An integrative, multi-scale, genome -wide model reveals the phenotypic landscape of Escherichia coli

Javier Carrera， Raissa Estrela

演讲：孙隽
翻译：熊丹

- Background
- Materials and Methods
- Results
- Discussion

Background

- The development of an integrative genome-scale model : Holy Grail

- Potential
  - discovery of novel properties and emerging behaviors
  - generating and testing predictable hypotheses
  - guiding experimentation
  - accelerating the in-depth understanding of cellular physiology

Early work

- E-cell

a modular software environment for whole-cell simulation that included organelle sub-models

More recently work

genome-scale simulations were performed to study complex phenomena

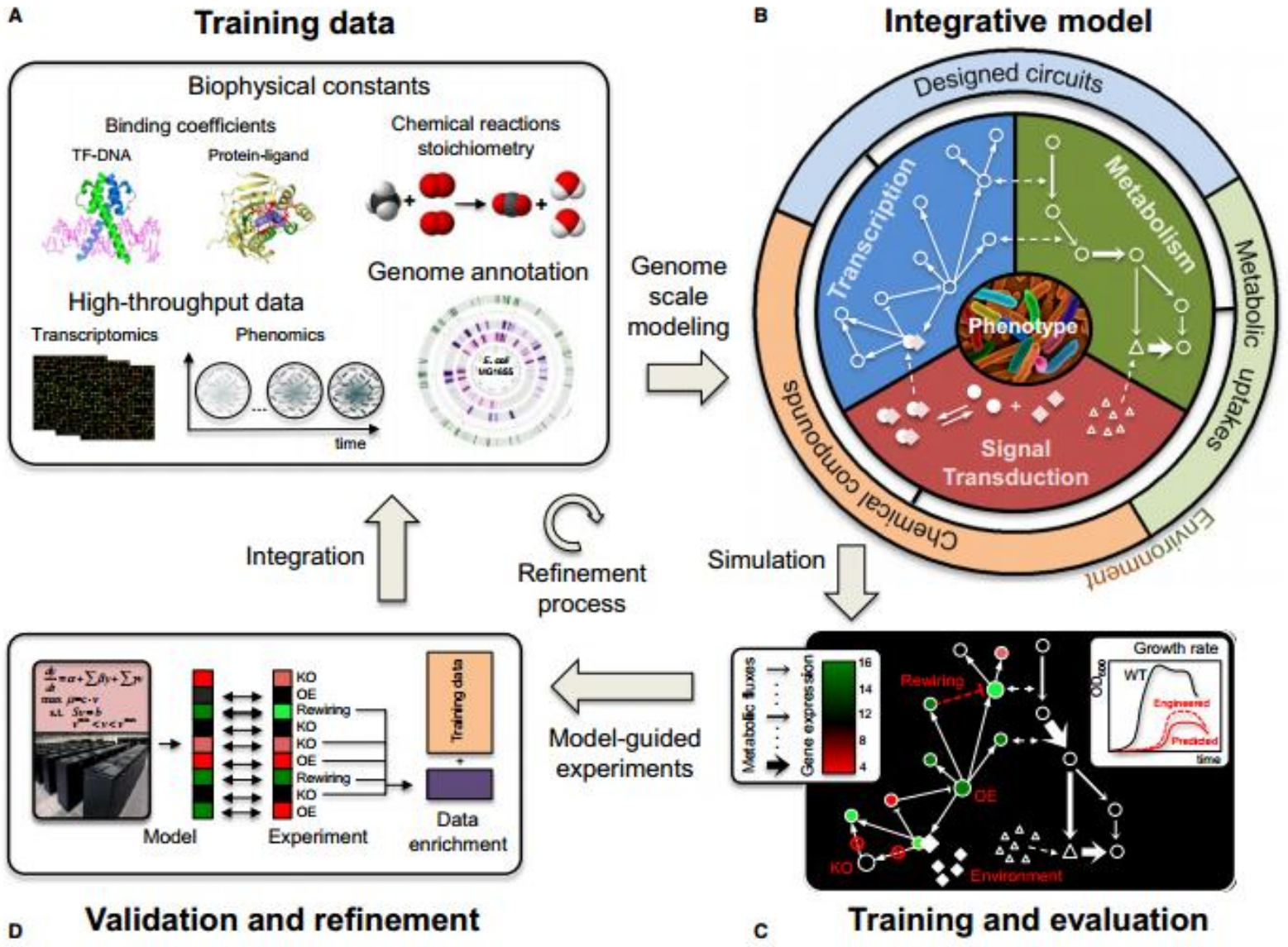
- the emergence of anticipatory behavior during evolution in varying environments
- the noise contributions of an inducible switch
- the effect of stochastic expression to metabolic variability

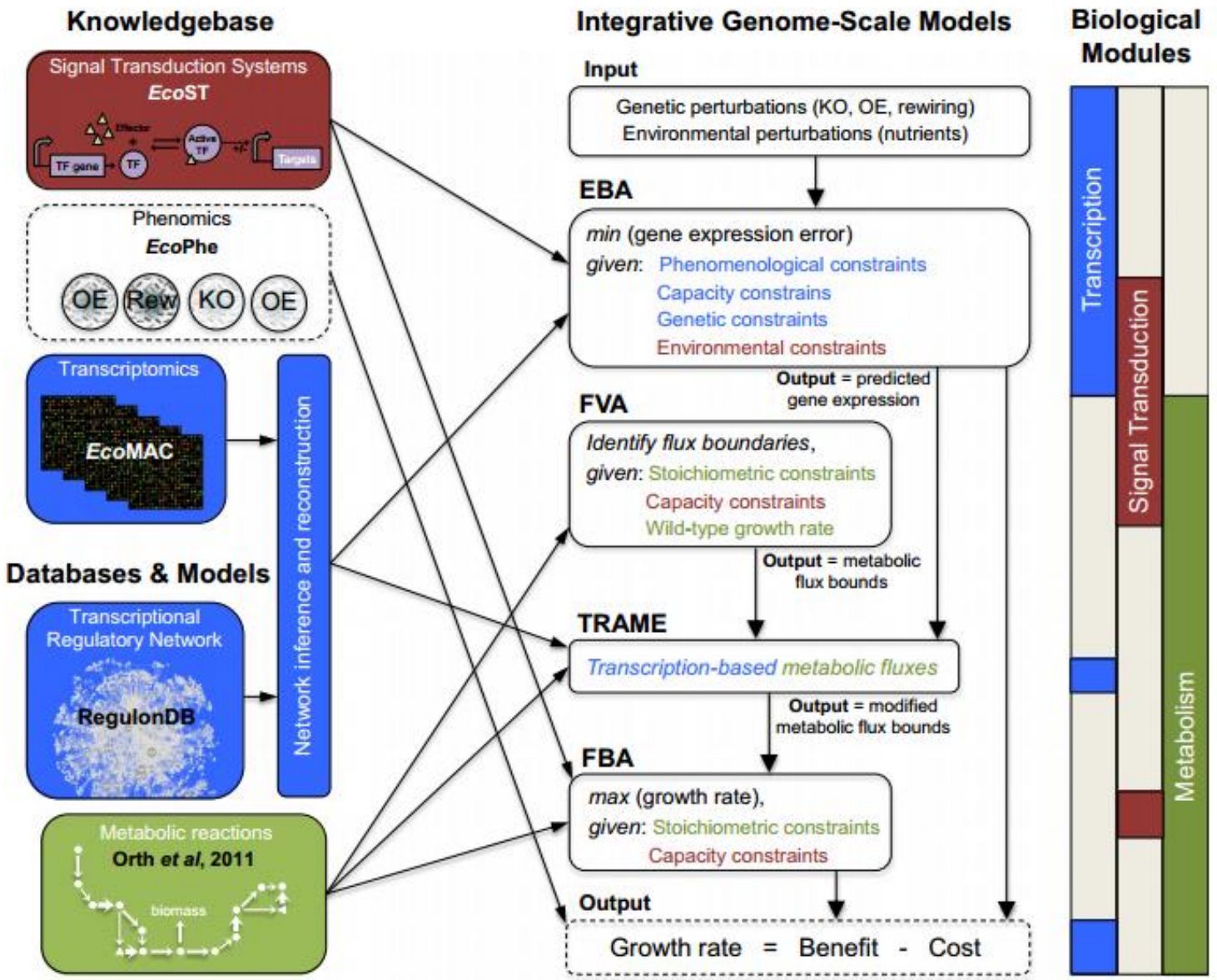
Our aim

To construct a phenomenological model for bacterial organisms that integrates multiple layers of biological organization.

Why *E.coli*

- The wealth of data and knowledge accumulated over the years
- The easiness to culture and manipulate experimentally
- Its importance in medical and biotechnological applications


A
Training data
Biophysical constants
Binding coefficients
TF-DNA
Protein-ligand
Chemical reactions stoichiometry
Genome annotation
High-throughput data
Transcriptomics
Phenomics
time
Genome scale modeling
B
Integrative model
Designed circuits
Transcription
Metabolism
Metabolic uptakes
Phenotype
Signal Transduction
Chemical compounds
Environment
Integration
Refinement process
Simulation
Model-guided experiments
Metabolic fluxes
Gene expression
Rewiring
Growth rate
WT
Engineered
Predicted
time
OE
KO
Environment
C
Training and evaluation
KO
OE
Rewiring
KO
KO
OE
Rewiring
KO
OE
Training data
Model
Experiment
Data enrichment
D
Validation and refinement


Knowledgebase
Signal Transduction Systems
EcoST
Effector
Active TF
TF gene
TF
Targets
Phenomics
EcoPhe
OE
Rew
KO
OE
Transcriptomics
EcoMAC
Databases & Models
Transcriptional Regulatory Network
RegulonDB
Metabolic reactions
Orth et al, 2011
biomass
Network inference and reconstruction
Integrative Genome-Scale Models
Input
Genetic perturbations (KO, OE, rewiring)
Environmental perturbations (nutrients)
EBA
min (gene expression error)
given: Phenomenological constraints
Capacity constrains
Genetic constraints
Environmental constraints
Output = predicted gene expression
FVA
Identify flux boundaries,
given: Stoichiometric constraints
Capacity constraints
Wild-type growth rate
Output = metabolic flux bounds
TRAME
Transcription-based metabolic fluxes
Output = modified metabolic flux bounds
FBA
max (growth rate),
given: Stoichiometric constraints
Capacity constraints
Output
Growth rate = Benefit - Cost
Biological Modules
Transcription
Signal Transduction
Metabolism

Materials and Methods

- Data
- Cellular sub-models
- Model integration

Data

- gene expression
  - from GEO, ASAP database
  - constructed a gene expression compendium of 4,189 genes over 2,198 arrays that were collected from 127 scientific articles
- signal transduction
  - A total of 328 transcription factors (TFs) and 1,357 enzymes were identified by using RegulonDB
  - identify 151 instances of signal transduction systems (STSs)
- Phenomics compendium
  - bacterial growth information for 616 of the arrays in EcoMAC by EcoPhe

Cellular sub-models

- Signal transduction model

$$y_{\mathrm{TF}} = y_{\mathrm{TF}}^{\mathrm{wt}} + \Omega(C_{\mathrm{TF}}^{\max} - C_{\mathrm{TF}}^{\min})\chi_{\mathrm{TF}}^{E}\frac{\Delta n_E}{\Delta n_E^{\max}}$$

- Transcriptional model and EBA

$$E = \frac{1}{2}\begin{bmatrix}\bar{y}_{\mathrm{TF}} & \bar{\varepsilon}_{\mathrm{TF}}\end{bmatrix}\bar{\bar{H}}\begin{bmatrix}\bar{y}_{\mathrm{TF}} \\ \bar{\varepsilon}_{\mathrm{TF}}\end{bmatrix} + \bar{f}\begin{bmatrix}\bar{y}_{\mathrm{TF}} \\ \bar{\varepsilon}_{\mathrm{TF}}\end{bmatrix}$$

- Metabolic model and Transcription-based Flux Enrichment

$$PV_{\min} \le v \le PV_{\max}, \text{ where } P_e = \left(\frac{y_e}{y_e^{\mathrm{wt}}}\right)^n$$

Model integration

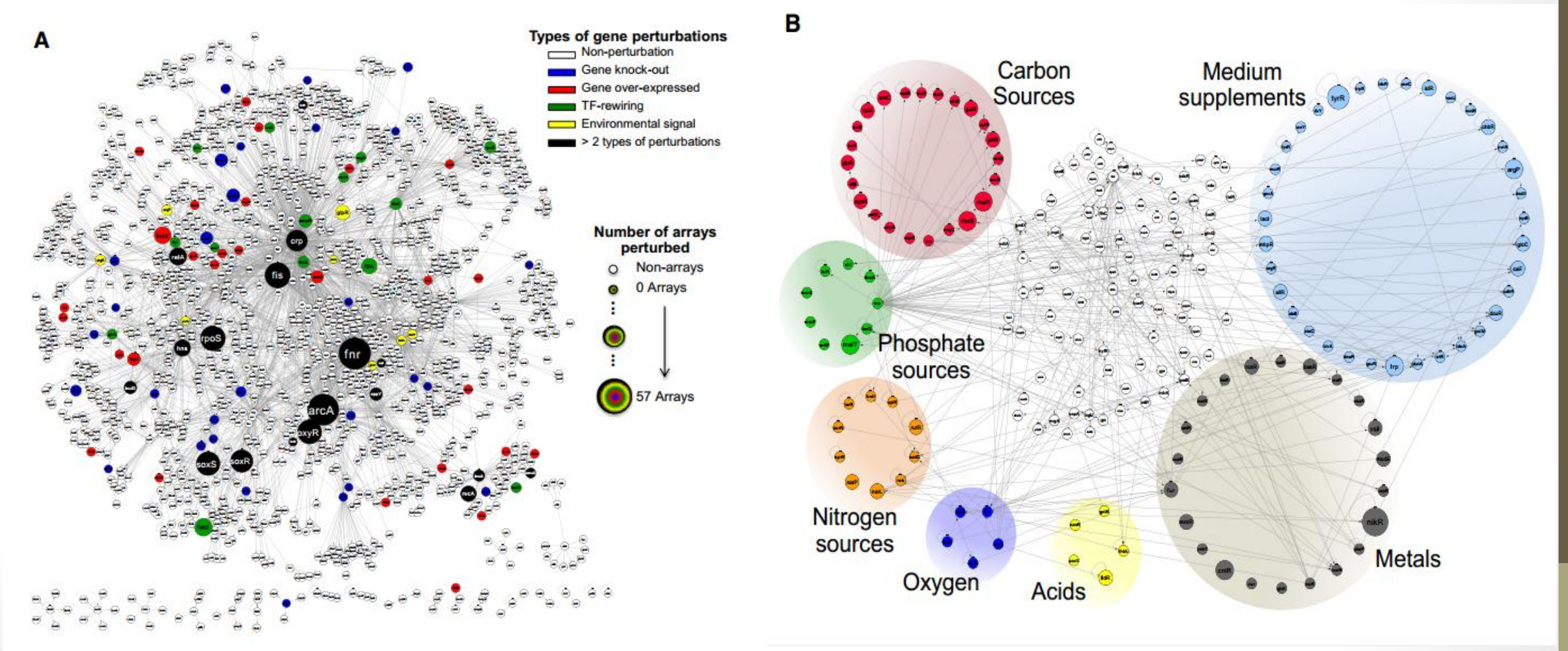
$$c = \frac{1}{N_G} \sum_g \left| \frac{\bar{y}_g - y_g^{\mathrm{WT}}}{y_g^{\mathrm{WT}}} \right|$$

$$\bar{\mu} = B - c$$

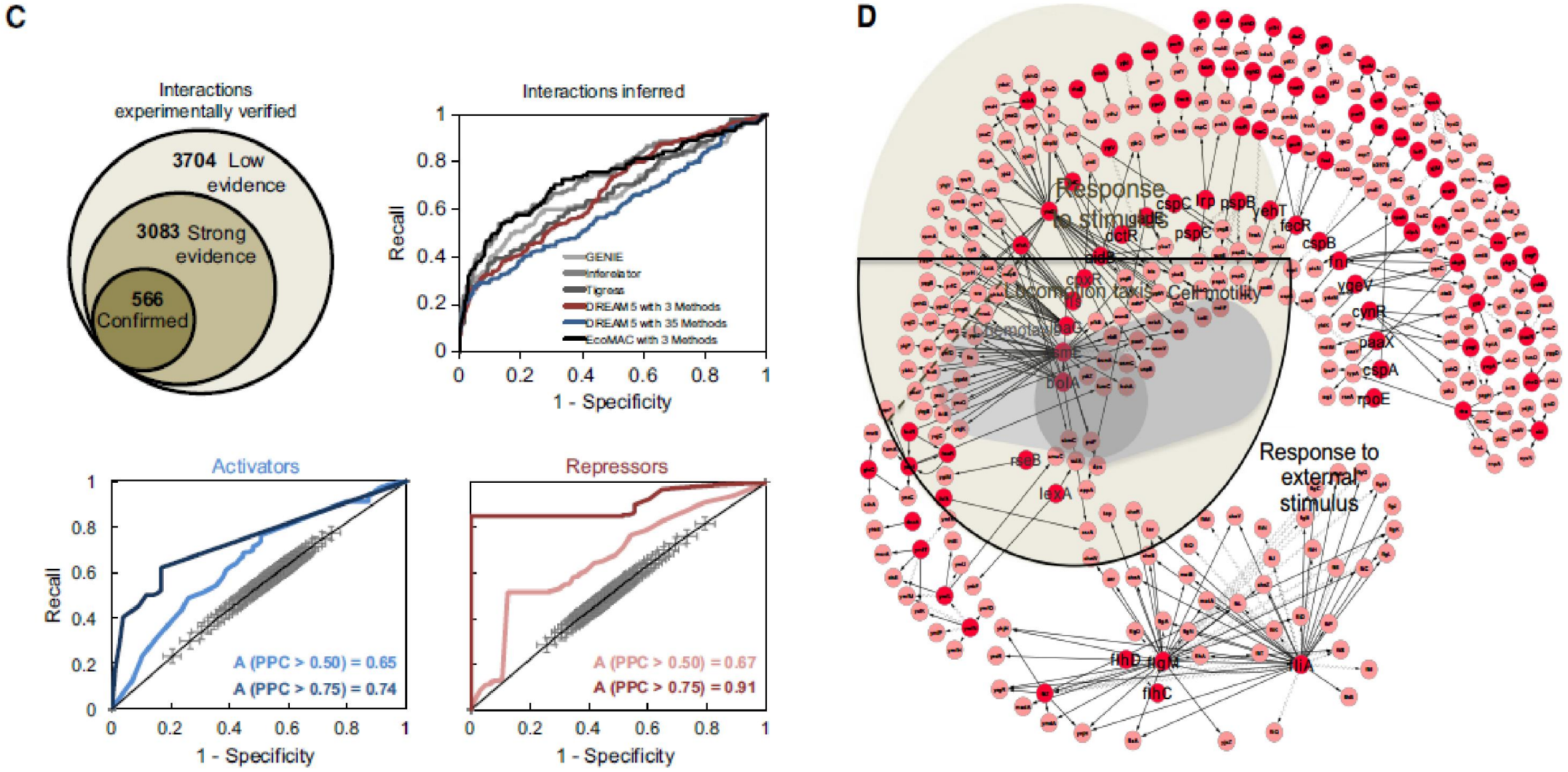
Results

- Genetic and environmental gene expression diversity
- An integrative knowledgebase as a base to regulatory network enrichment
- Expression Balance Analysis
- Phenotypic predictions in an integrated model
- Model enrichment through targeted experimentation

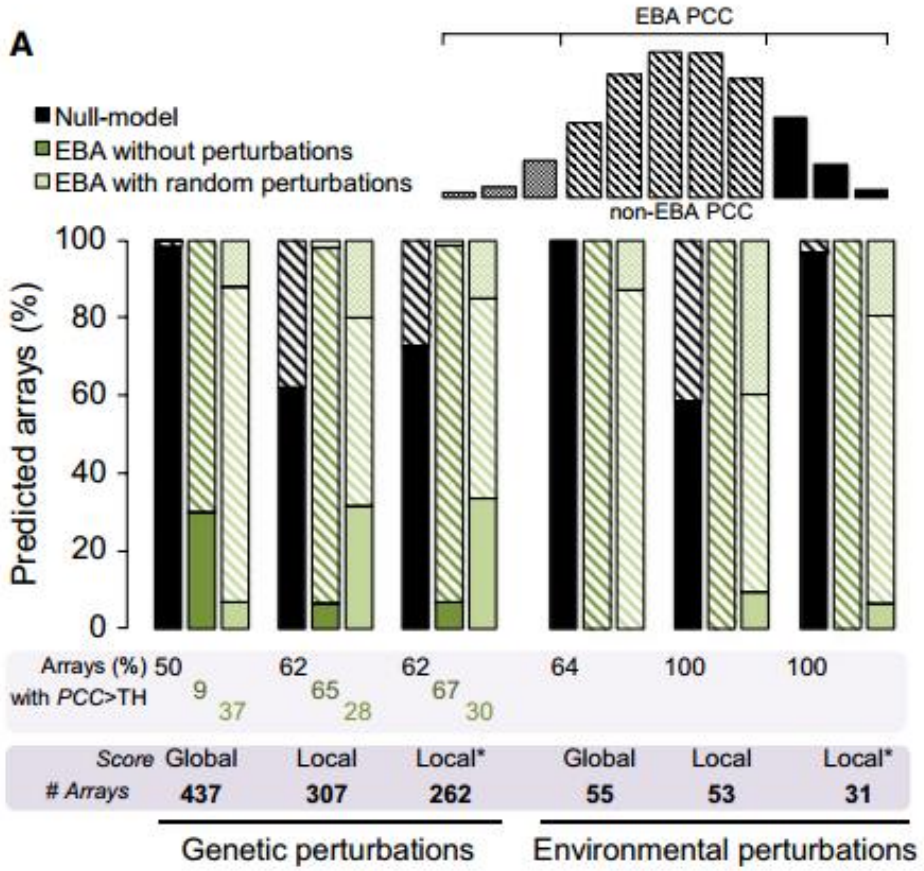
Genetic and environmental gene expression diversity

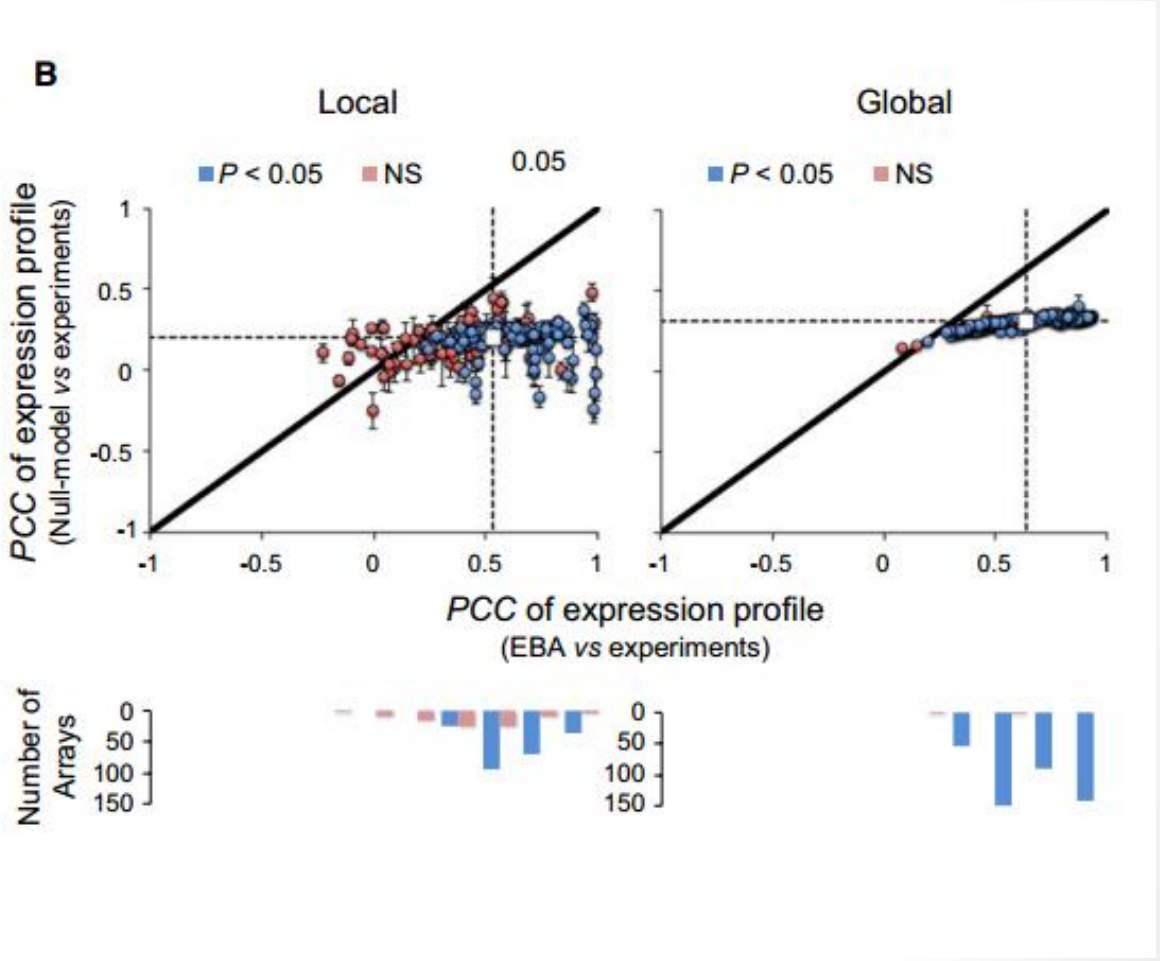


An integrative knowledgebase as a base to regulatory network enrichment

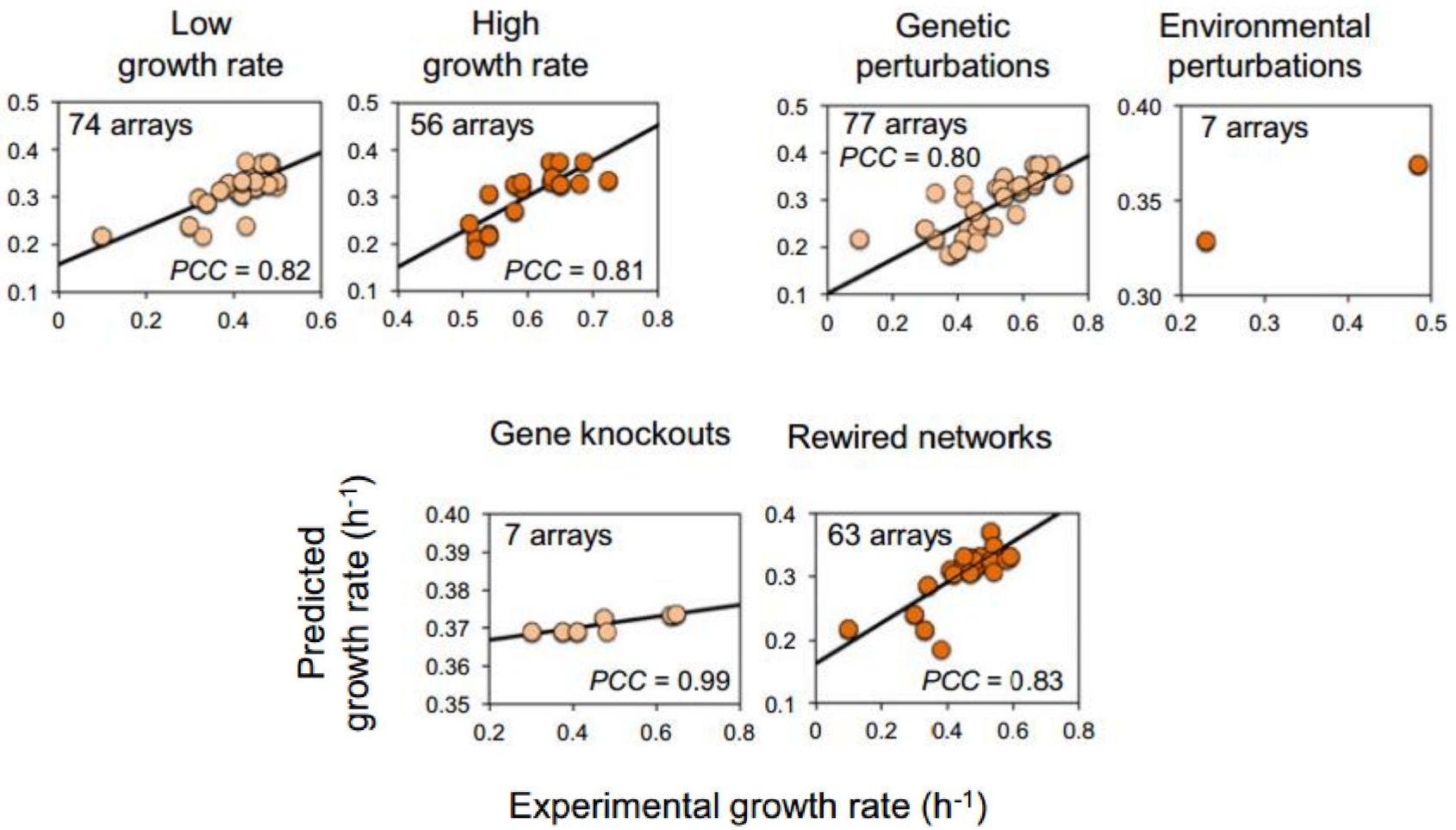


Expression Balance Analysis

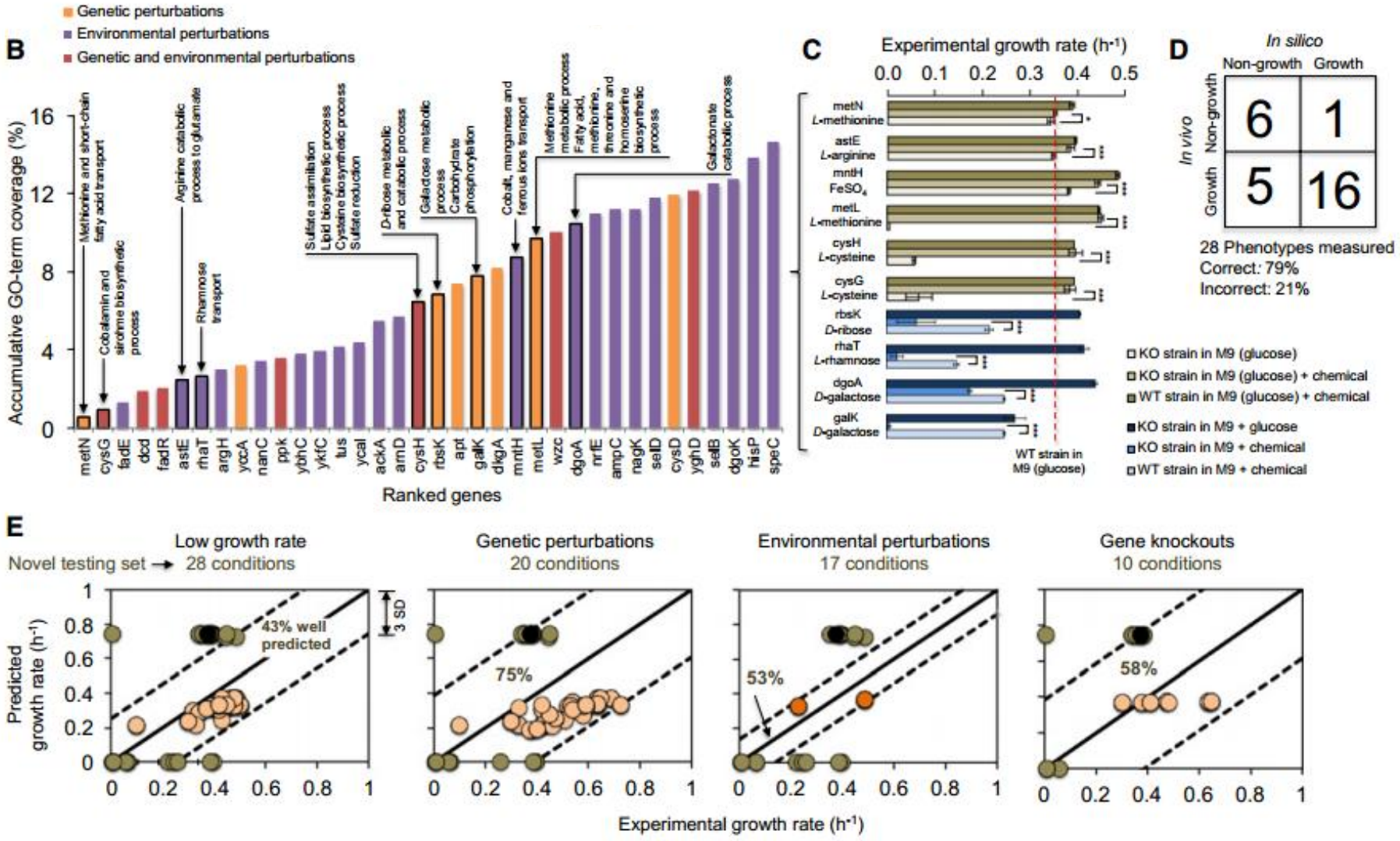




Phenotypic predictions in an integrated model



Model enrichment through targeted experimentation



Discussion

- Advantage
  - the creation of a signal transduction network (EcoST)
  - its integration to the transcriptional and metabolic network through constraint modeling

- Disadvantage
  - Coverage
  - the severe bias to negative samples in the ground truth

Inspiration

- Binding site
- Funtion

Thanks for your attention