

## Supporting Information

### **S1 Fig. Hierarchical clustering of stress condition differential expression correlation.**

Dendrogram on the left shows the results of hierarchical clustering using complete linkage and Euclidean distance of the pairwise sample PCC. The table on the right summarizes the organ/cell type, the stress condition, and the time point for each dataset.

### **S2 Fig. Within and between family PCC distribution**

Boxplots summarize the PCC comparing TFBMs within each family, and between families. The boxes are color coded by family.

### **S3 Fig. Under/overrepresentation of TF family like PWMs in the full pCREs.**

Over and under-representation of TF families in the organ-stress pCREs as defined by significant match to a TFBM using the between family PCC distributions. Bars represent the Log2 odds-ratio between the significant matching pCREs and the TFBMs for a particular family. Black bars signify a significant enrichment (Fisher Exact Test,  $p < 0.05$ ), while grey bars are not significant.

### **S4 Fig. PR curves for models based only on the full pCREs that are found in at least one combinatorial rule**

Same as **Figure 5**, except that only pCREs involved in a combinatorial rule are included in the models. All other full pCREs are excluded from this analysis. Note that the PR curves are highly similar to those in **Figure 5**, and the performance as measured by AUC-ROC is highly comparable.

### **S5 Fig. Summarization of the distance between pairs of motifs in combinatorial rules.**

The distance between pairs of motifs in the same rule for all instances of the rule on all of the promoters of all genes in the *A. thaliana* genome for the **(A)** root and **(B)** shoot rule sets. Dots represent the median distance for a rule, and the edges of the ribbon represent the 25<sup>th</sup> and 75<sup>th</sup> percentile of distances. The color of each dot represents the significance (Mann-Whitney, adjusted  $p < 0.05$ ) of the distance distribution compared the random background distribution. Rows are sorted from lowest median distance to highest median distance. **(C)** The log2 ratio of median distance for shoot rule instances on root up-regulated genes and non-responsive genes, ordered from low to high values, with each dot representing one rule. **(D)** Same as **(C)**, but for shoot rules on shoot up-regulated and non-responsive genes.

### **S6 Fig. Summary of cell-type salt pCREs**

Each column of the heat map represents a cell type, each row represents a cell-type salt pCRE, and the values represent a significant over-representation on the promoters of salt up-regulated genes for that cell type (red) or no over-representation (white). Bar plot on top represents the counts of motifs specific to each cell type.

### **S7 Fig. Average PCC between pCREs from different pCRE sets**

Heatmap summarizes the average PCC between pairs of pCREs from two sets of pCREs. The pCRE set show are all six cell-type specific pCRE sets, the combination of pCREs that are over-represented on promoters of genes from up-regulated by salt in more than 1 but less than 6 cell types (COMBO), pCREs over-represented on promoters of genes up-regulated by salt in all 6 cell-types (CELL GEN), along with the general pCREs (ORGAN GEN), the root pCREs (R), and the shoot pCREs (S).

### **S8 Fig. Overlap in salt stress up-regulated genes correctly predicted using models based on cell-type salt pCREs and full pCREs.**

Venn diagram on the left describes the overlap in genes up-regulated by salt in the Stele cells that were correctly predicted by the full pCRE set (“O”) and the cell-type pCREs (“C”). Diagram on the left is the same, except these are predictions of genes that are only up-regulated in the stele during salt stress. Numbers outside of the circles represent the percentage of genes up-regulated by salt in stele that were not predicted correctly using models based on either pCRE set.

### **S1 Table. Summary of TFs that are up-regulated in an organ specific or globally under salt stress.**

Six columns are in this table. Column 1 gives the AGI number for TF up-regulated at 3 hours of salt stress. Column 2 specifies if this TF is up-regulated specifically in either the root or shoot, or globally. Column 3 gives the family of the TF as it was assigned in Weirauch et al. [16]. Column 4 comes from Weirauch et al.[16], and specifies if the TF has an experimentally determined TFBM (D), an TFBM inferred from DNA binding domain similarity (I), or if no TFBM was assigned (D). Finally columns 5 and 6 specify which, if any, TFBMs assigned to this TF are overrepresented on the promoters of root or shoot up-regulated genes.

### **S2 Table. Summary of transcription factor binding motifs used for modeling salt stress up-regulation**

TFBM summary: The motif id from CisBP, International Union of Pure and Applied Chemists (IUPAC) sequence generated for this study, and associated TF for the TFBM in the dataset. The TF are

labeled with the AGI number, whether or not the TFBM was matched to the TF through experimental determination (D) or inference (I), and the TF family. The last column represents all the families represented by TF for this TFBM.

Root Enriched TFBM and Shoot enriched TFBM: Same format as the TFBM summary, but only featuring the TFBMs that are enriched in either the root salt stress up-regulated or shoot salt stress up-regulated genes. In some cases, the IUPAC sequence looks identical for several motifs. In this case, only one representative of these motifs were used for the modeling.

**S3 Table. Summary of full pCRE set.**

Table includes the IUPAC sequence of the pCREs identified in for root and shoot salt up-regulated genes, along with which subset the pCRE belongs to (root, shoot, or general pCRE).

**S4 Table: Overlap in binary predictions of genes from models based on different pCRE sets.**

The first sheet summarizes the genes overlapping in predictions for models predicting root up-regulated genes. Columns 1 and 2 describe which pCRE sets are being considered. Column 3 gives the median percentage overlap of correctly predicted up-regulated genes. Columns 4 and 5 describe the median percentage overlap between multiple runs of the same pCRE based model. Column 6 gives the *p* value for a Mann-Whitney test of the percentage overlap between multiple runs using a pCRE set 1 based model against the percentage overlap between runs of models based on pCRE set 1 and pCRE set 2. Column 7 makes the same test, but it focuses on pCRE set 2.

**S5 Table: Summary of SVM models generated.**

This table lists four things about each model: the (1) up-regulated genes and (2) non-responsive gene sets that were being predicted, (3) the promoter features used as the basis of the model (pCRE set, TFBM set, or pCRE rule set), and (4) the average AUC-ROC for the model.

**S6 Table: SVM model parameters used in grid search**

Listing of each parameter used along with a comma delimited list of the possible values used in the grid search.

**S7 Table: Parameters used to discover root rules with CBA.** The following table lists the values of confidence, support, and the numbers of non-responsive and up-regulated genes used for CBA runs to discover combinatorial rules.

**S8 Table: Summary of full pCRE and TFBM motif correlation**

Between Family Best Match: Each full pCRE is represented with its best match to a TFBM based upon the lowest adjusted p-value using the “between family” distributions to compare PCC.