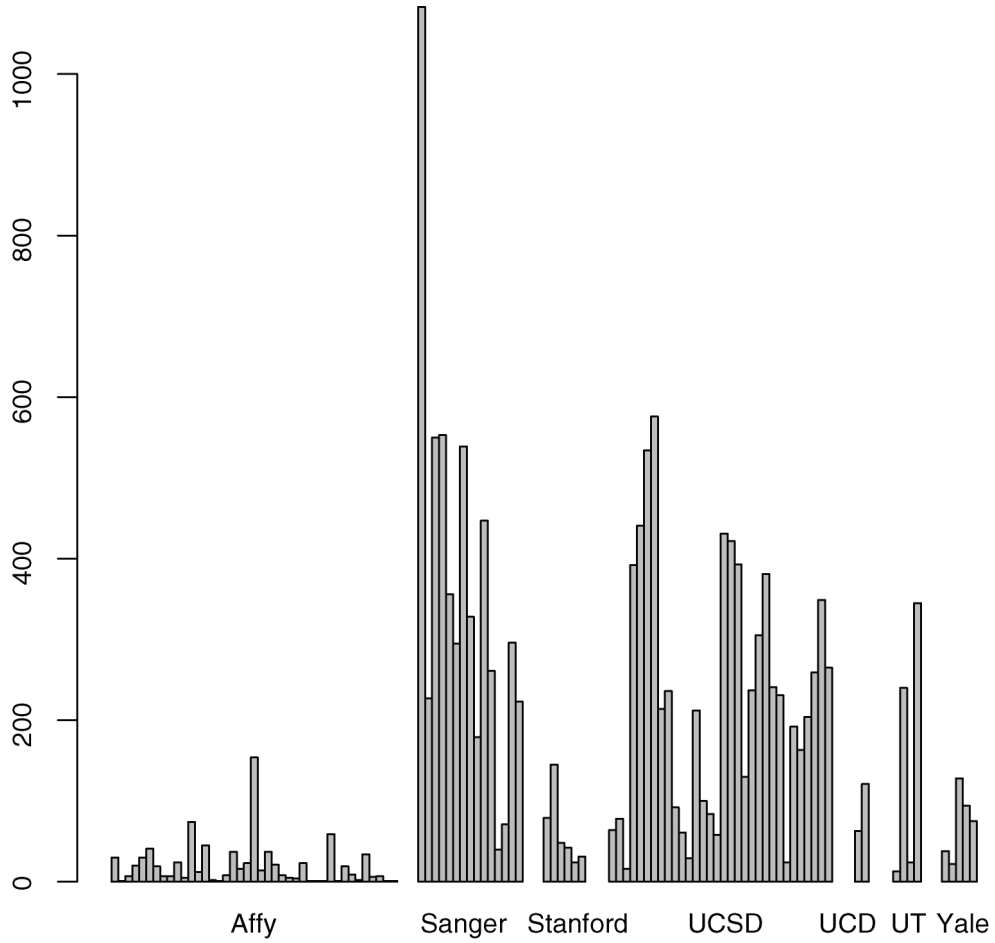
Supplementary material

Figures



Supplementary figure 1. The number of TREs identified by each of the 105 ChIP-chip experiments. Bars are aggregated by different laboratories, which are labeled by the abbreviated names of their affiliated institutions. They are Affymetrix Inc., Sanger Institute, Stanford University, University of California at San Diego, University of California at Davis, University of Texas, and Yale University.

Tables

Supplementary table 1. Four microarray designs used for ChIP-chip experiments.

Name	Category	Resolution(bp)	Coverage(Mb)
Affymetrix (AFFX)	oligo tiling	25	15.5
NimbleGen (Ng)	oligo tiling	38	14.9
UCSD	PCR tiling	~500	13.3
Sanger	PCR tiling	~1000	23.7

Supplementary table 2. Summary of TR datasets

Platform	Contributor	Cell line	Transcription factor or marker	Note
AFFX	Affymetrix	HL60	Brg1 CEBPe CTCF H3K27tm H3ac H4ac P300 PU1 RARecA SIRT1 PolII	0, 2, 8 and 32 hours after retinoic acid treatment
			TFIIB	
		ME180	p63	+/- Actinomycin D
Ng	UCSD	HeLa	H3ac H4ac H3K4dm H3K4tm PolII TAF250	+/- gamma interferon (0 and 30 min)
	Yale	HeLa	BAF155 BAF170 cJun Fos TAF250	
	Yale NASA	HeLa	STAT1	different array designs
	UT UCDavis	HeLa	cMyc	UT:stimulated/quiescent
	UT	2091	E2F4 cMyc	

	Stanford	HCT116 Jurkat K562	Sp1 Sp3	
Sanger	Sanger	HeLa GM06990	H3ac H4ac H3K4m H3K4dm H3K4tm	
		K562	H3ac H4ac H3K4dm H3K4tm	
UCSD	UCSD	HeLa	H3ac H4ac H3K4dm H3K4tm PolII TAF250 Suz12 STAT1 H3K27tm	+/- gamma interferon (0 and 30 min)
		IMR90	PolII TAF250 H3ac H3K4dm	
		HCT116 THP1	PolII TAF250	
STAGE	UT	2091	cMyc	

Biplots

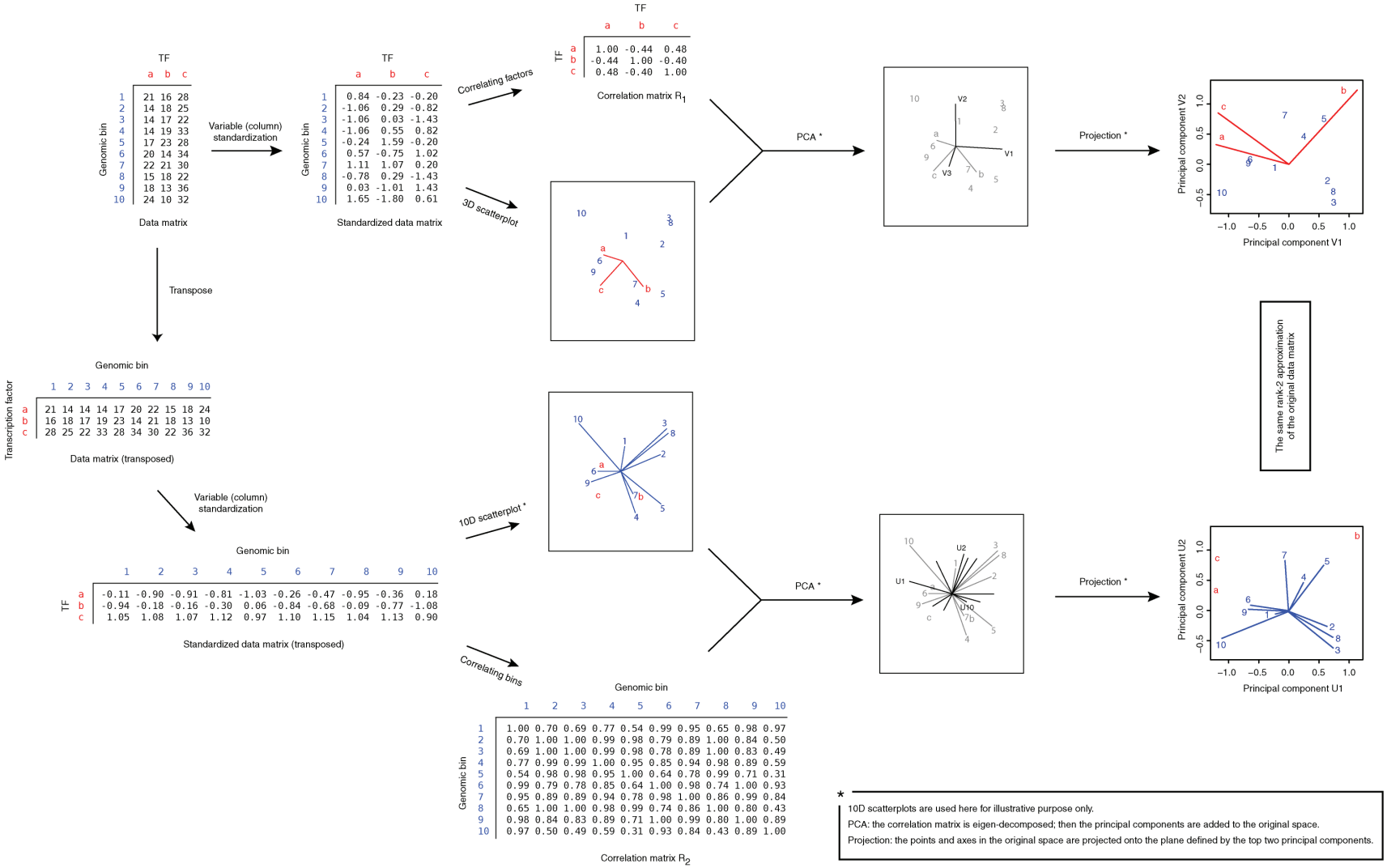
A biplot is a two-dimensional representation of a data matrix showing a point for each of the n observation vectors (rows of the data matrix) along with a point for each of the p variables (columns of the data matrix). The prefix 'bi' refers to the two kinds of points; not to the dimensionality of the plot. The method presented here could, in fact, be generalized to a three-dimensional (or higher-order) biplot. Biplots were introduced by Gabriel (1971) and have been discussed at length by Gower and Hand (1996).

We applied the biplot procedure to the following toy data matrix to illustrate how a biplot can be generated and interpreted. See the figure on the next page. Here we have three variables (transcription factors) and ten observations (genomic bins). We can obtain a two-dimensional plot of the observations by plotting the first two principal components of the TF-TF correlation matrix R_1. We can then add a representation of the three variables to the plot of principal components to obtain a biplot. This shows each of the genomic bins as points and the axes as linear combination of the factors.

The great advantage of a biplot is that its components can be interpreted very easily. First, correlations among the variables are related to the angles between the lines, or more specifically, to the cosines of these angles. An acute angle between two lines (representing two TFs) indicates a positive correlation between the two corresponding variables, while obtuse angles indicate negative correlation. Angle of 0 or 180 degrees indicates perfect positive or negative correlation, respectively. A pair of orthogonal lines represents a correlation of zero. The distances between the points (representing genomic bins) correspond to the similarities between the observation profiles. Two observations that are relatively similar across all the variables will fall relatively close to each other within the two-dimensional space used for the biplot. The value or score for any observation on any variable is related to the perpendicular projection form the point to the line.

Gabriel, K. R. (1971), "The Biplot Graphical Display of Matrices with Application to Principal Component Analysis," Biometrika, 58, 453–467.

Gower, J. C., and Hand, D. J. (1996), Biplots, London: Chapman & Hall.


TF
a b c
Genomic bin
1 21 16 28
2 14 18 25
3 14 17 22
4 14 19 33
5 17 23 28
6 20 14 34
7 22 21 30
8 15 18 22
9 18 13 36
10 24 10 32
Data matrix
Variable (column) standardization
TF
a b c
1 0.84 -0.23 -0.20
2 -1.06 0.29 -0.82
3 -1.06 0.03 -1.43
4 -1.06 0.55 0.82
5 -0.24 1.59 -0.20
6 0.57 -0.75 1.02
7 1.11 1.07 0.20
8 -0.78 0.29 -1.43
9 0.03 -1.01 1.43
10 1.65 -1.80 0.61
Standardized data matrix
Correlating factors
TF
a b c
a 1.00 -0.44 0.48
b -0.44 1.00 -0.40
c 0.48 -0.40 1.00
Correlation matrix R1
3D scatterplot
PCA *
Projection *
Principal component V2
Principal component V1
Transpose
Genomic bin
Transcription factor
1 2 3 4 5 6 7 8 9 10
a 21 14 14 14 17 20 22 15 18 24
b 16 18 17 19 23 14 21 18 13 10
c 28 25 22 33 28 34 30 22 36 32
Data matrix (transposed)
Variable (column) standardization
Genomic bin
TF
1 2 3 4 5 6 7 8 9 10
a -0.11 -0.90 -0.91 -0.81 -1.03 -0.26 -0.47 -0.95 -0.36 0.18
b -0.94 -0.18 -0.16 -0.30 0.06 -0.84 -0.68 -0.09 -0.77 -1.08
c 1.05 1.08 1.07 1.12 0.97 1.10 1.15 1.04 1.13 0.90
Standardized data matrix (transposed)
10D scatterplot *
PCA *
Projection *
Principal component U2
Principal component U1
Correlating bins
Genomic bin
1 1.00 0.70 0.69 0.77 0.54 0.99 0.95 0.65 0.98 0.97
2 0.70 1.00 1.00 0.99 0.98 0.79 0.89 1.00 0.84 0.50
3 0.69 1.00 1.00 0.99 0.98 0.78 0.89 1.00 0.83 0.49
4 0.77 0.99 0.99 1.00 0.95 0.85 0.94 0.98 0.89 0.59
5 0.54 0.98 0.98 0.95 1.00 0.64 0.78 0.99 0.71 0.31
6 0.99 0.79 0.78 0.85 0.64 1.00 0.98 0.74 1.00 0.93
7 0.95 0.89 0.89 0.94 0.78 0.98 1.00 0.86 0.99 0.84
8 0.65 1.00 1.00 0.98 0.99 0.74 0.86 1.00 0.80 0.43
9 0.98 0.84 0.83 0.89 0.71 1.00 0.99 0.80 1.00 0.89
10 0.97 0.50 0.49 0.59 0.31 0.93 0.84 0.43 0.89 1.00
Correlation matrix R2
The same rank-2 approximation of the original data matrix
* 10D scatterplots are used here for illustrative purpose only.
PCA: the correlation matrix is eigen-decomposed; then the principal components are added to the original space.
Projection: the points and axes in the original space are projected onto the plane defined by the top two principal components.