Legends for supplementary figures and tables

Supplemental Table 1. The Arabidopsis predicted interactome. This is a multi-sheet file containing the raw source data for the construction of interologs. The complete interactome (Arabidopsis_Interactome sheet) is presented as protein-protein pairs, with the confidence value (CV), expression correlation coefficient (PCC), and subcellular localization of each protein. Note that all hetero-dimers are presented twice, as A interacts with B, and B interacts with A so that one needs only search column A to identify a protein of interest. The interacting proteins page lists all proteins in our dataset, the number of interactions (hub size), and whether mutations of these genes are lethal (according to Tzafrir et al., 2004). The analysis_details sheet summarizes the species orthologous interactions were found in, and the amount and type of experimental support used to build the CV. The raw sources for each orthologous interaction, including the pubmed ID of the experimental source and database the interaction was taken from is listed on the Sources of Interactions sheet. A functional analysis of chloroplast localized Arabidopsis interologs, and all interacting proteins grouped by hub size is presented in the final two sheets.

Supplemental table 2. Sources of microarray expression data. This table lists all microarray experiments used from the At Gen Express dataset to generate the co-expression maps used in the Botany Array resource and this work. The original experiment code and ATGE experiment IDs are listed, and raw data can be obtained from (http://www.weigelworld.org/resources/microarray/AtGenExpress/)

Supplemental table 3. Resolution of conflicting localizations in SUBA. 88 cases where there was disagreement between MS and GFP data, and 9 cases where MS and GFP data were overridden by other experimental approaches are listed, and the original conflicting data is presented, along with our resolution decision.

Supplemental Figure 1. RNA_splicing network expanded by predicted interactions. Proteins with known, experimentally determined interactions (blue lines) from the BIND dataset formed an initial set. This was expanded one layer outwards by identifying all proteins which are predicted to interact with proteins from the initial set. All predicted interactions are rated by CV (line thickness) and co-expression (line color). Nodes are color coded with predicted subcellular localizations and sized according to the number of predicted interacting protein partners throughout the entire predicted interactome.

Supplemental Figure 2. RHO-RAB network expanded by predicted interactions. Proteins with known, experimentally determined interactions (blue lines) from the BIND dataset formed an initial set. This was expanded one layer outwards by identifying all proteins which are predicted to interact with proteins from the initial set. All predicted interactions are rated by CV (line thickness) and co-expression (line color). Nodes are color coded with predicted subcellular localizations and sized according to the number of predicted interacting protein partners throughout the entire predicted interactome.

Supplemental Figure 3. Homeobox STM / KNAT / BELL shoot apical meristem forming regulator network expanded by predicted interactions. Proteins with known, experimentally determined interactions (blue lines) from the BIND dataset formed an initial set. This was expanded one layer

outwards by identifying all proteins which are predicted to interact with proteins from the initial set. All predicted interactions are rated by CV (line thickness) and co-expression (line color). Nodes are color coded with predicted subcellular localizations and sized according to the number of predicted interacting protein partners throughout the entire predicted interactome.

Supplemental Figure 4. Distribution and construction of the Confidence Value.

The arbitrary confidence value was built from the product of total experimental support (N; blue bars), number of species with orthologous interaction (S) and support by different types of experiments (E). The rationale is that a wide variety of evidence is more convincing than repetition using the same methods, and so should receive a higher score. Note log scale in Y-axis.

Supplemental Figure 5. Analysis of hub size.

Top panel: The distribution of interacting proteins was ranked on a linear scale (v.s. the class based scale presented in figure 2), showing an exponential decrease in frequency for increasingly larger hubs (proteins with multiple partners). Lower panel: Interacting proteins in 3 different categories were ranked for fraction of lethal or indespensible genes (according to Tzafrir et al., 2004). The molecular functions of large hubs according to GO annotation were enriched for protein, nucleic acid and nucleotide binding (asterisks) when compared to the whole genome (see supplementary table 1 for numbers).