ATF2-BATF*	4.50E-02 3.60E-02 5.40E-02	Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.Newman-Keating-2003.txt	
		BCL3-EP300	8.40E-03 1.20E-02 5.60E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		EBF1-PAX5	2.00E-01 2.20E-01	Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		EP300-ATF2	9.90E-03	Physical_interactions.IREF-OPHID.txt	
		EP300-EBF1	5.70E-02	Physical_interactions.IREF-OPHID.txt	
		EP300-NFATC1	1.80E-02 1.20E-02	Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		EP300-TCF12	8.20E-03 9.50E-02 7.40E-03 2.40E-02 5.30E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		EP300-STAT5A	4.00E-03 3.90E-03	Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		EP300-SP1	2.20E-03 3.40E-03 2.40E-03 1.70E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		IRF4-BATF	1.80E-01 1.30E-01 5.40E-01	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.Ravasi-Hayashizaki-2010.txt	
		MEF2A-EP300	1.00E-02 1.30E-02 8.30E-03 1.70E-02 6.00E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		MEF2C-EP300	7.30E-03 1.10E-02 6.90E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt	

Supplemental Table 5. (continued)

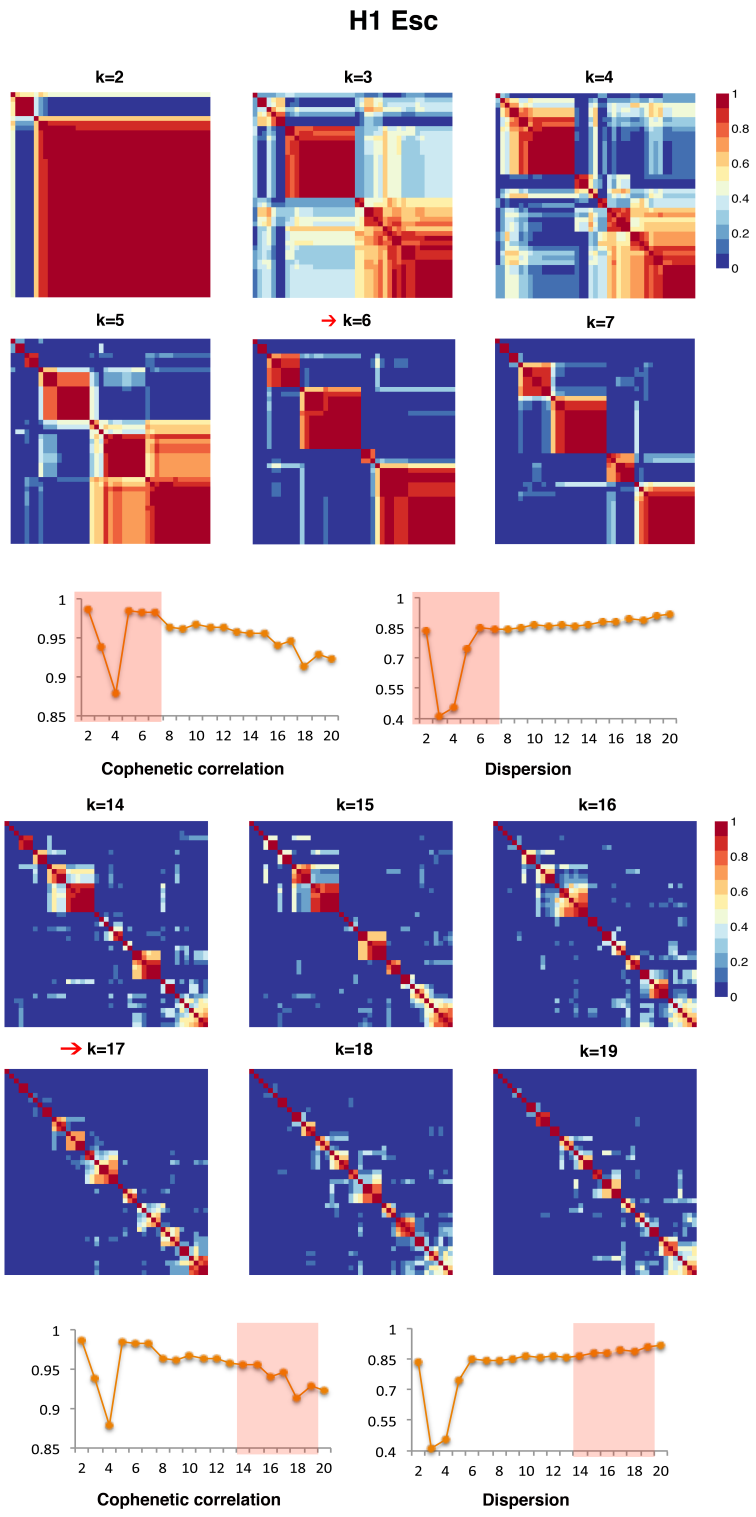
Cell Type (#complexes)	Predicted Complex ID (#interactions)	Physical interaction	GeneMANIA scores	Source of interaction (Interaction network files)	p-value
GM12878 (4)	G-4 (18)	*MEF2C-SP1*	1.20E-02 1.70E-02 1.50E-02 2.00E-02 1.20E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		SPI1-IRF4	5.30E-02 5.10E-02 6.30E-02 1.10E-01 6.20E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		SPI1-NFATC1	9.00E-02 5.30E-02	Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		RUNX3-EP300	1.20E-02 1.50E-02 8.60E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		TCF3-EP300	4.90E-03 7.40E-03 3.50E-03 7.60E-03 2.80E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		TCF3-TCF12	2.30E-02 4.60E-02 1.60E-02	Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
K562 (4)	K-1 (5)	*HDAC1-CHD1*	1.20E-02 1.50E-02 2.60E-02 3.90E-02 1.40E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	0.024
		HDAC1-SAP30	5.40E-03 9.80E-03 1.10E-02 2.50E-02 6.20E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		HDAC1-ZBTB7A	1.30E-02 1.70E-02 1.20E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		YY1-HDAC1	3.30E-03 7.40E-03 4.90E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		YY1-SAP30	2.30E-02 4.40E-02 4.60E-02 7.90E-02 2.60E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
	K-2 (23)	*HDAC1-CHD1*	1.20E-02 1.50E-02 2.60E-02 3.90E-02 1.40E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		HDAC1-SP1	1.70E-03 3.00E-03 3.60E-03 5.90E-02 1.90E-01 5.60E-03 2.00E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-MINT.txt Physical_interactions.IREF-MPPI.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		HDAC1-GATA2	1.40E-01 8.00E-03	Physical_interactions.IREF-DIP.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		HDAC1-FOSL1	2.00E-02 1.80E-02 2.00E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		HDAC1-PML	2.40E-03 4.20E-03 5.50E-03 1.10E-02 2.80E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		HDAC1-HDAC2	1.20E-03 2.40E-03 4.20E-03 4.00E-02 6.30E-03 1.80E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-INTACT.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		SP1-HDAC2	2.60E-03 4.40E-03 5.90E-03 9.40E-03 2.90E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		SPI1-HDAC1	5.40E-03 7.60E-03 4.80E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		SPI1-GATA2	5.40E-02 2.00E-01 5.70E-02 4.70E-02 1.50E-01 3.20E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-DIP.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		EP300-HDAC1	1.00E-03 3.90E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-OPHID.txt	
		EP300-SP1	2.20E-03 3.40E-03 2.40E-03 1.70E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	

Supplemental Table 5. (continued)

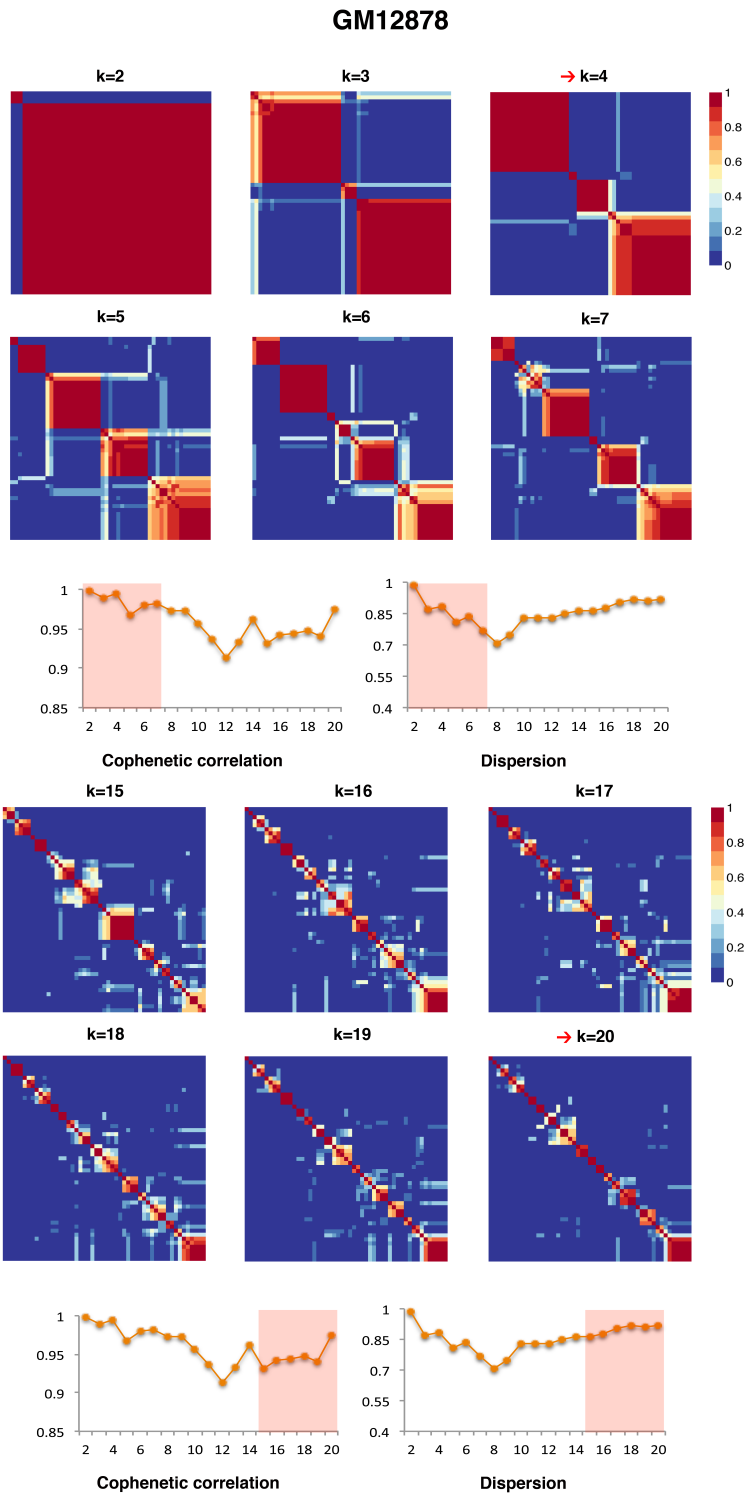
Cell Type (#complexes)	Predicted Complex ID (#interactions)	Physical interaction	GeneMANIA scores	Source of interaction (Interaction network files)	p-value
K562 (4)	K-2 (23)	*EP300-GATA2*	1.00E-02 3.60E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-OPHID.txt	
		EP300-PML	3.60E-03 2.50E-03	Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		TAF7-SP1	8.80E-02 2.80E-02 1.10E-02	Physical_interactions.IREF-DIP.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		BCL3-HDAC1	6.60E-03 6.50E-02 1.10E-02 6.40E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		BCL3-EP300	8.40E-03 1.20E-02 5.60E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		EZH2-HDAC1	4.60E-03 5.90E-03 7.70E-03 1.20E-02 5.70E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		EZH2-HDAC2	7.10E-03 8.60E-03 8.50E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		MEF2A-EP300	1.00E-02 1.30E-02 8.30E-03 1.70E-02 6.00E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		PML-SP1	5.00E-03 7.70E-03 7.70E-03 1.60E-02 4.70E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		SP1-GATA2	2.40E-02 3.20E-02 3.10E-02 1.00E-01 1.90E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		SP1-HDAC2	3.70E-03 6.20E-03 9.00E-03 4.20E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		BCLAF1-CHD1	3.10E-01 3.80E-01 2.00E-01	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
	K-3 (8)	*ZBTB7A-HDAC2*	2.00E-02 2.50E-02 1.80E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		REST-SIN3A	1.50E-02 2.90E-02 1.90E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		REST-HDAC2	1.40E-02 2.40E-02 1.70E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		USF1-FOSL1	2.40E-01 1.50E-01 2.50E-01 2.80E-01 1.80E-01 1.50E-01	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-DIP.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		SIN3A-ZBTB7A	2.20E-02 3.00E-02 2.00E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		SIN3A-HDAC2	2.10E-03 1.30E-01 4.10E-03 7.10E-03 5.70E-02 1.40E-02 3.00E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-INTACT.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		YY1-MAX	1.50E-02 2.30E-02 2.00E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
		YY1-HDAC2	5.20E-03 9.80E-02 1.10E-02 1.80E-02 2.00E-02 7.30E-03	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-BIND.txt Physical_interactions.IREF-GRID.txt Physical_interactions.IREF-HPRD.txt Physical_interactions.IREF-OPHID.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	
	K-4 (1)	*CTCF-RAD21*	1.30E-01 7.10E-01 6.20E-02	Physical_interactions.BIOGRID-SMALL-SCALE-STUDIES.txt Physical_interactions.IREF-MINT.txt Physical_interactions.IREF-SMALL-SCALE-STUDIES.txt	



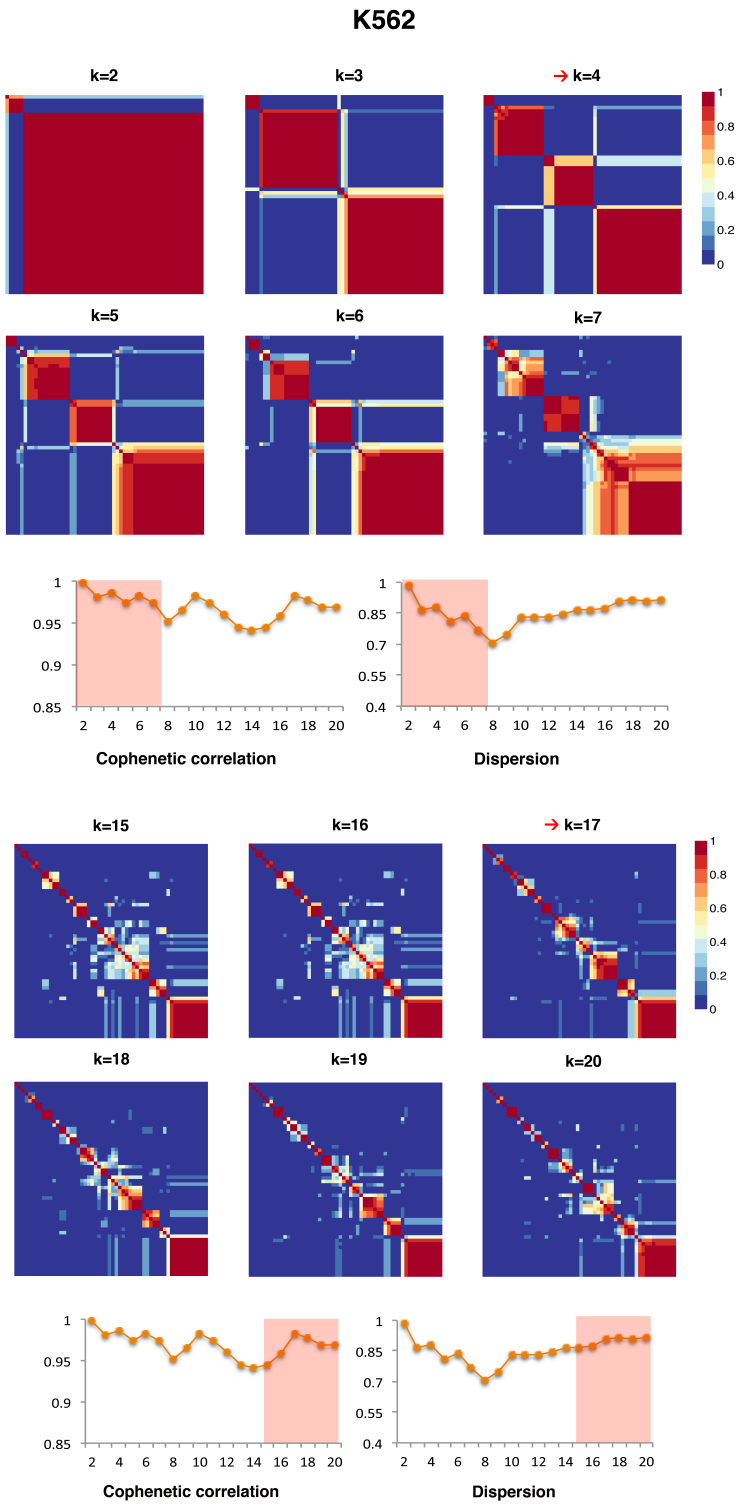
Supplementary Figure 1. Quantitative measures estimated for multiple NMF runs using different values of rank *k*. The *cophenetic correlation* and the *dispersion* coefficients are both based on the consensus matrix (i.e., the average of connectivity matrices) and measure the stability and reproducibility of the clusters obtained from NMF for a certain value of *k*, respectively. The *explained variance* measure evaluates how well the NMF model reconstructs the original data. The *sparseness* measure, for both the basis and mixture coefficient matrices, shows whether an NMF representation encodes much of the data using only few components, which makes the encoding easy to interpret. The large empty circles and squares in each graph correspond to the chosen levels of k for the low and high NMF runs discussed in the main text (Figure 3) and in Supplementary material (Supplementary Figure 15) respectively.



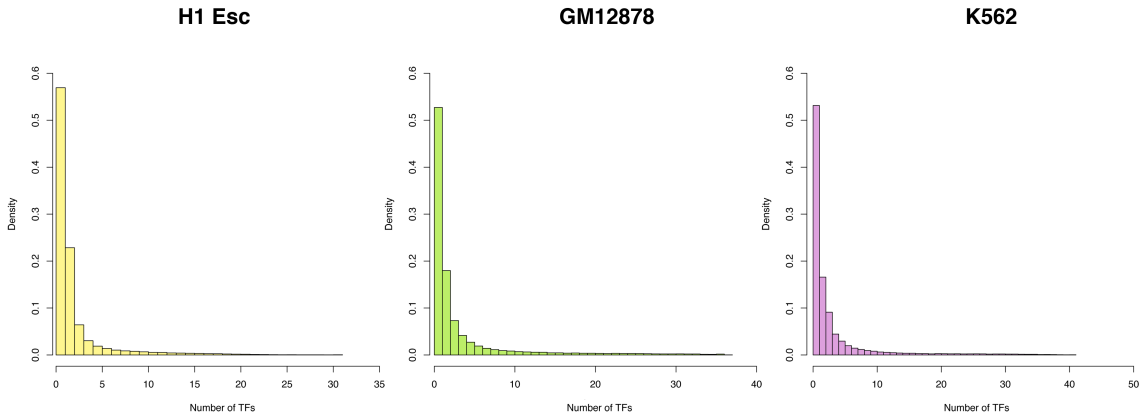
Supplementary Figure 2. Consensus matrices for the H1 Esc dataset. Consensus matrices averaging 100 connectivity matrices at k=2,3,4,5,6,7 and k=14,15,16,17,18,19. The cophenetic correlation and dispersion qualitative measures for the corresponding ranks are highlighted in the respective plots. The red arrows indicate the ranks that were selected for the "low-rank" NMF (Supplementary text, Supplementary Figure 15) and the "high-rank" NMF (main text, Figure 3).



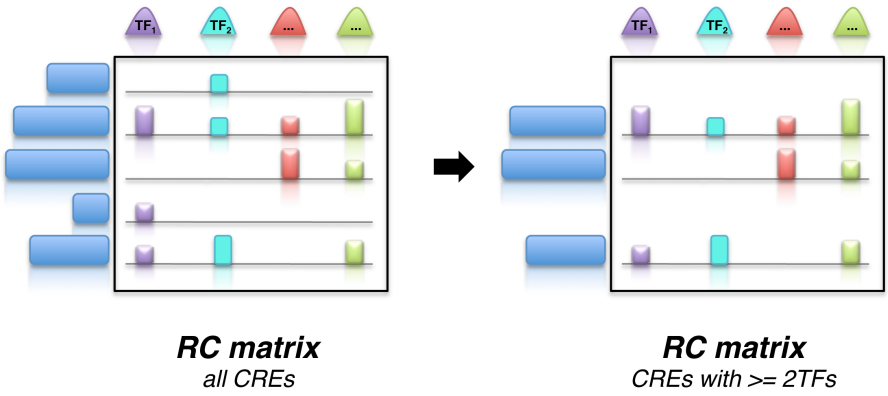
Supplementary Figure 3. Consensus matrices for the GM12878 dataset. Consensus matrices averaging 100 connectivity matrices at k=2,3,4,5,6,7 and k=15,16,17,18,19,20. The cophenetic correlation and dispersion qualitative measures for the corresponding ranks are highlighted in the respective plots. The red arrows indicate the rank that were selected for the the "low-rank" NMF (Supplementary text, Supplementary Figure 15) and the "high-rank" NMF (main text, Figure 3).



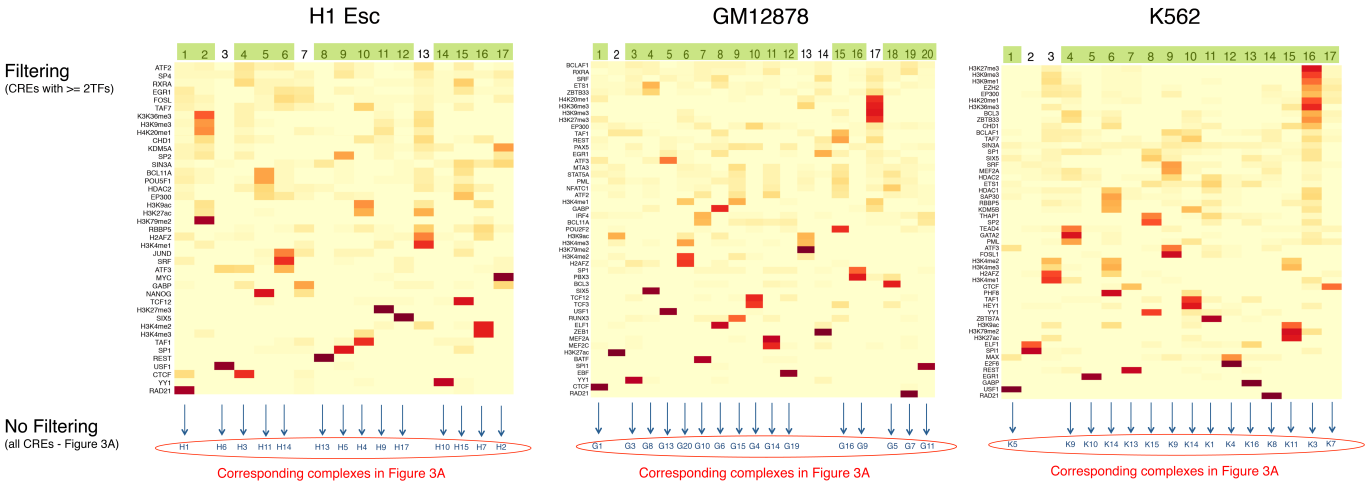
Supplementary Figure 4. Consensus matrices for the K562 dataset. Consensus matrices averaging 100 connectivity matrices at k=2,3,4,5,6,7 and k=15,16,17,18,19,20. The cophenetic correlation and dispersion qualitative measures for the corresponding ranks are highlighted in the respective plots. The red arrows indicate the rank that were selected for the the “low-rank” NMF (Supplementary text, Supplementary Figure 15) and the “high-rank” NMF (main text, Figure 3).



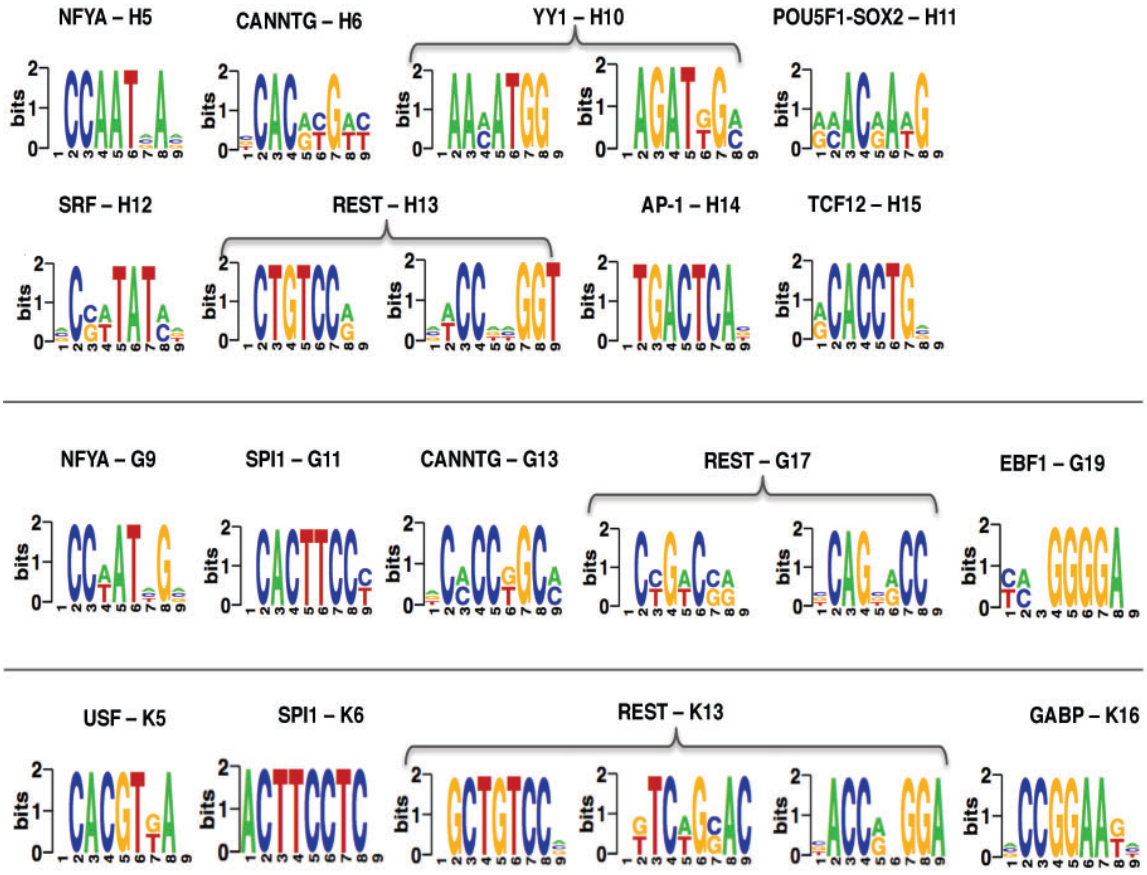
Supplementary Figure 5. Histograms of the number of TFs bound at the CREs in each of the three human cell lines. CREs bound by at least 1 TF are quite frequent (almost 85%), while more than 45% of the CREs are bound by >1 TF. Overall, the histograms show that there are many CREs bound by more than one TF, suggesting that the CREs are potentially regions where protein complexes bind.



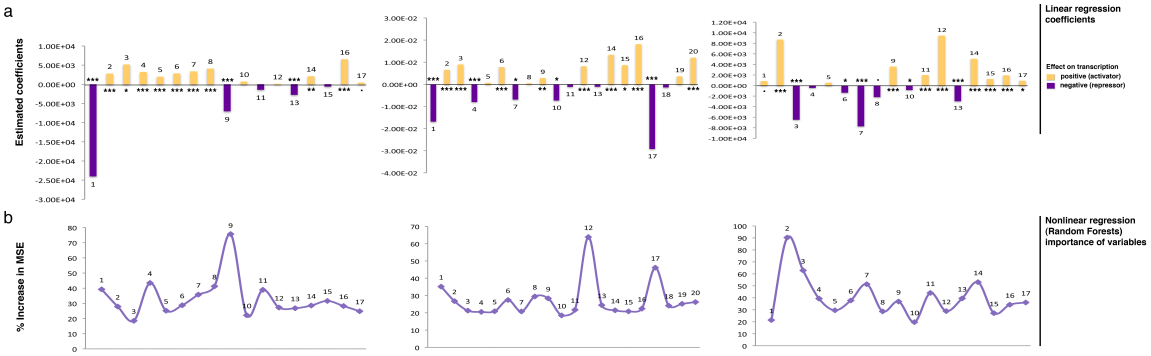
Supplementary Figure 6. Filtering of the RC matrix. The CREs bound only by one TF were filtered out from the RC matrix for the analysis described in Supplementary text. We kept only the CREs bound by at least 2 TFs, and NMF analysis was then performed on the filtered RC matrix.



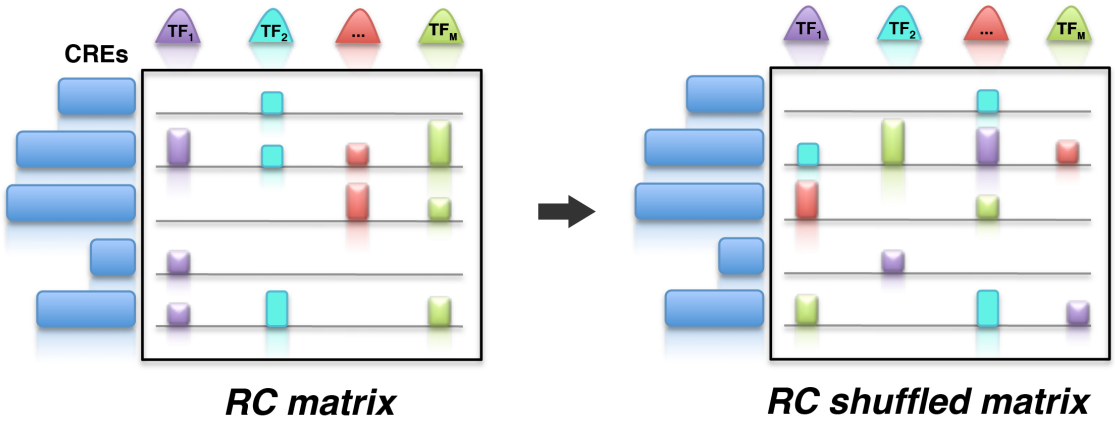
Supplementary Figure 7. Protein complexes revealed after filtering of the RC matrix. The heatmaps displaying the NMF coefficient matrices for the H1 ESC, GM12878 and K562 datasets, obtained after removing CREs that are bound by less than 2 TFs. 82% (14/17), 80% (16/20) and 88% (15/17) of the complexes identified from the filtered datasets are highly similar (>0.92 Pearson's correlation) to the complexes identified from the corresponding non-filtered datasets in the previous analysis (Figure 3a of the manuscript).



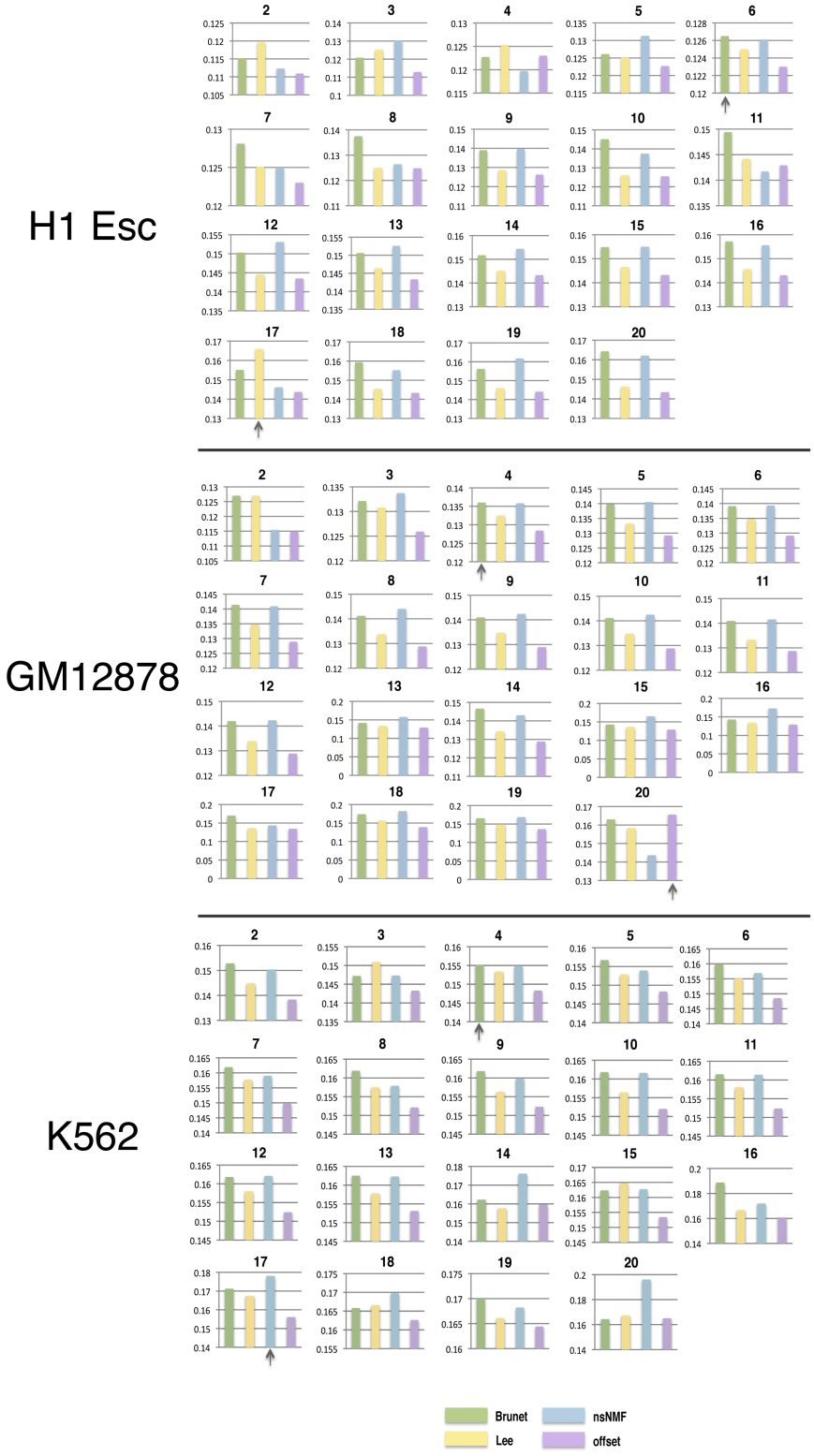
Supplementary Figure 8. Regulatory motifs analysis. Enriched motifs for the complex-specific CREs of Figure 3c.



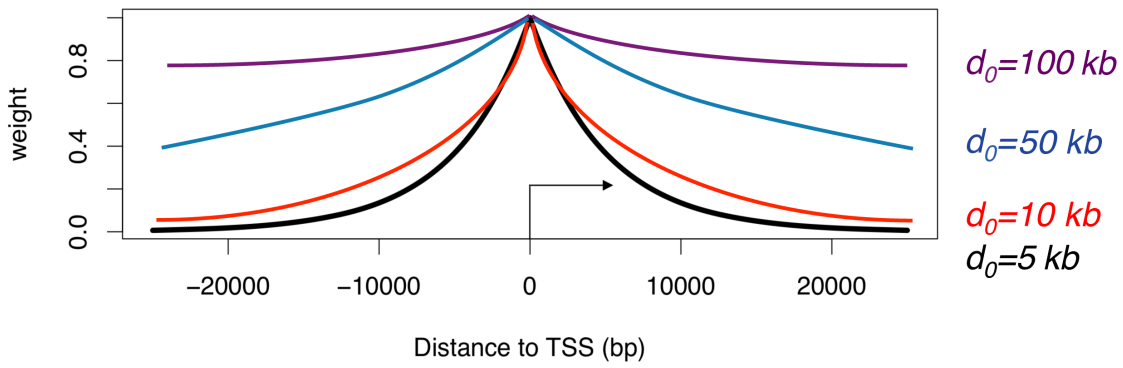
Supplementary Figure 9. The contribution of complexes to gene expression in the linear and nonlinear regression models. (a) The linear regression coefficients for each complex, also shown in Figure 3. The size of the coefficient corresponds to the size of the effect that each complex has on gene expression, and the sign of the coefficient (positive or negative) gives the direction of the effect. Statistical significance of the estimated coefficients is coded as: '***': 0<p<0.001, '**': 0.001<p<0.01, '*': 0.01<p<0.05, '.': 0.05<p<0.1, ' ': 0.1<p<1. (b) The importance of each complex in random forests regression measured by the increased mean square error that represents the deterioration of the predictive accuracy of the model when each component-predictor is replaced by random noise.



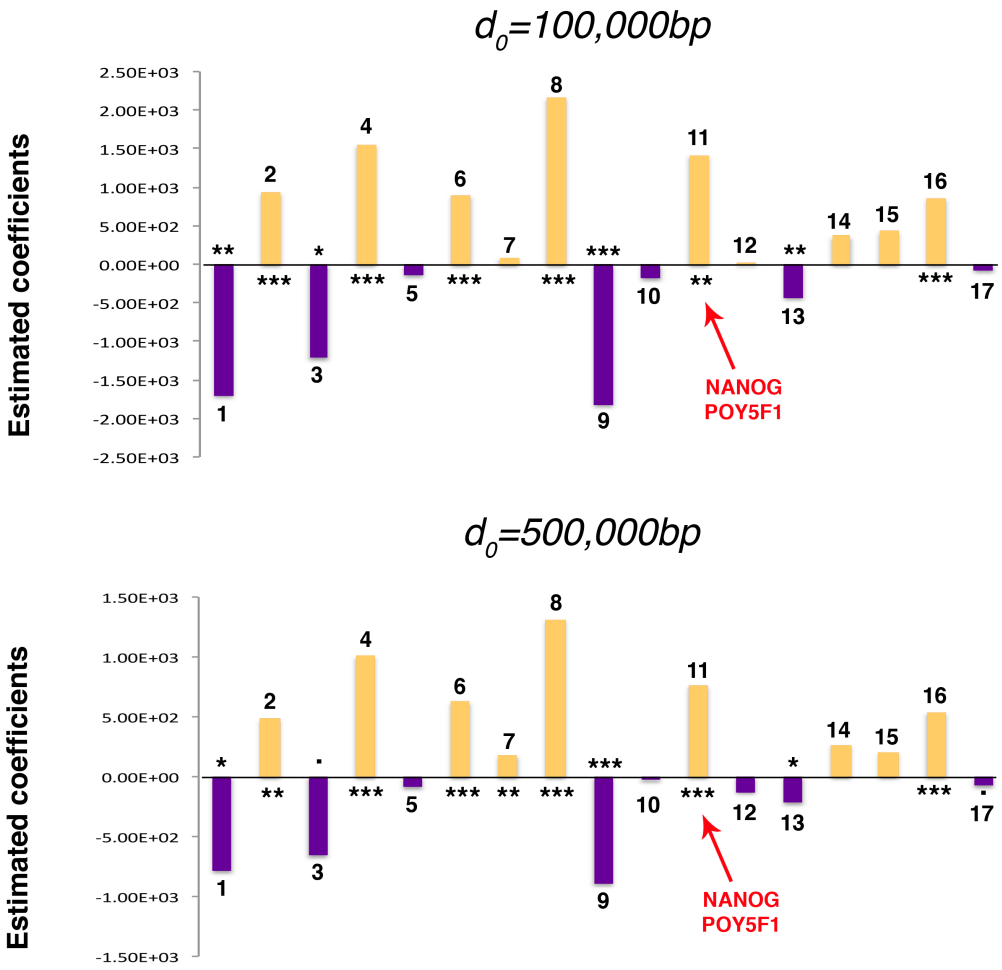
Supplementary Figure 10. Shuffling of the RC matrix. The ChIP-seq read densities were randomly shuffled within each CRE: the read density of a ChIP-seq experiment was randomly assigned to another experiment. NMF was subsequently performed on the shuffled RC matrix, and regression analysis followed using the BIS values of the new complexes as explanatory variables, in order to assess the accuracy of the predictive model.



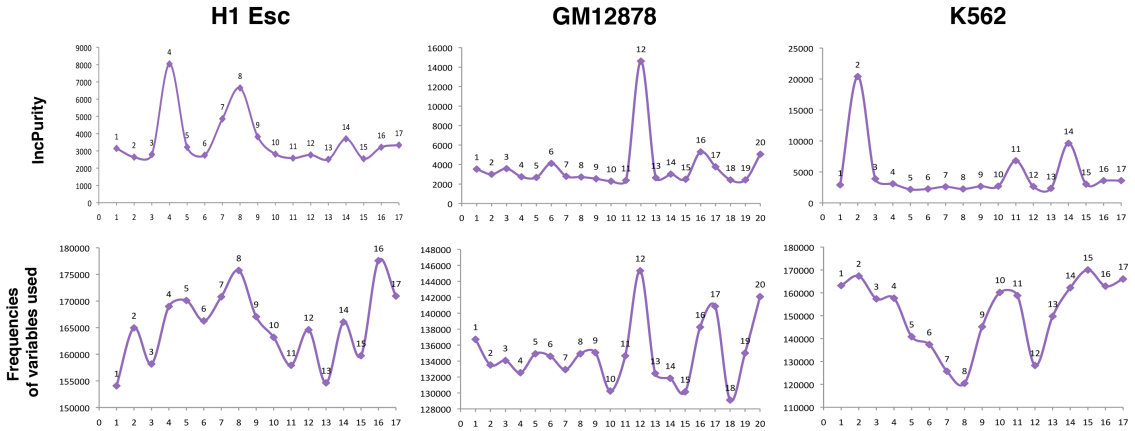
Supplementary Figure 11. NMF runs with multiple algorithms for each k. The coefficient of determination (R^2) is shown for NMF runs with different algorithms supported by the *NMF* package in R. The arrows indicate the NMF runs that we show and discuss in the paper.



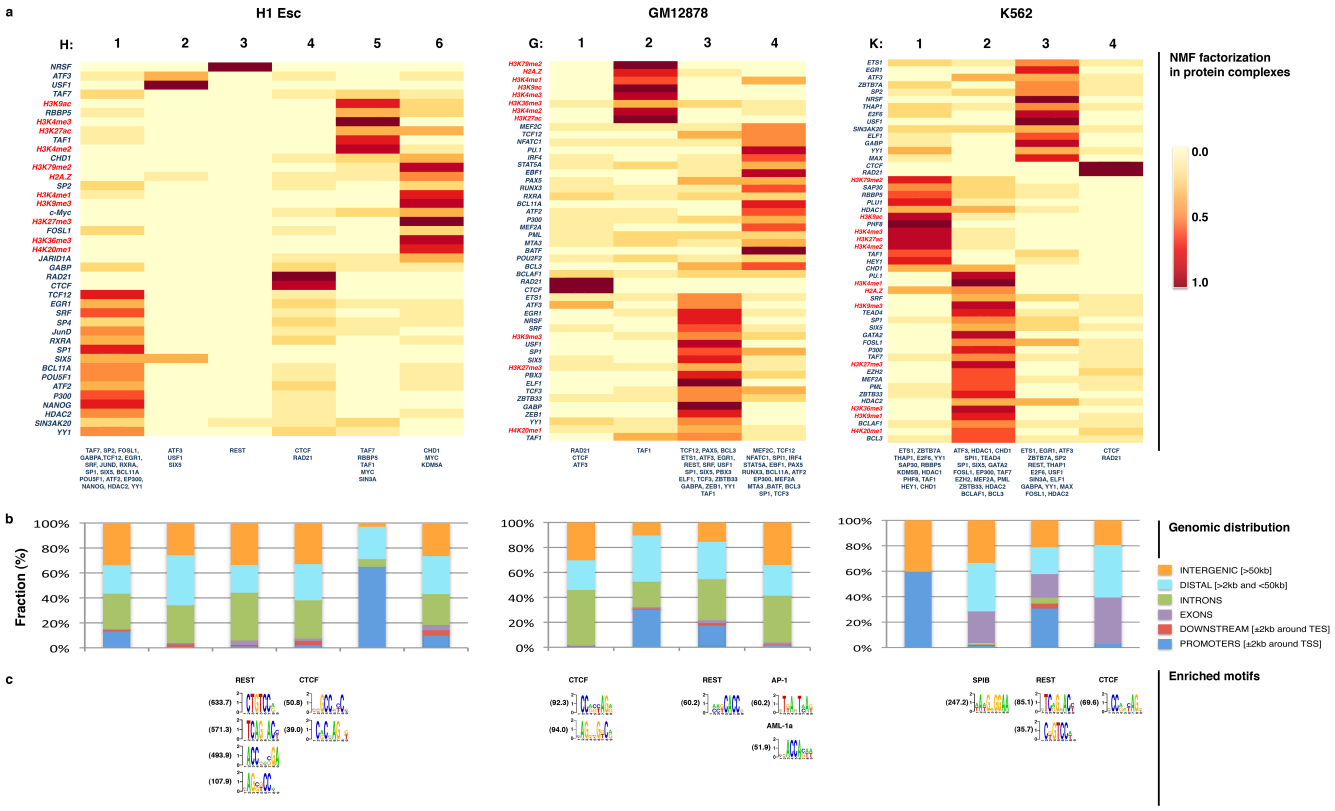
Supplementary Figure 12. The exponential function in the BIS score formula is affected by d_0. The larger the d_0, the more distant CREs will influence the promoter and eventually BIS score. In this study, we set d_0 to 5,000bp because at this value the model obtained the highest R^2 value a showed better quality of the overall fit.



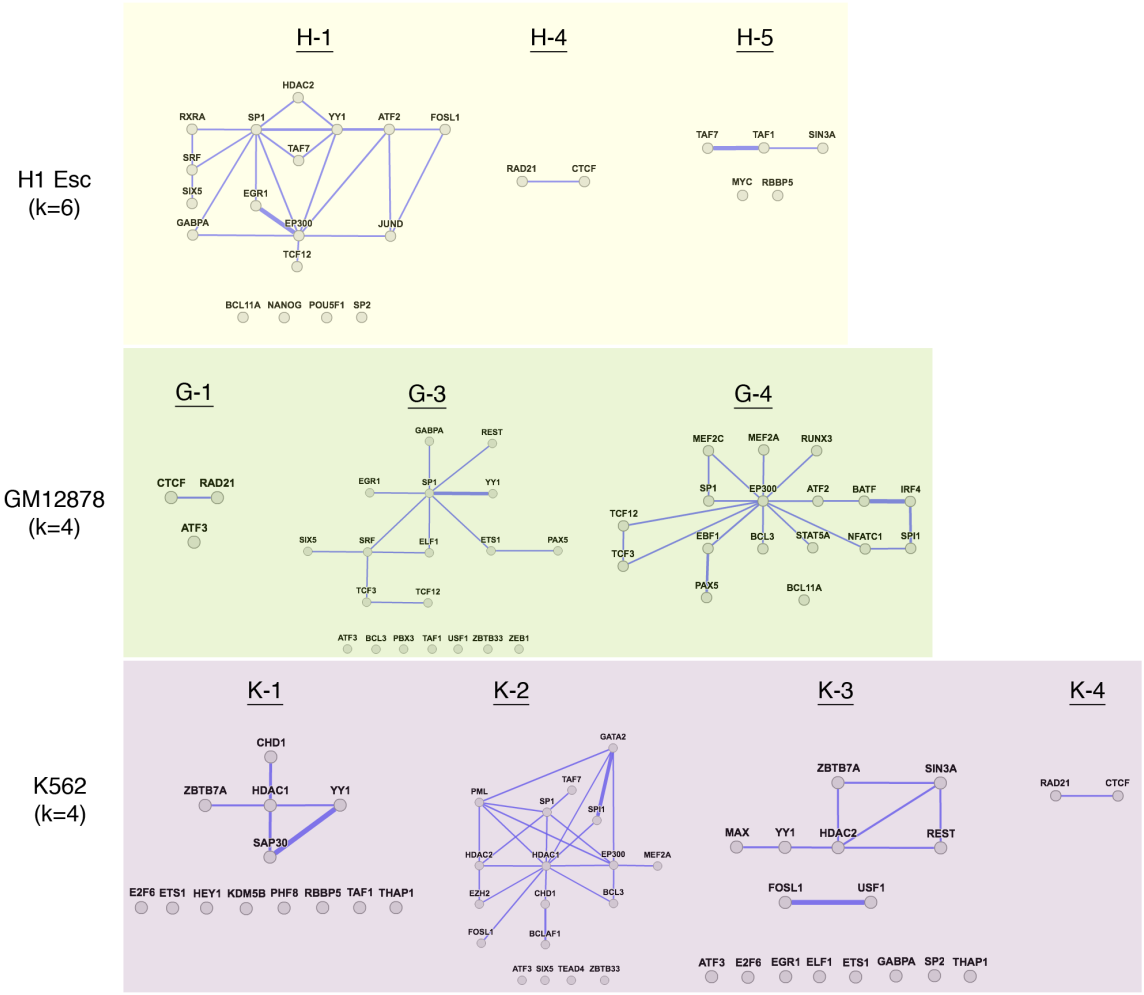
Supplementary Figure 13. **Linear regression coefficients of the complexes after changing the d_0 parameter.** The d_0 parameter in the BIS formula is changed to very high values (100 kb– upper panel, 500 kb– lower panel), in order to allow more distal complexes that influence the promoter. Now, distal regulatory elements are not penalized resulting in new coefficients for the POU5F1-SOX2-NANOG complex that are now positive. This indicates that after altering the d_0 parameter that defines the influence on the promoters of distal CREs and the complexes that bind them, this complex has an activating effect on its target genes, which is also supported by the literature. The size of the coefficient corresponds to the size of the effect that each complex has on gene expression, and the sign of the coefficient (positive or negative) gives the direction of the effect. Statistical significance of the estimated coefficients is coded as: '***': 0<p<0.001, '**': 0.001<p<0.01, '*': 0.01<p<0.05, '.': 0.05<p<0.1, ' ': 0.1<p<1.



Supplementary Figure 14. Importance measures for the complexes-predictors in random forests regression. *IncPurity* is measured by residual sum of squares and shows the decrease in node impurities from permutation of each predictor. *Frequencies of variables used* measures how many times each predictor is used in the forest.



Supplementary Figure 15. Protein complexes revealed at low rank factorization. (a) The heatmaps visualize the coefficients matrices for the three cell types. The columns represent the detected complexes and the rows the reads binding pattern of the corresponding TF/HM experiment for each complex. The cells depict the relative contribution of a complex to each experiment. All experiments are shown on the left of the heatmaps (HMs red, TFs blue). To facilitate the identification of factors within the complexes, the names of the TFs with coefficient > 0.3 are shown below each complex. **(b)** Genomic distribution of the complex-specific CREs and categorization in promoters (±2kb around TSS), downstream extremities (±2kb around TES), exons, introns, distal (>2kb and <50kb), and intergenic regions (>50kb). **(c)** Regulatory motifs analysis revealed the most enriched motifs for the complex-specific CREs. The predicted motif logos are shown along with z-scores in parentheses.



Supplementary Figure 16. Physical interaction networks for the low NMF protein complexes. Physical interactions from geneMANIA are shown for the protein complex factors found at low rank factorization for the three cell types.