Figure 2: EGRIN 2.0 Model Validation

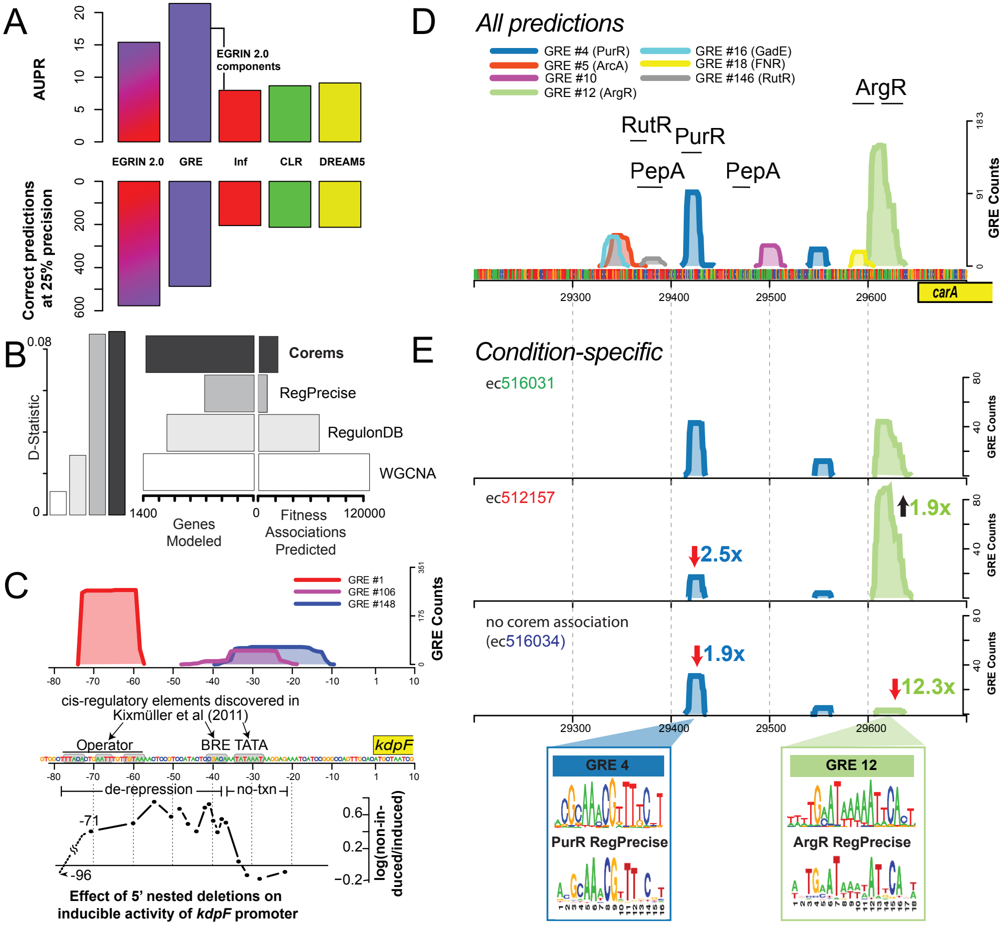

Figure 2. EGRIN 2.0 Model Validation.

(A) EGRIN 2.0 performance on experimentally-validated gold-standard network. Comparison of EGRIN 2.0 model components ("GRE": GRE-only; "Inf": Inferelator-only) to CLR and the DREAM5 community ensemble network, against RegulonDB (strong evidence code). (Top) Area under the precision-recall curve (AUPR) and (Bottom) number of correct predictions at 25% precision.

(B) Enrichment of similar fitness effects within gene modules. (Left) Magnitude of enrichment for gene pairs with similar fitness consequences, assessed by one-tailed KS-test (KS D-statistic). (Right) Number of genes and gene-pairs predicted by each method. Comparison methods include EGRIN 2.0 corems, co-expression modules from WGCNA, and regulons from databases (RegPrecise and RegulonDB).

(C) Promoter architecture of the H. salinarum kdpFABC promoter predicted by the EGRIN 2.0 model. (Top) Frequency of GRE alignment to each position in the kdpFABC promoter. GREs are indicated by shaded lines. (Middle) Genome sequence marked with putative functions by (19). (Bottom) Transcriptional activity measurements from truncated promoters used by authors to validate these sites.

(D) Predicted architecture of the E. coli carA promoter across all ensemble predictions (as in C). Horizontal bars above peaks mark experimentally characterized TF binding sites (RegulonDB). Significant GRE matches to characterized E. coli binding sites in RegulonDB are indicated in parentheses.

(E) Condition-specific states of the carA promoter in E. coli. Variation in conditional discovery of GREs (counts and fold-change relative to ec516031, top) suggests when they are "active" across three different subsets of experimental conditions in the carA promoter. (Bottom). Condition subsets correspond to co-regulation of carA with genes in the nucleotide and pyrimidine corems (ec516031, ec512157) or environments where carA is not co-regulated with genes in any corem (ec516034). Motif logos for GRE #4 (PurR) and GRE #12 (ArgR) from the EGRIN 2.0 predictions compared to logos from RegPrecise.

See also Figure S2.