

# Construction of a Genome-Wide Regulatory Network for the Metal Reducing Microbe *Shewanella oneidensis* MR-1

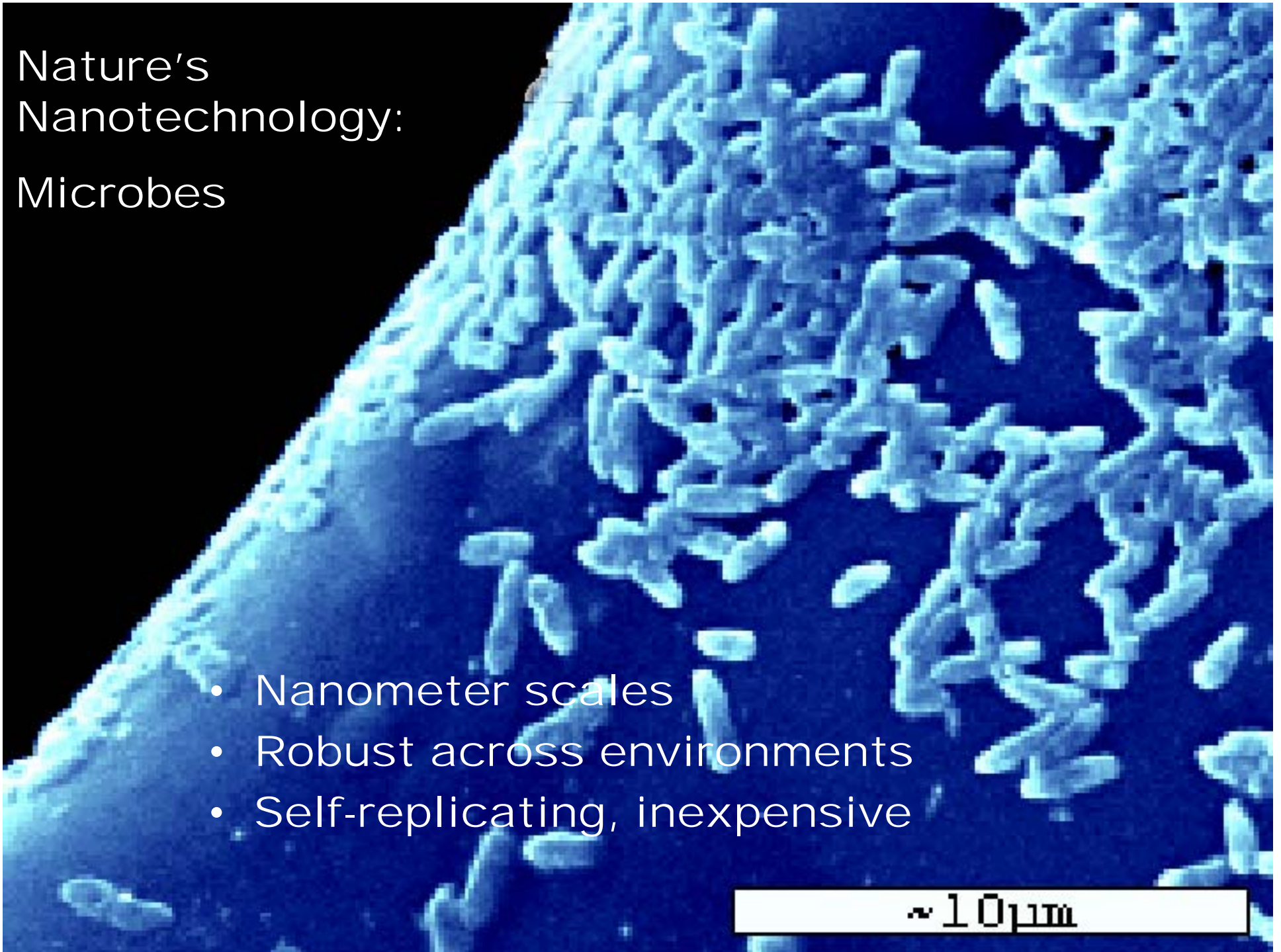
Michael Driscoll  
Bioinformatics Program  
Boston Univeristy  
CSGF Conference 2006



Nature's  
Nanotechnology:  
Microbes

- Nanometer scales
- Robust across environments
- Self-replicating, inexpensive

$\sim 10\mu\text{m}$



# Synthetic Biology (I): Engineering Microbes

- H. Boyer → *E.coli* to produce insulin (1978)
- A. Chakrabarty → a petroleum eating *Pseudomonas* bacterium (1980)
- J. Keasling → *E.coli* to produce artemisinin, an anti-malarial (2006)

Synthetic Biology (II):  
Engineering  
*Shewanella*  
*oneidensis* MR-1

A metal-breathing microbe

- Bioremediation
- Bacterial fuel cells
- Bacterial sensors

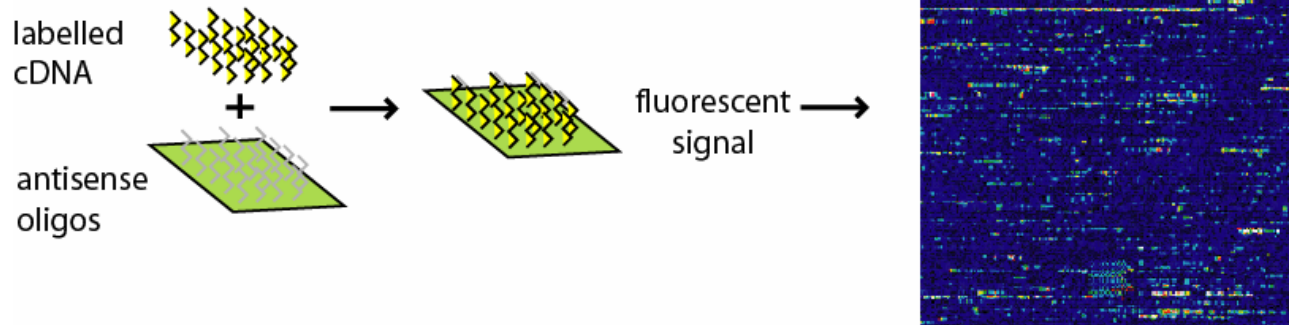
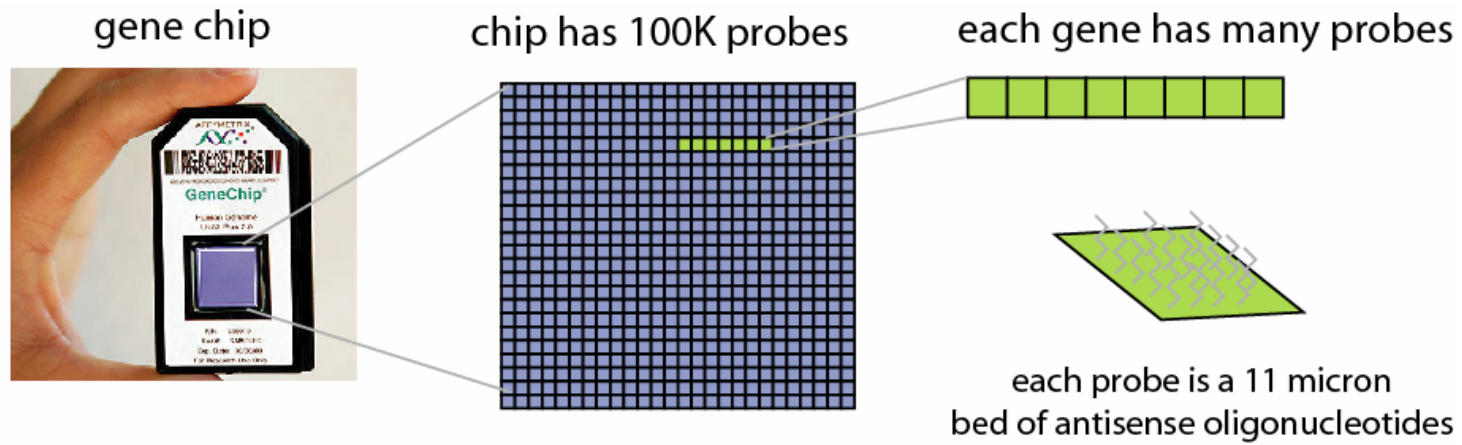


# Synthetic Biology (II): Engineering *Shewanella* *oneidensis* MR-1

- identify gene networks involved in metal reduction
- build a model for genome regulatory network
- observe genome-wide network in action, with microarrays

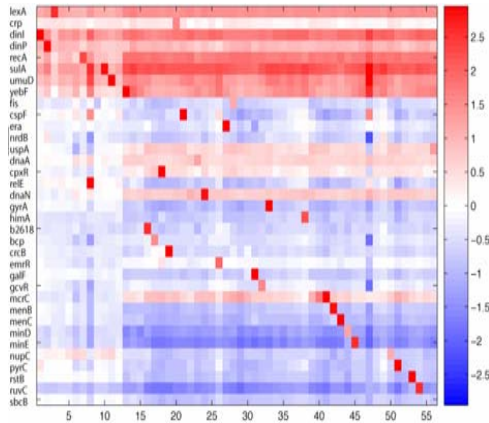


# *Shewanella* MR-1 Affymetrix chip



# Learning networks from expression data

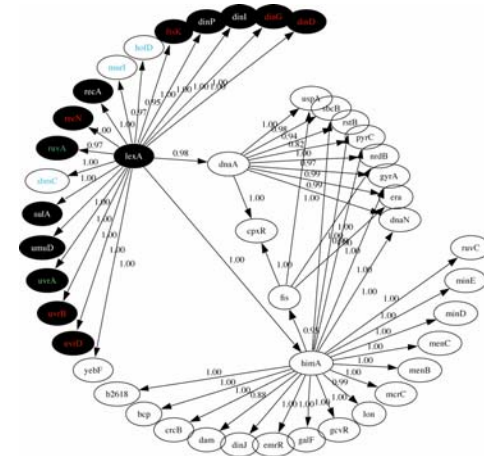
expression & genomic data



machine learning algorithms



regulatory network model

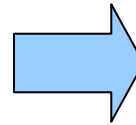
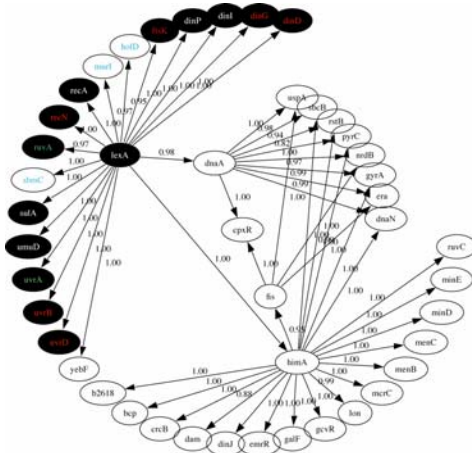


Physical Models

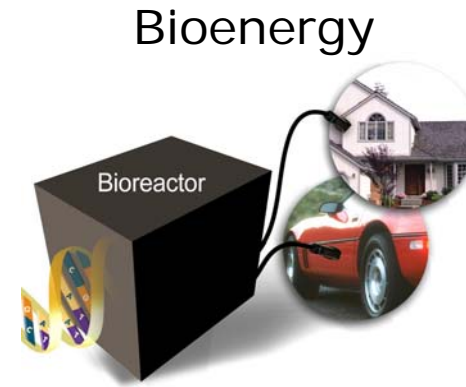
Influence Models

# Shewanella Project: Map genes regulating respiration

1) Map the genes regulating metal reduction and anaerobic respiration

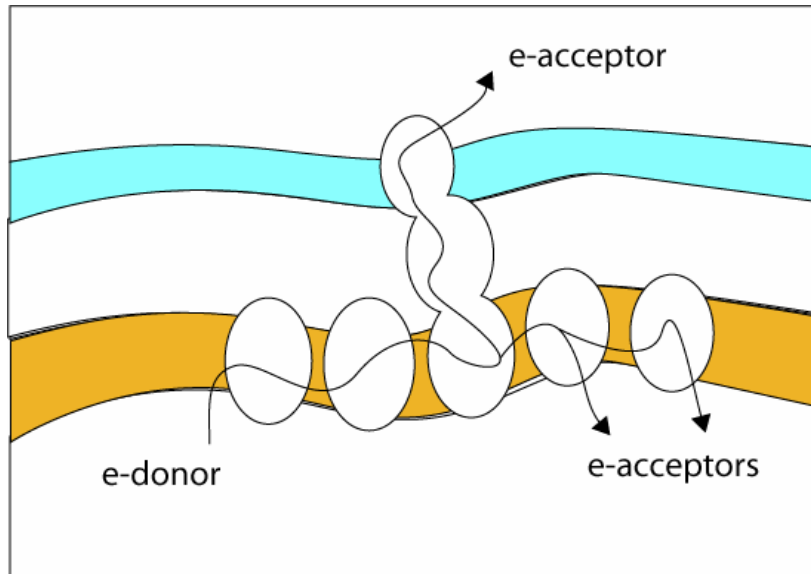


2) Engineer *Shewanella's* respiratory pathways for bioremediation and microbial fuel cells

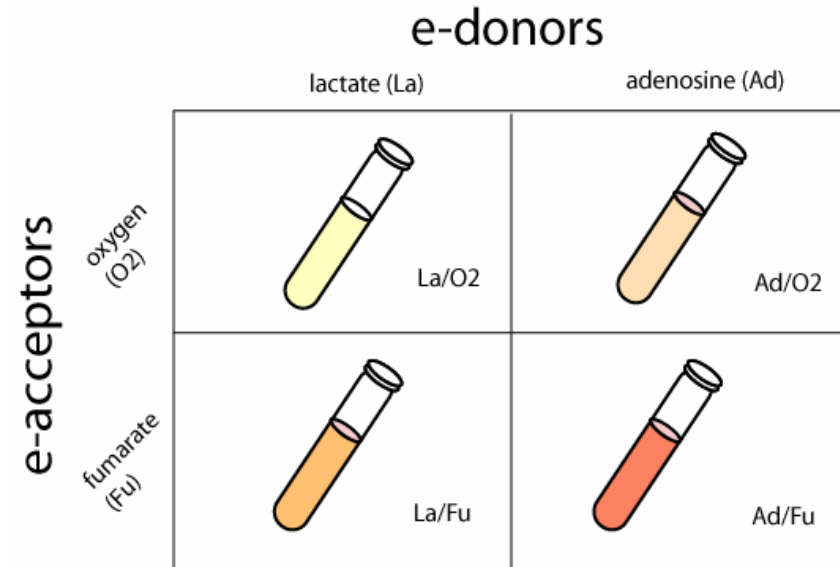




# Reconstructing *Shewanella*'s transcriptional regulatory networks

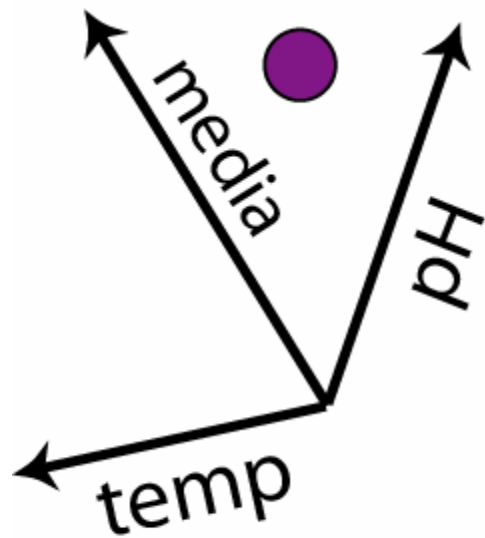


***Shewanella* uses a variety of e-donors and e-acceptors**



**For multiple e-donors and e-acceptors were profiled**

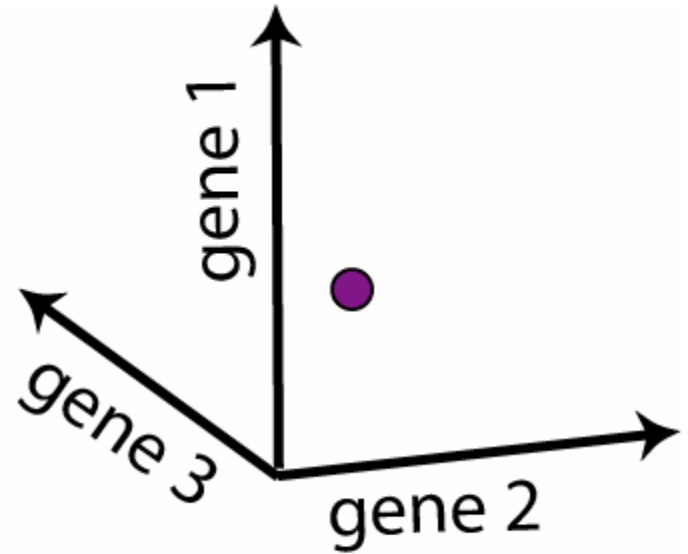
finding conditions which generate the  
“most informative” expression states



condition space

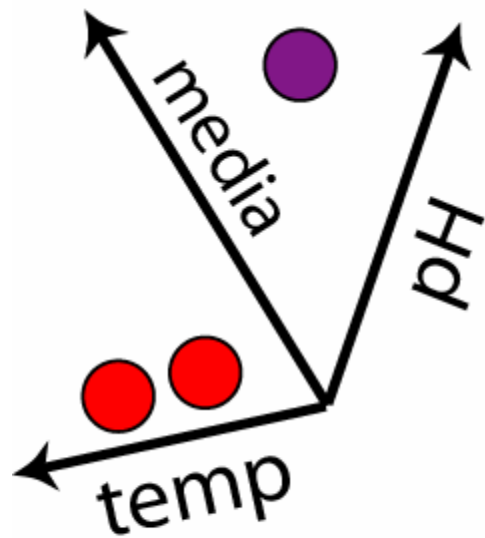
$$f(\cdot)$$

A horizontal arrow pointing from the condition space to the expression space, with the function notation  $f(\cdot)$  above it.



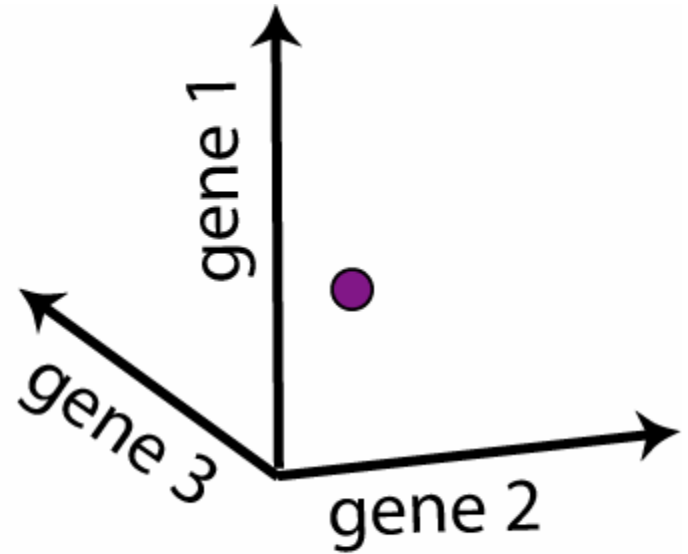
expression space

finding conditions which generate the “most informative” expression states



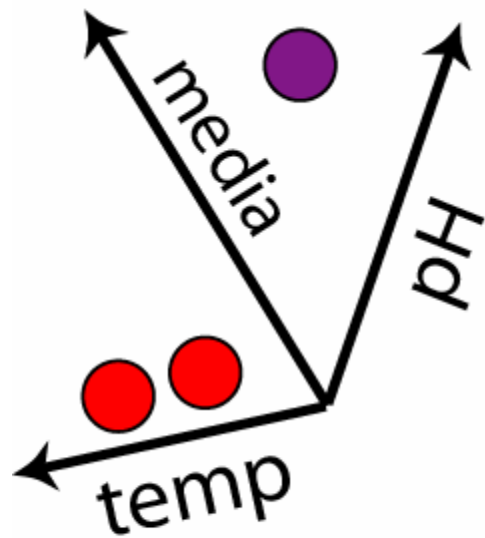
condition space

$$f(\cdot)$$



expression space

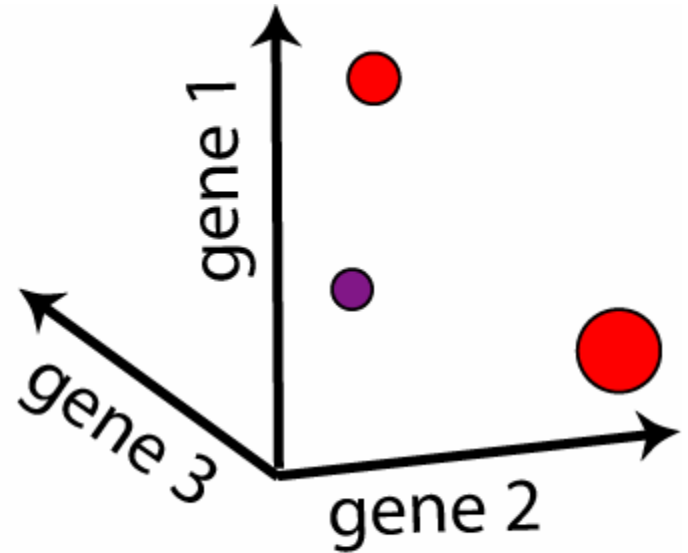
finding conditions which generate the “most informative” expression states



condition space

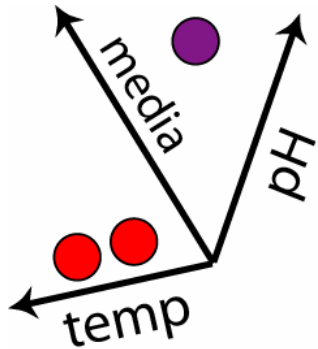
$$f(\cdot)$$

A horizontal arrow pointing from the condition space plot to the expression space plot, with the function notation  $f(\cdot)$  written above it.



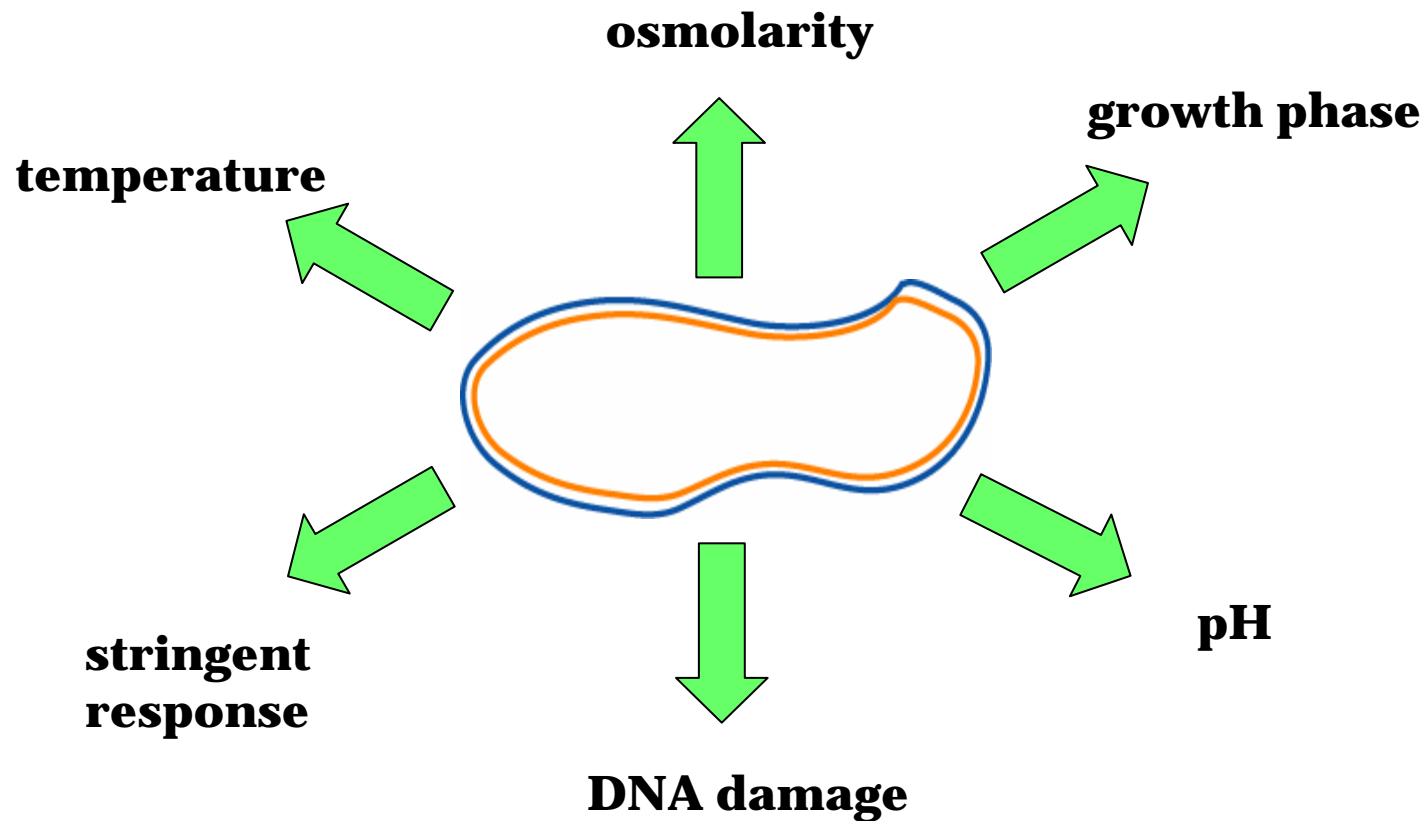
expression space

# experimental approach



- (I) determine range of condition to explore based on literature
- (II) perform growth profiles on this range to check for viable growth
- (III) choose four to eight values to profile
- (IV) grow cells at each value, harvest, and run expression profile
- (V) analyze expression profiles and choose “most informative” values for further experiments

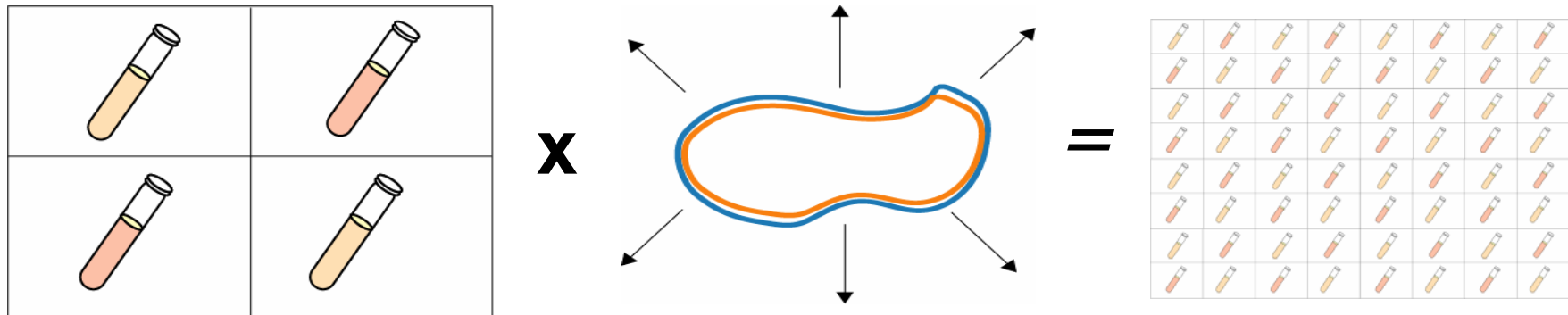
# environmental perturbations



Environmental perturbations have been selected based on our experience with *E.coli* in identifying which conditions were most informative in the reconstruction of the global transcriptional network.

# environmental perturbations

**Challenge: High-dimensionality of condition space**



**media conditions  
containing  $K$  e-donors,  
and  $L$  e-acceptors**

**$M$  2-level  
environmental  
perturbations**

**$K \times L \times 2^M$   
experimental  
conditions**

**Solution: Partial Factorial Design**

**This approach isolates the *main effects* of variables. Full independence assumption allows exploration of sample space with  $K + L + M$  experiments; assuming at most 2 interacting variables allows for  $(K+L+M)(K+L+M-1)/2$  experiments.**

# e-donors & e-acceptors: growth profiling

Biolog chip data for *Shewanella*

<b>activity</b>	<b>carbon source</b>	<b>type</b>
247	Adenosine	C-Source, nucleoside
245	Inosine	C-Source, nucleoside
239	Uridine	C-Source, nucleoside
221	L-Lactic Acid	C-Source, carboxylate
208	Pyruvic Acid	C-Source, carboxylate
207	Methyl Pyruvate	C-Source, carboxylate
198	N-Acetyl-D-Glucosamine	C-Source, carbohydrate
166	Glycyl-L-Glutamic Acid	C-Source, amino acid
164	2'-Deoxy Adenosine	C-Source, nucleoside
140	Glycyl-L-Aspartic Acid	C-Source, amino acid
70	Thymidine	C-Source, nucleoside
64	$\alpha$ -Ketobutyric Acid	C-Source, carboxylate
61	Tween 20	C-Source
48	Tween 40	C-Source



TABLE 1 The Fnr and Arc modulons<sup>a</sup>

Gene(s)	Targets	Effects of:	
	Enzyme and/or function	Fnr <sup>b</sup>	ArcA
<i>aceB</i>	Isocitrate lyase	ND	(-)
<i>acn</i>	Aconitase	ND	(-)
<i>aeg-46.5</i>	Putative periplasmic nitrate reductase	+	0
<i>ansB</i>	L-Asparaginase II	+	ND
<i>arcA</i>	ArcA	+	+
<i>aspA-dcuA</i>	L-Aspartase and dicarboxylate transport	+	ND
<i>cea</i>	Colicin E1	+	ND
<i>cob</i>	Cobalamin biosynthesis	ND	+
<i>cydAB</i>	Cytochrome <i>d</i> oxidase	-	+
<i>cyoABCDE</i>	Cytochrome <i>o</i> oxidase	-	-
<i>dmsABC</i>	Dimethyl sulfoxide reductase	+	0
<i>dcuB-fumB</i>	Dicarboxylate transport and fumarase B	+	ND
<i>fadB</i>	3-Hydroxyacyl coenzyme A	ND	(-)
<i>fdnGHI</i>	Formate dehydrogenase-N	+	(-)
<i>feoAB</i>	Iron (II) transport	+	ND
<i>focA-pfl</i>	Formate transport and pyruvate-formate lyase	+	+
<i>fnr</i>	Fnr	-	ND
<i>fumA</i>	Fumarase A (aerobic)	(0)	(-)
<i>frdABCD</i>	Fumarate reductase	+	(0)
<i>glpACB</i>	Glycerol-3-phosphate dehydrogenase (anaerobic)	+	0
<i>glpD</i>	Glycerol-3-phosphate dehydrogenase (aerobic)	0	-
<i>glpTQ</i>	Glycerol 3-phosphate transport	+	ND
<i>gltA</i>	Citrate synthase	ND	-
<i>hemA</i>	Glutamyl-tRNA dehydrogenase	-	+
<i>hyaA-F</i>	Hydrogenase 1	0	+
<i>hypBCDE-fhlA</i>	Hydrogenase activities and formate regulation	+	ND
<i>icd</i>	Isocitrate dehydrogenase	ND	(-)
<i>ild(lctD)</i>	L-Lactate dehydrogenase	0	-
<i>mdh</i>	Malate dehydrogenase	ND	(-)
<i>narGHJI</i>	Nitrate reductase	+	0
<i>narK</i>	Nitrite extrusion protein	+	0
<i>narX</i>	NarX sensor protein	+	ND
<i>ndh</i>	NADH dehydrogenase II (aerobic)	-	ND
<i>nikA-E (hydC)</i>	Nickel transport	+	ND
<i>nirBDC</i>	NADH-dependent nitrite reductase	+	ND

blue = *fnr* target in  
Shewanella (by  
homology)

TABLE 1 The Fnr and Arc modulons<sup>a</sup>

Gene(s)	Targets	Effects of:	
	Enzyme and/or function	Fnr <sup>b</sup>	ArcA
<i>aceB</i>	Isocitrate lyase	ND	(-)
<i>acn</i>	Aconitase	ND	(-)
<i>aeg-46.5</i>	Putative periplasmic nitrate reductase	+	0
<i>ansB</i>	L-Asparaginase II	+	ND
<i>arcA</i>	ArcA	+	+
<i>aspA-dcuA</i>	L-Aspartase and dicarboxylate transport	+	ND
<i>cea</i>	Colicin E1	+	ND
<i>cob</i>	Cobalamin biosynthesis	ND	+
<i>cydAB</i>	Cytochrome <i>d</i> oxidase	-	+
<i>cyoABCDE</i>	Cytochrome <i>o</i> oxidase	-	-
<i>dmsABC</i>	Dimethyl sulfoxide reductase	+	0
<i>dcuB-fumB</i>	Dicarboxylate transport and fumarase B	+	ND
<i>fadB</i>	3-Hydroxyacyl coenzyme A	ND	(-)
<i>fdnGHI</i>	Formate dehydrogenase-N	+	(-)
<i>feoAB</i>	Iron (II) transport	+	ND
<i>focA-pfl</i>	Formate transport and pyruvate-formate lyase	+	+
<i>fnr</i>	Fnr	-	ND
<i>fumA</i>	Fumarase A (aerobic)	(0)	(-)
<i>frdABCD</i>	Fumarate reductase	+	(0)
<i>glpACB</i>	Glycerol-3-phosphate dehydrogenase (anaerobic)	+	0
<i>glpD</i>	Glycerol-3-phosphate dehydrogenase (aerobic)	0	-
<i>glpTQ</i>	Glycerol 3-phosphate transport	+	ND
<i>gltA</i>	Citrate synthase	ND	-
<i>hemA</i>	Glutamyl-tRNA dehydrogenase	-	+
<i>hyaA-F</i>	Hydrogenase 1	0	+
<i>hypBCDE-fhlA</i>	Hydrogenase activities and formate regulation	+	ND
<i>icd</i>	Isocitrate dehydrogenase	ND	(-)
<i>ild(lctD)</i>	L-Lactate dehydrogenase	0	-
<i>mdh</i>	Malate dehydrogenase	ND	(-)
<i>narGHJI</i>	Nitrate reductase	+	0
<i>narK</i>	Nitrite extrusion protein	+	0
<i>narX</i>	NarX sensor protein	+	ND
<i>ndh</i>	NADH dehydrogenase II (aerobic)	-	ND
<i>nikA-E (hydC)</i>	Nickel transport	+	ND
<i>nirBDC</i>	NADH-dependent nitrite reductase	+	ND

TABLE 1 The Fnr and Arc modulons<sup>a</sup>

Gene(s)	Targets Enzyme and/or function	Effects of:	
		Fnr <sup>b</sup>	ArcA
<i>aceB</i>	Isocitrate lyase	ND	(-)
<i>acn</i>	Aconitase	ND	(-)
<i>aeg-46.5</i>	Putative periplasmic nitrate reductase	+	0
<i>ansB</i>	L-Asparaginase II	+	ND
<i>arcA</i>	ArcA	+	+
<i>aspA-dcuA</i>	L-Aspartase and dicarboxylate transport	+	ND
<i>cea</i>	Colicin E1	+	ND
<i>cob</i>	Cobalamin biosynthesis	ND	+
<i>cydAB</i>	Cytochrome <i>d</i> oxidase	-	+
<i>cyoABCDE</i>	Cytochrome <i>o</i> oxidase	-	-
<i>dmsABC</i>	Dimethyl sulfoxide reductase	+	0
<i>dcuB-fumB</i>	Dicarboxylate transport and fumarase B	+	ND
<i>fadB</i>	3-Hydroxyacyl coenzyme A	ND	(-)
<i>fdnGHI</i>	Formate dehydrogenase-N	+	(-)
<i>feoAB</i>	Iron (II) transport	+	ND
<i>focA-pfl</i>	Formate transport and pyruvate-formate lyase	+	+
<i>fnr</i>	Fnr	-	ND
<i>fumA</i>	Fumarase A (aerobic)	(0)	(-)
<i>frdABCD</i>	Fumarate reductase	+	(0)
<i>glpACB</i>	Glycerol-3-phosphate dehydrogenase (anaerobic)	+	0
<i>glpD</i>	Glycerol-3-phosphate dehydrogenase (aerobic)	0	-
<i>glpTQ</i>	Glycerol 3-phosphate transport	+	ND
<i>gltA</i>	Citrate synthase	ND	-
<i>hemA</i>	Glutamyl-tRNA dehydrogenase	-	+
<i>hyaA-F</i>	Hydrogenase 1	0	+
<i>hypBCDE-fhlA</i>	Hydrogenase activities and formate regulation	+	ND
<i>icd</i>	Isocitrate dehydrogenase	ND	(-)
<i>ild(lctD)</i>	L-Lactate dehydrogenase	0	-
<i>mdh</i>	Malate dehydrogenase	ND	(-)
<i>narGHJI</i>	Nitrate reductase	+	0
<i>narK</i>	Nitrite extrusion protein	+	0
<i>narX</i>	NarX sensor protein	+	ND
<i>ndh</i>	NADH dehydrogenase II (aerobic)	-	ND
<i>nikA-E (hydC)</i>	Nickel transport	+	ND
<i>nirBDC</i>	NADH-dependent nitrite reductase	+	ND

blue = *fnr* target in  
Shewanella (by  
homology)

yellow = *arcA* target  
in Shewanella (by  
homology)

# ArcA Regulon

*14 of 15 operons  
are regulated  
similarly to E.coli*

conserved operon	regulonDB	z-change
aceBAK	-	0.2
acnB	-	-1.2
aldA	-	-2.9
cydAB	+	1.4
cyoABCDE	-	-3.5
fadBA	-	-0.5
focA-pflB	ND	3.5
gltA	-	-1.1
hyaABCDEF	+	4.0
mdh	-	-1.7
moeAB	+	1.4
ndh	-	0.3
nuoABCEFGHIJKLMN	-	-1.1
lpdA	-	-1.3
sdhCDAB-b0725-sucABCD	-	-0.9

# *Shewanella* anaerobic conditions activate known *E.coli* regulons

## Fnr conserved regulon

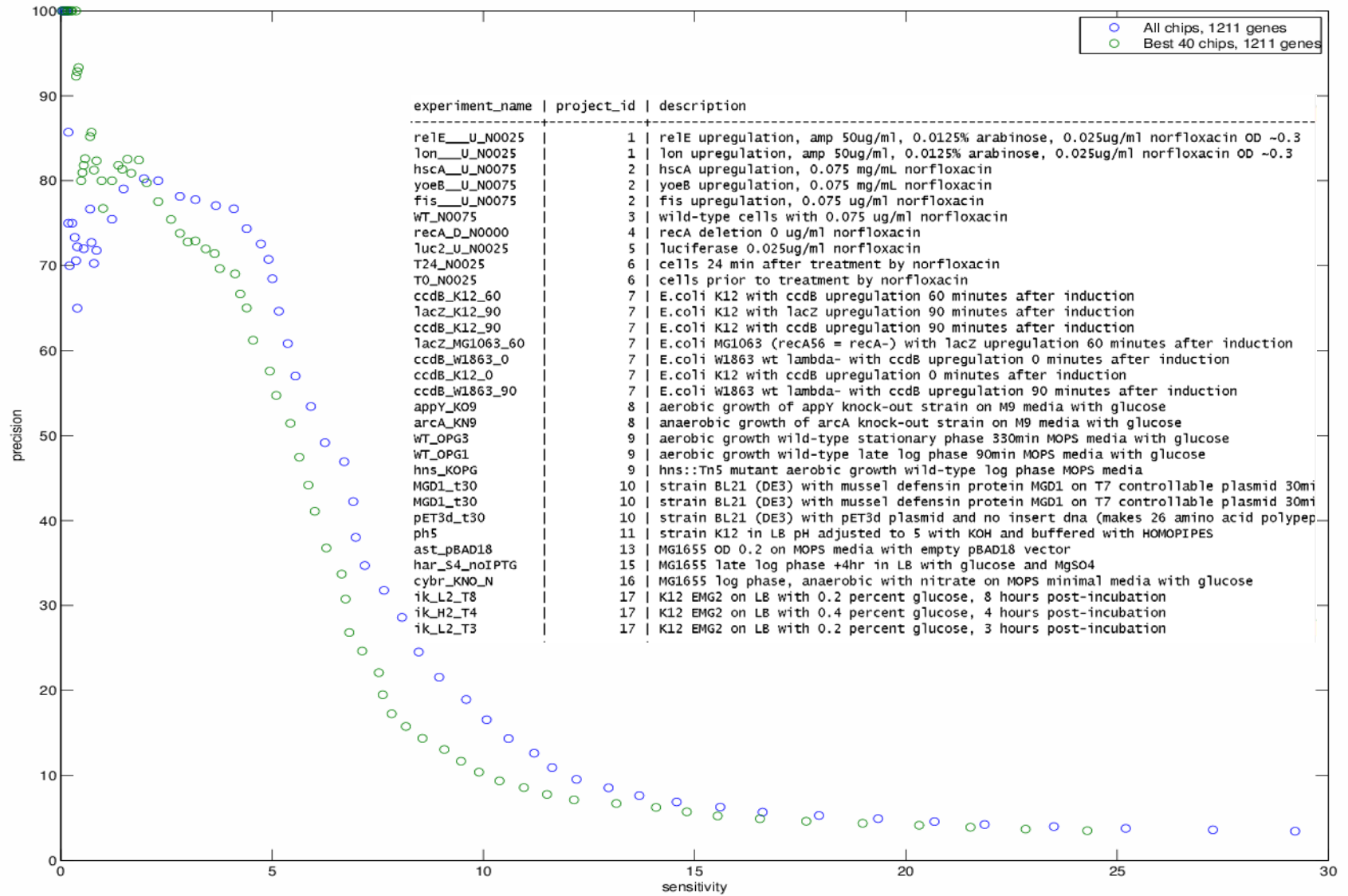
target operon	RegulonDB regulation	<i>Shewanella</i> array (Fu/O2)
nrdDG	+	2.7
adhE	+	3.4
arcA	+	1.4
cydDC	+	1.7
dcuB-fumB	+	2.1
dmsABC	+	2.9
fdnGHI	+	1.7
focA-pflB	+	3.5
hypBCDE	+	3.2
moaABCDE	+	0.5
moeAB	+	1.4
napFDAGHBC	+	-0.7
nrfABCDEFGFG	+	-1.0
cydAB	-	1.4
cyoABCDE	-	-3.5
ndh	-	0.3
nuoA-N	-	-1.1
sdhCDAB--sucA-D	-	-0.8

## ArcA conserved regulon

target operon	RegulonDB regulation	<i>Shewanella</i> array (Fu/O2)
aceBAK	-	0.2
acnB	-	-1.2
aldA	-	-2.9
cydAB	+	1.4
cyoABCDE	-	-3.5
fadBA	-	-0.5
focA-pflB	nd	3.5
gltA	-	-1.1
hyaABCDEF	+	4.0
mdh	-	-1.7
moeAB	+	1.4
ndh	-	0.3
nuoA--N	-	-1.1
lpdA	-	-1.3
sdhCDAB--sucABCD	-	-0.8

14/18 and 13/15 targets, respectively, are regulated as in *E.coli*

# Our *E.coli* data set can inform our *Shewanella* experiments



## **Acknowledgements**

**Tim Gardner and Gardner Lab at BU**

**Jim Collins and the Collins Lab at BU**

**Shewanella Federation**

**Department of Energy GTL Program**

**Krell Institute**

