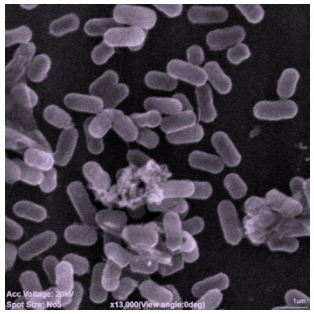
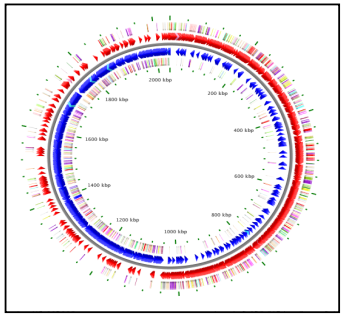
Adding Power to Phenomic Studies:
Novel Statistical Approaches for PM Data

Joseph Sturino, Ph.D.
Nutrition and Food Science Department
Texas A&M University





Sodium Chloride Toxicity

Biocatalysts used to produce novel compounds

- Chemostat fermentation (Batch):

  ACID ACCUMULATION → Toxicity

- Adjust pH (NaOH)

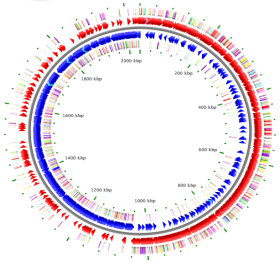
  Acid Neutralized, but Toxicity Still Observed

Hypothesized: Acid + Base = NaCl Accumulation → **Osmotic Toxicity**

Used comparative phenomics to investigate

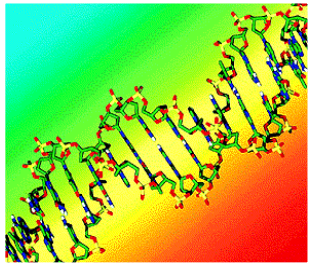
Functional Genomic Approach

genomics

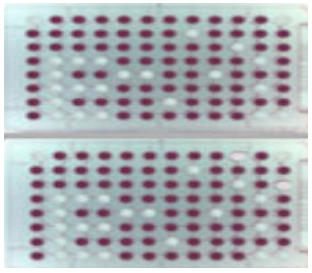


hypothesis

molecular biology



phenomics



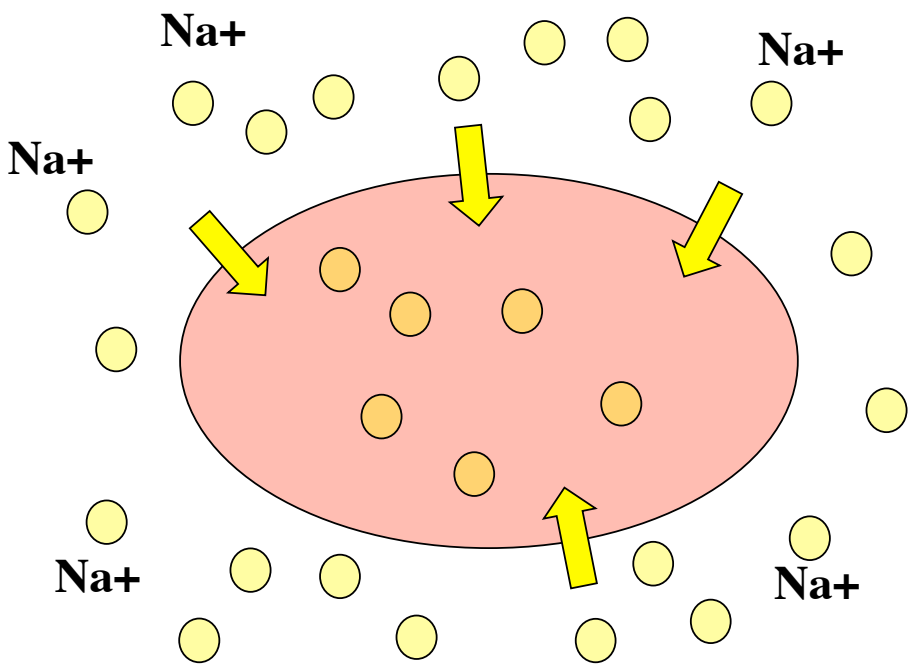
PM09: Osmolytes

A1 NaCl 1%	A2 NaCl 2%	A3 NaCl 3%	A4 NaCl 4%	A5 NaCl 5%	A6 NaCl 5.5%	A7 NaCl 6%	A8 NaCl 6.5%	A9 NaCl 7%	A10 NaCl 8%	A11 NaCl 9%	A12 NaCl 10%
B1 NaCl 6%	B2 NaCl 6% + Betaine	B3 NaCl 6% + N-N Dimethyl glycine	B4 NaCl 6% + Sarcosine	B5 NaCl 6% + Dimethyl sulphonyl propionate	B6 NaCl 6% + MOPS	B7 NaCl 6% + Ectoine	B8 NaCl 6% + Choline	B9 NaCl 6% + Phosphoryl choline	B10 NaCl 6% + Creatine	B11 NaCl 6% + Creatinine	B12 NaCl 6% + L- Carnitine
C1 NaCl 6% + KCl	C2 NaCl 6% + L-proline	C3 NaCl 6% + N-Acetyl L-glutamine	C4 NaCl 6% + β-Glutamic acid	C5 NaCl 6% + γ–Amino -n- butyric acid	C6 NaCl 6% + Glutathione	C7 NaCl 6% + Glycerol	C8 NaCl 6% + Trehalose	C9 NaCl 6% + Trimethylamine -N-oxide	C10 NaCl 6% + Trimethylamine	C11 NaCl 6% + Octopine	C12 NaCl 6% + Trigonelline
D-1 Potassium chloride 3%	D2 Potassium chloride 4%	D3 Potassium [illegible]	D4 Potassium Chloride [illegible]	D5 Sodium sulfate 2%	D6 Sodium sulfate 3%	D7 Sodium sulfate 4%	D8 Sodium sulfate 5%	D9 Ethylene glycol 5%	D10 Ethylene glycol 10%	D11 Ethylene glycol 15%	D12 Ethylene glycol 20%
E1 Sodium formate 1%	E2 Sodium formate 2%	E3 Sodium [illegible] 3%	E4 Sodium [illegible] 4%	E5 Sodium formate 5%	E6 Sodium [illegible] 6%	E7 Urea 2%	E8 Urea 3%	E9 Urea 4%	E10 Urea 5%	E11 Urea 6%	E12 Urea 7%
F1 Sodium Lactate 1%	F2 Sodium Lactate 2%	F3 Sodium Lactate 3%	F4 Sodium Lactate 4%	F5 Sodium Lactate 5%	F6 Sodium Lactate 6%	F7 Sodium Lactate 7%	F8 Sodium Lactate 8%	F9 Sodium Lactate 9%	F10 Sodium Lactate 10%	F11 Sodium Lactate 11%	F12 Sodium Lactate 12%
G1 Sodium Phosphate pH 7 20mM	G2 Sodium Phosphate pH 7 50mM	G3 Sodium Phosphate pH 7 100mM	G4 Sodium Phosphate pH 7 200mM	G5 Sodium Benzoate pH 5.2 20mM	G6 Sodium Benzoate pH 5.2 50mM	G7 Sodium Benzoate pH5.2 100mM	G8 Sodium Benzoate pH 5.2 200mM	G9 Ammonium sulfate pH8 10mM	G10 Ammonium sulfate pH 8 20mM	G11 Ammonium sulfate pH 8 50mM	G12 Ammonium sulfate pH8 100mM
H1 Sodium Nitrate 10mM	H2 Sodium Nitrate 20mM	H3 Sodium Nitrate 40mM	H4 Sodium Nitrate 60mM	H5 Sodium Nitrate 80mM	H6 Sodium Nitrate 100mM	H7 Sodium Nitrite 10mM	H8 Sodium Nitrite 20mM	H9 Sodium Nitrite 40mM	H10 Sodium Nitrite 60mM	H11 Sodium Nitrite 80mM	H12 Sodium Nitrite 100mM

Osmotic Stress Model System

Escherichia coli

Wild-type EC25113

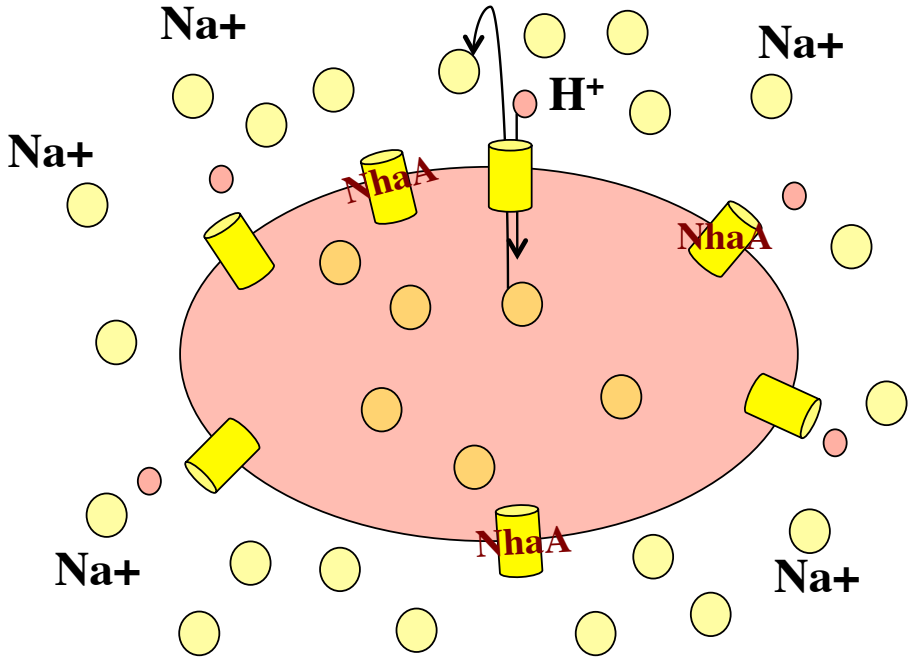


Wood. 2006. Osmosensing by Bacteria *Sci. STKE* (357): 43

Model System

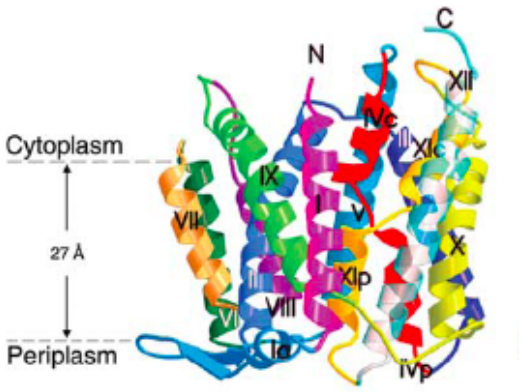
Escherichia coli

Wild-type EC25113

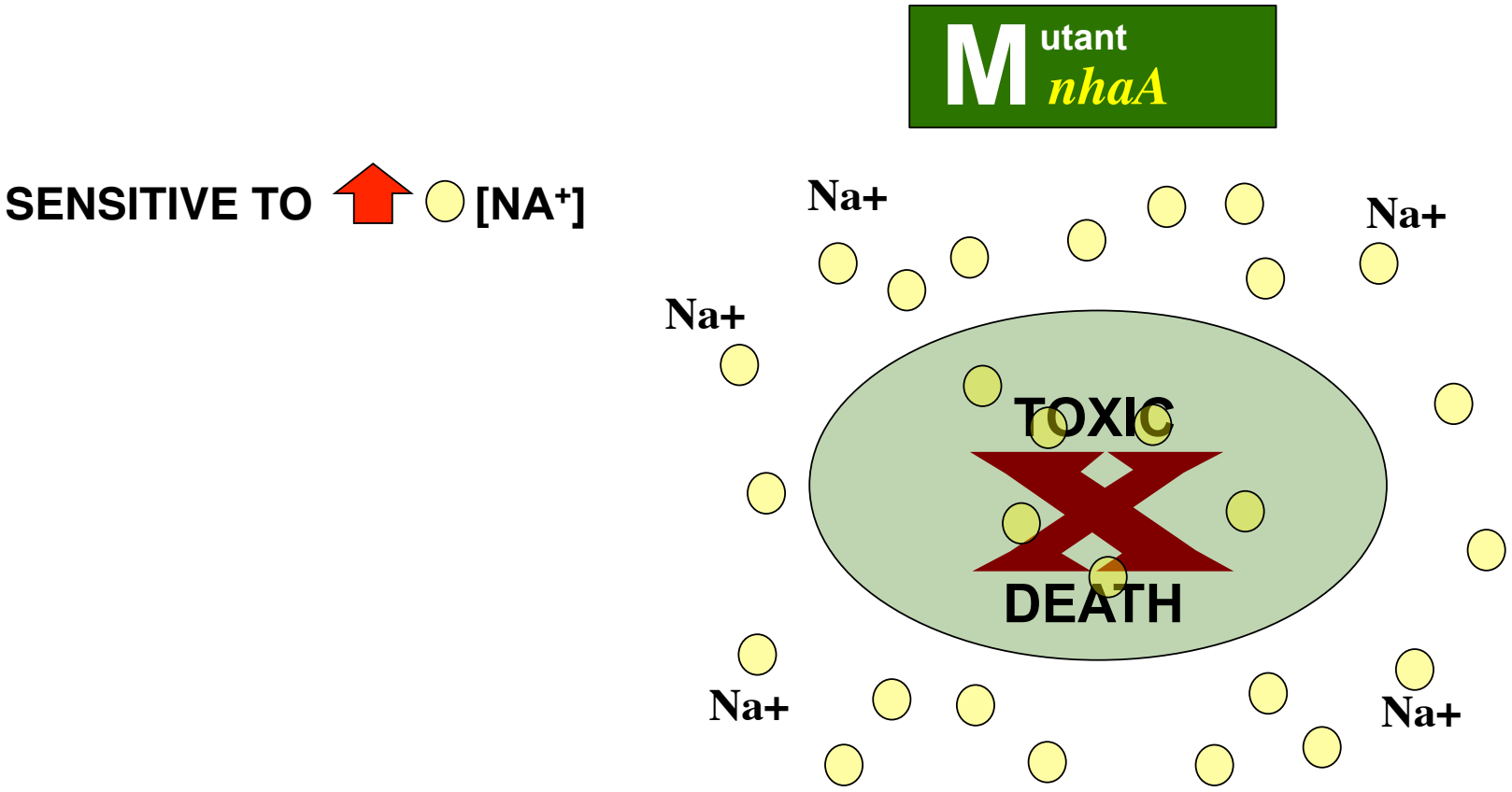


NhaA

Primary (Na^+-H^+) Antiporter



Model System



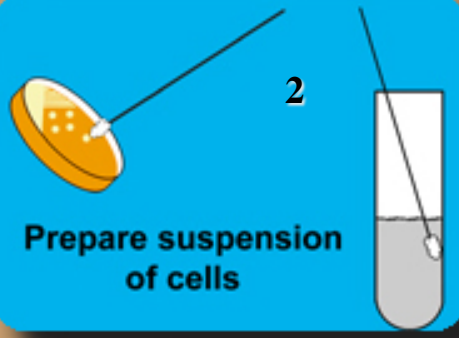
The Assay

Overview

1

Grow cells

2



Prepare suspension of cells

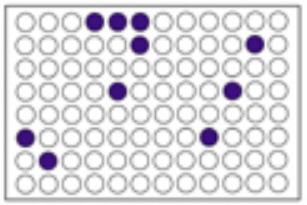
3

Inoculate MicroPlates

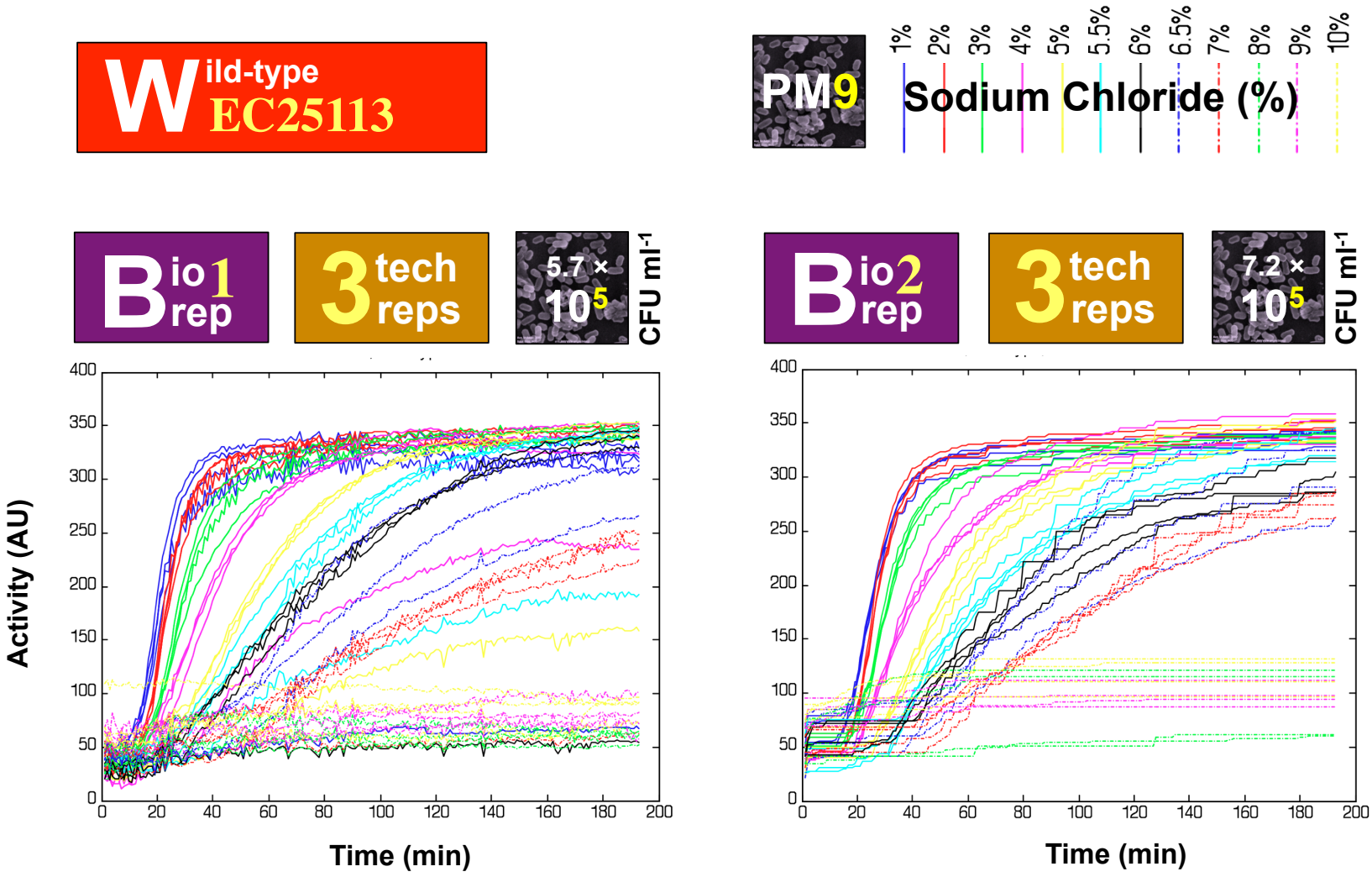
4

Incubation

5

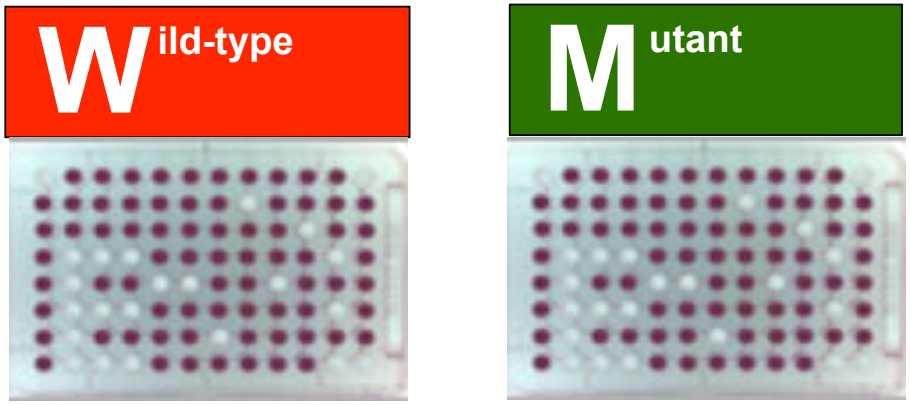


Interpretation



Robust, but inter-replicate differences found

Parametric Module



- **Compare two microorganisms**
- **Calculates summary values**
  - Area under the curve
  - Min/maximum signal intensity
  - Maximum slope
  - Lag time

Does Not Include Hypothesis Testing

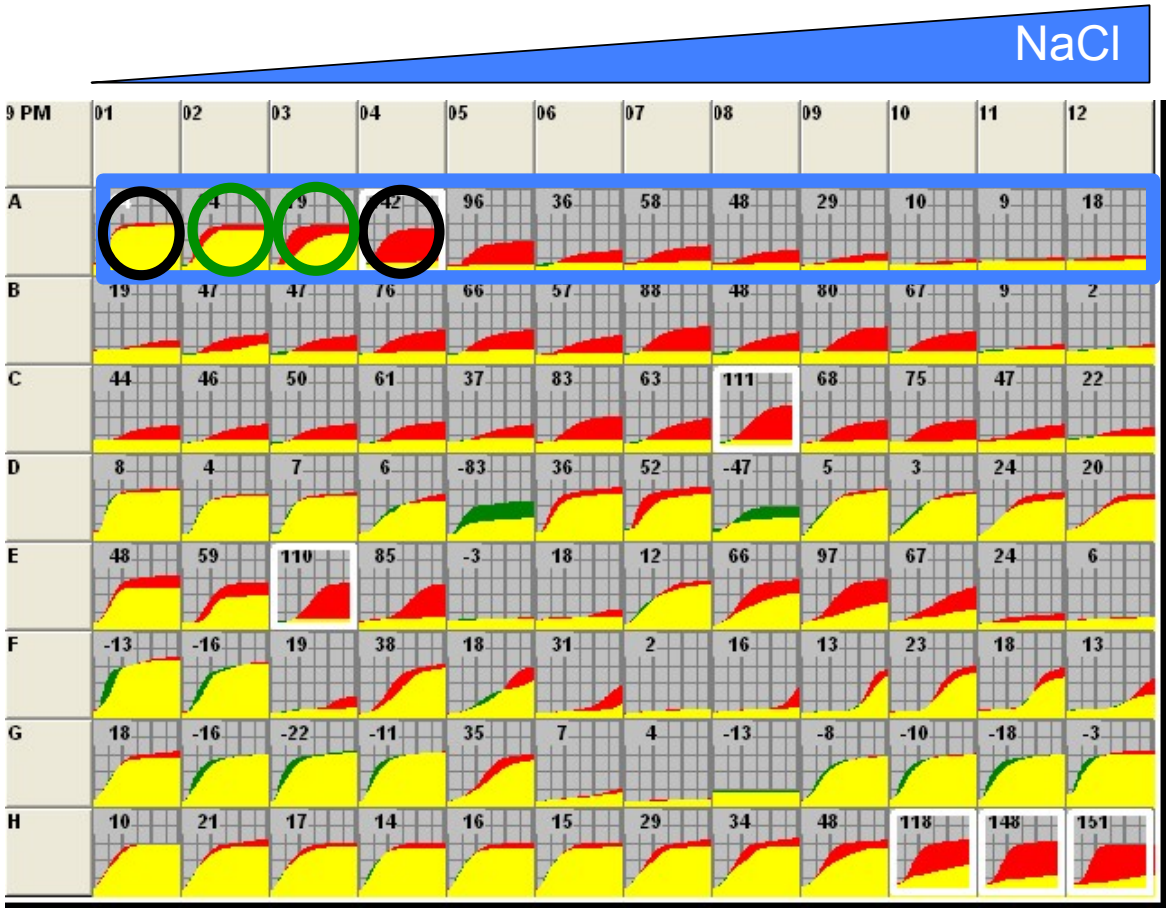
H_0 NULL H_1 ALT

Gradient Differences Unclear

Wild-type EC25113

Mutant *nhaA*

OVERLAP BOTH



Two strains under same conditions

W

WM

?

Study Objective

Dr. Raymond Carroll
Texas A&M University

Develop simple but robust statistical methodologies

Enable sound biological inferences from PM data

The International Journal of Biostatistics

Volume 6, *Issue* 1 2010 *Article* 29

Sturino *et al.*, 2010

Statistical Methods for Comparative Phenomics Using High-Throughput Phenotype Microarrays

Dr. Ivan Zorych
Columbia University

Dr. Nikolay Bliznyuk
University of Florida

(www.r-project.org)

Hypothesis Testing

Permutation-based statistics:

H_0 NULL

$f1(t) = f2(t)$ for all t

2 Organisms have exchangeable phenotypes

H_1 ALT

$f1(t) \neq f2(t)$ for all t

Nonparametric Permutation

$Y_{ij}(t)$ defined as observed phenotype curve for:

Organism $i = 1, ..., I = 2$ and

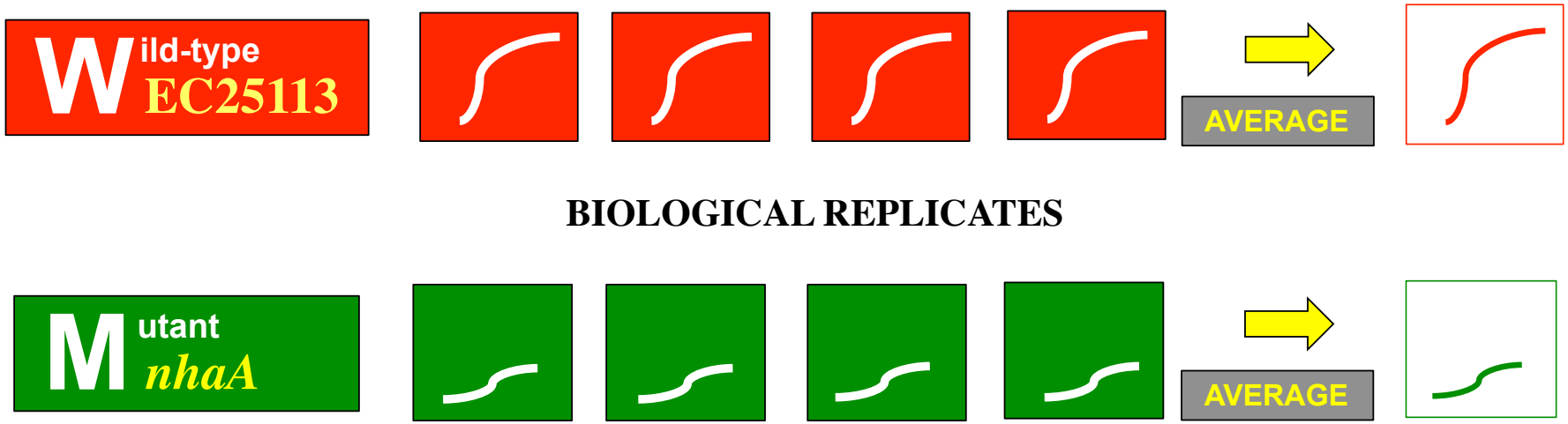
Replicate $j = 1, ..., J$,

Sample curve (mean) in organism i is:

MEAN CURVES

$$\hat{f}_i(t) = \mathrm{J}^{-1} \sum_{\mathrm{j}=1}^{\mathrm{J}} \{\mathrm{Y}_{\mathrm{ij}}(\mathrm{t})\}$$

Nonparametric Permutation



Calculate overall ***squared difference*** between mean curves:

$$S = \int \{\hat{f}_{1\cdot}(t) - \hat{f}_{2\cdot}(t)\}^2 dt.$$

Permutation Device

Take all permutations of the indices (i, j) such that:

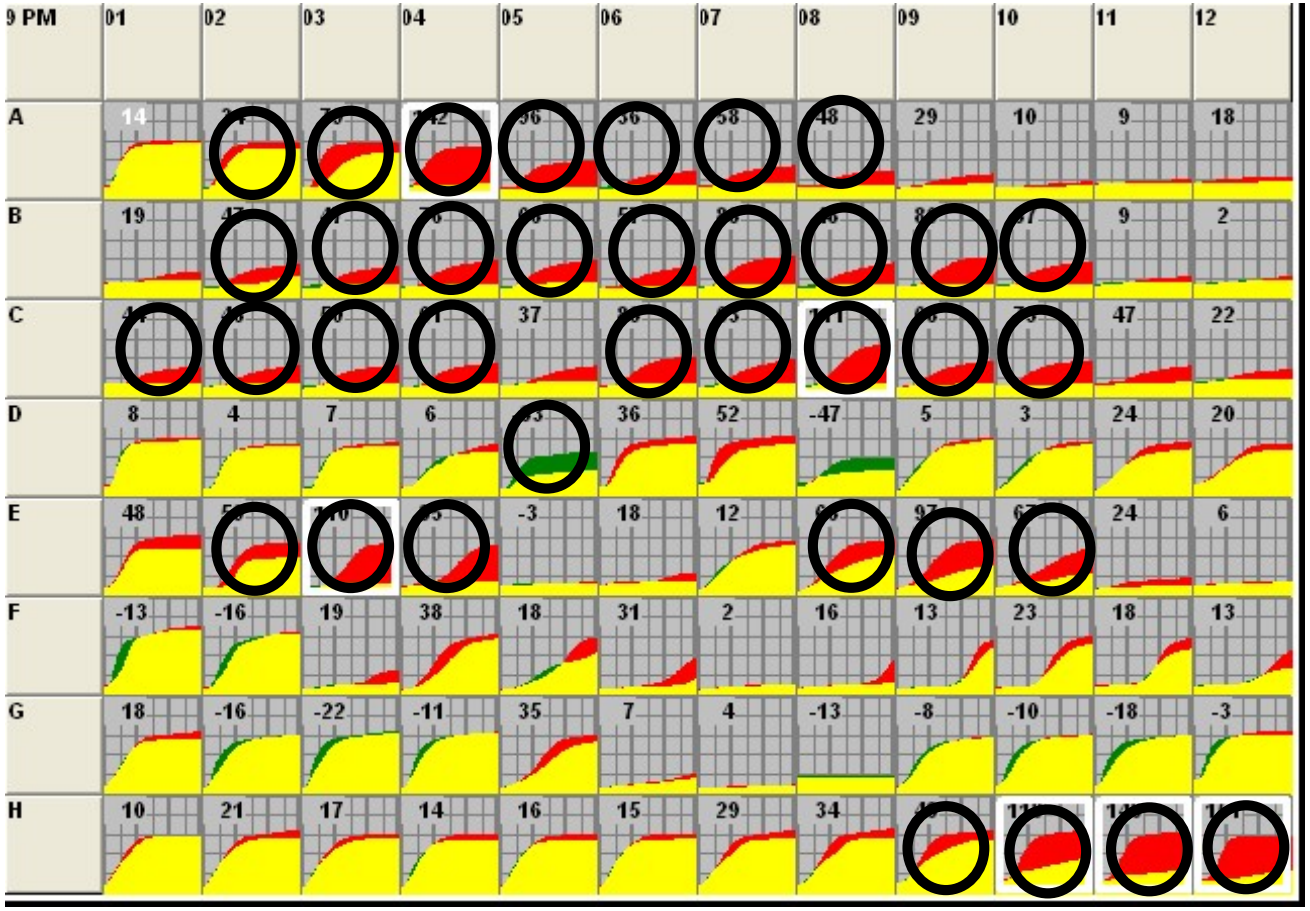
Let the number of unique permutations be **B**.

Then for each of the b = 1, ..., B unique permutations,

Recompute the test statistic (1) and record it as S_b.

p-value: $$p = B^{-1} \sum_{b=1}^{B} I(\mathcal{S}_b > \mathcal{S}).$$

Effective Differentiation



○ P < 0.05

Alternative Tests

MEDIAN CURVES

$$\tilde{f}_i(t) = \text{median}_{j \in J}\{Y_{ij}(t)\}$$

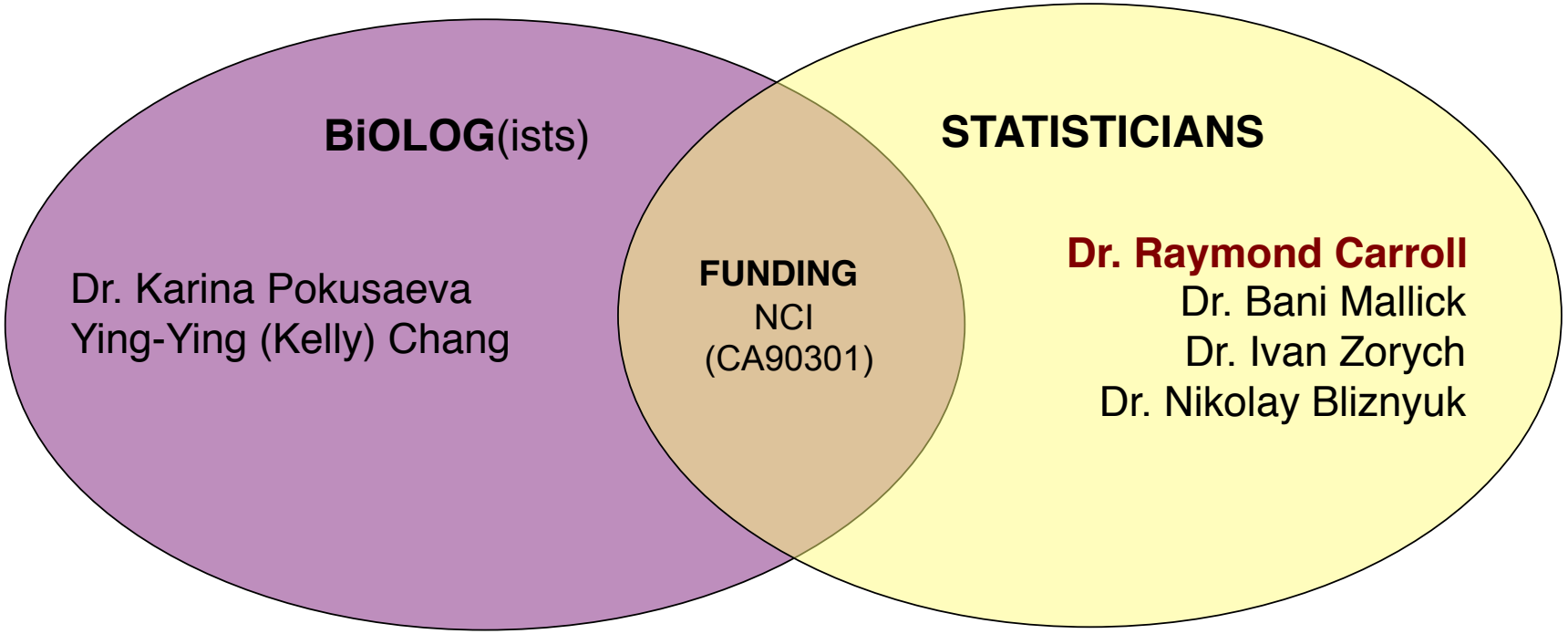
When error-adjusted, test can be carried out as with mean curves

Effective for high data skewness

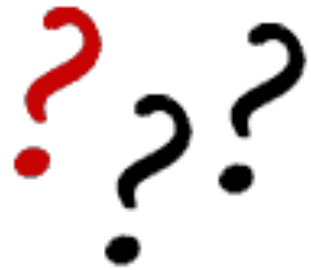
CURVES AREA

fPCA

Acknowledgments



joseph.sturino@tamu.edu



Questions