Evaluation of Gene Expression for Sediment Toxicity Identification Evaluation

Progress Report
April 7, 2010

Steven M. Bay

Southern California Coastal Water Research Project

Chris Vulpe

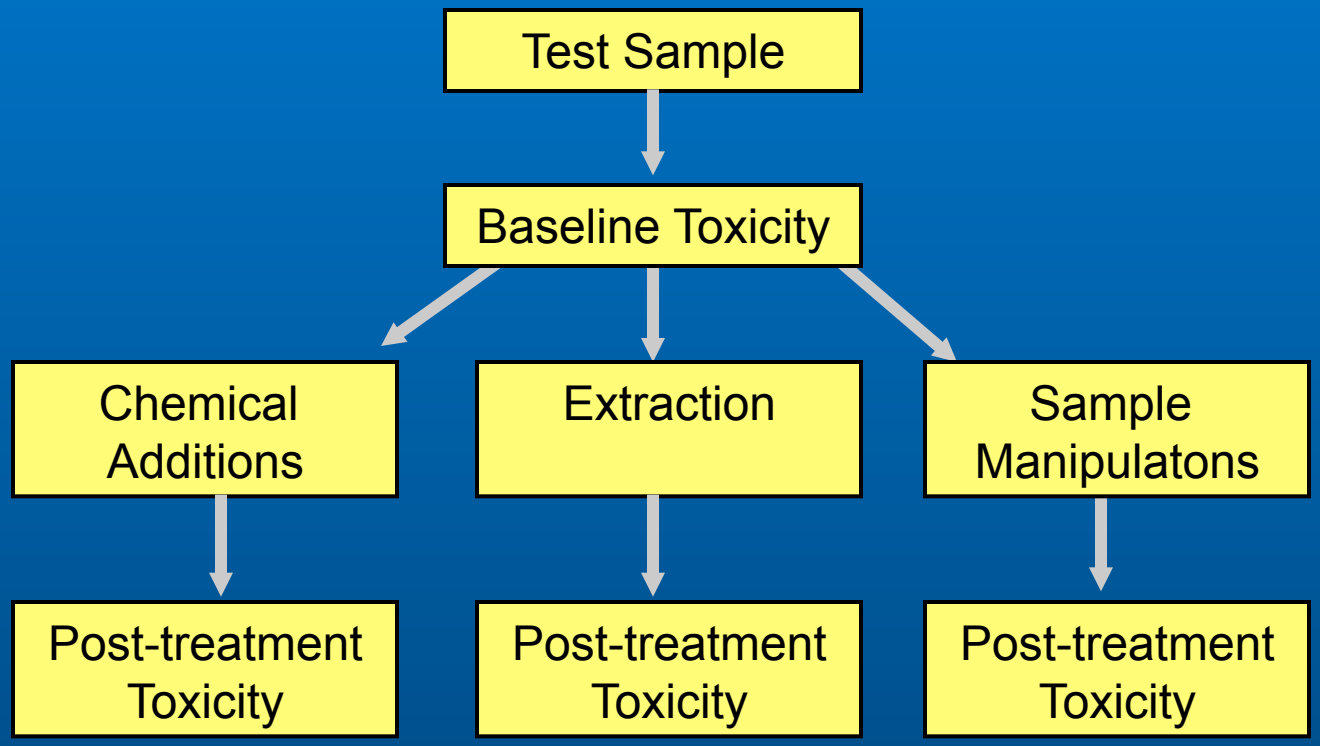
University of California Berkeley

Project Goal and Objectives

Develop improved tools for determining the contaminants responsible for impaired sediment quality

- **Develop a molecular stressor identification method based on gene expression in amphipods**
- **Evaluate the ability of the molecular method to identify toxicants in sediments**

Toxicant Identification Evaluation (TIE)
Traditional Approach



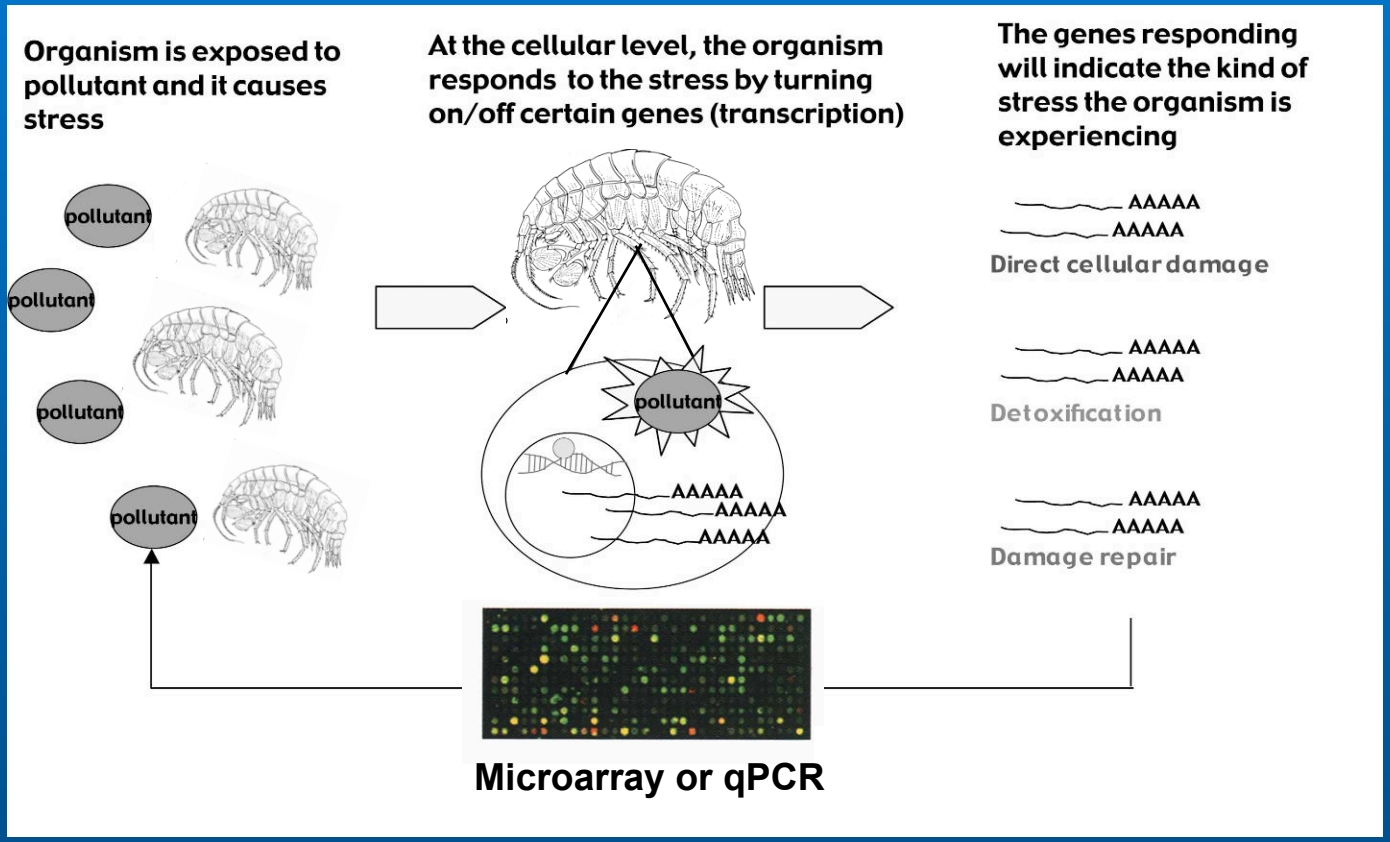
Various contaminant-specific treatments applied to sample

Changes in toxicity following sample treatments indicates type of toxicant

Better Stressor Identification Methods Are Needed

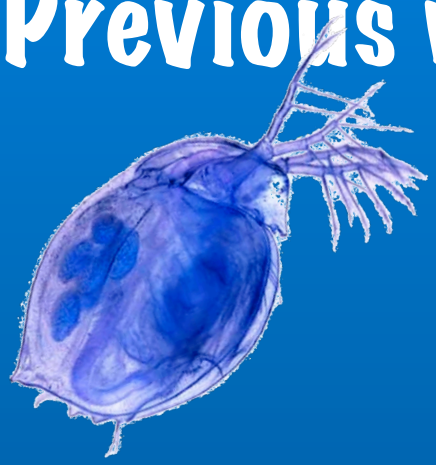
- **TIE results are frequently inconclusive or nonspecific**
  - **Chemical treatments have limited specificity**
- **Limited range of application**
  - **Need highly toxic sediments**
- **TIEs not applicable to resident organisms**
  - **Rely on laboratory manipulations of sediment**
- **Limited ability to identify new types of chemical stressors**
  - **Have to determine chemical characteristics first**

Molecular TIE Approach

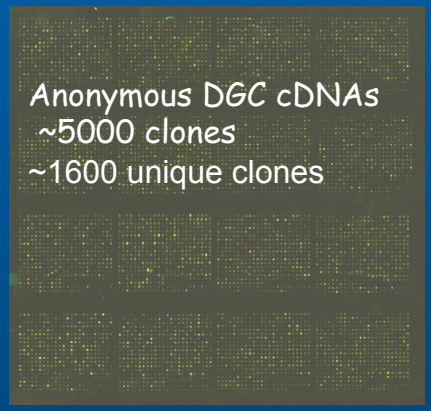


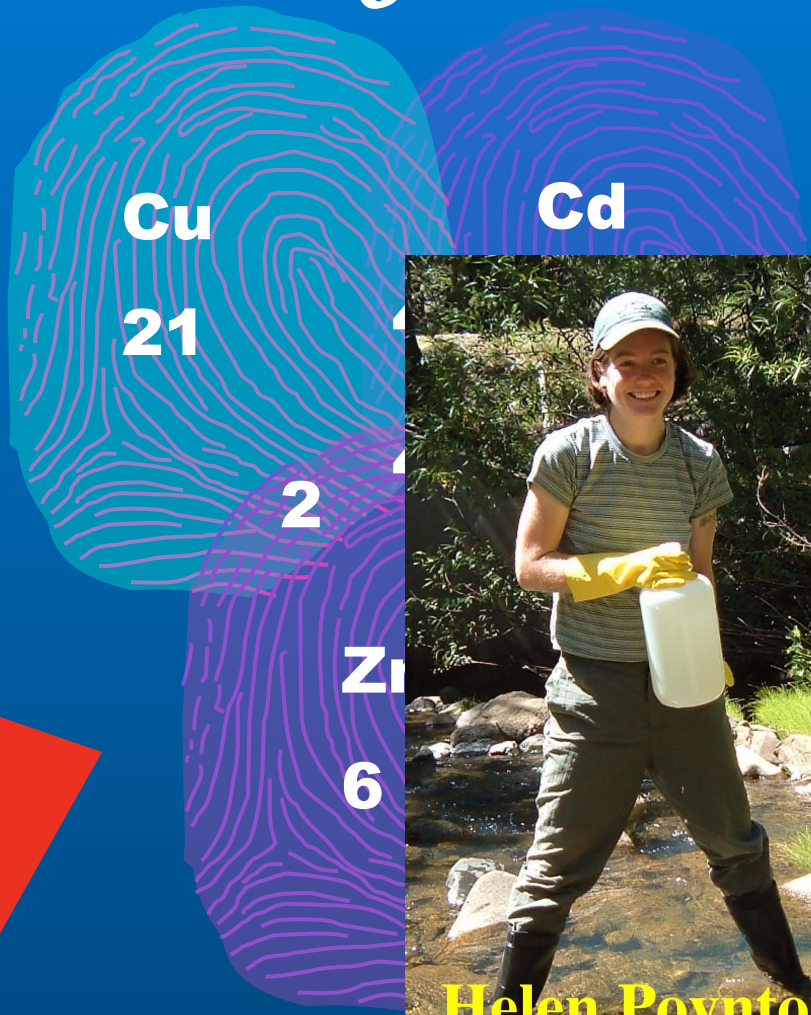
- Simultaneous evaluation of multiple genes provides a contaminant-specific "fingerprint" of toxicant exposure and effect = greater specificity
- Direct assessment of organism response with few or no chemical manipulations of sample = higher relevance
- Sublethal endpoints increases sensitivity of method = wider applicability

Previous work with *D.magna*



Exposed adult *Daphnia magna*
24 hrs at 1/10 LC50 acute.



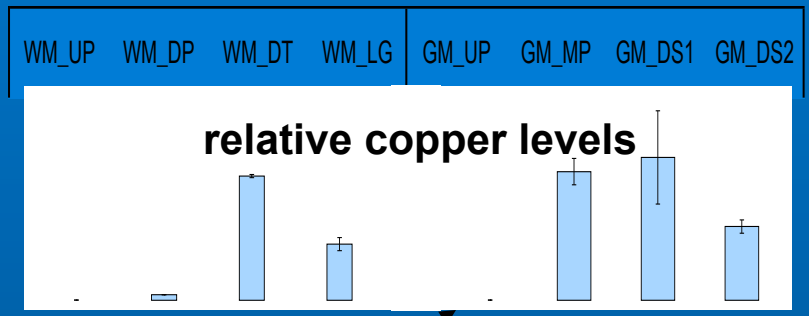


Each metal had a specific expression profile

Poynton et al. *Environmental Science and Technology,* Feb. 1, 2007

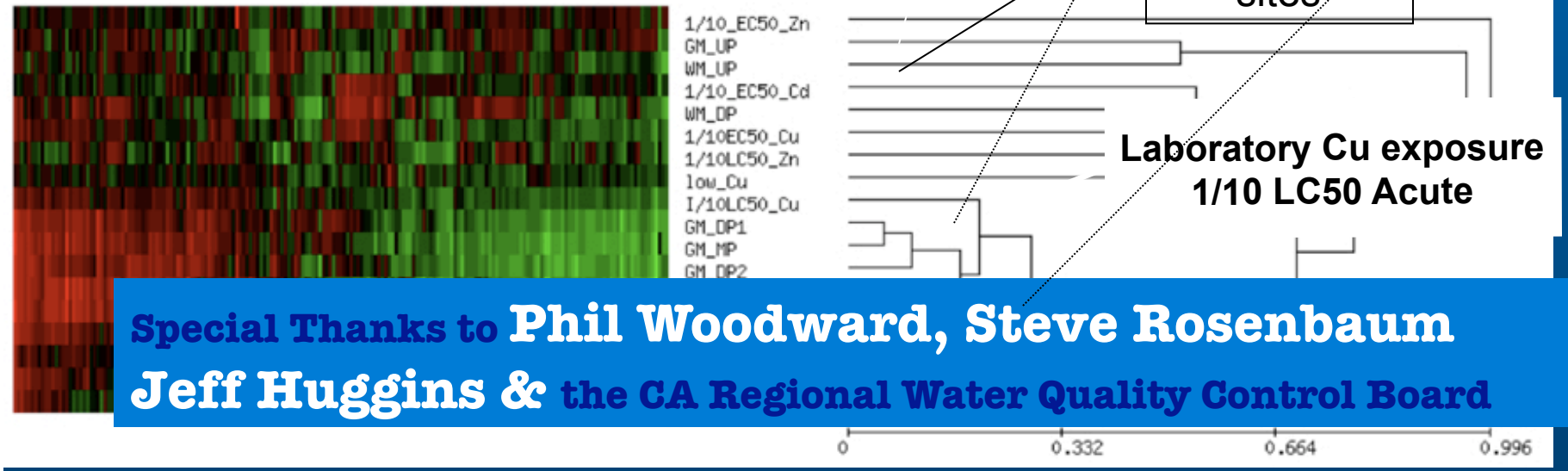
New predictive (of field exposure) biomarkers?

Field samples collected from two abandoned copper mines



Upstream sites

Downstream sites



Laboratory Cu exposure
1/10 LC50 Acute

Special Thanks to Phil Woodward, Steve Rosenbaum Jeff Huggins & the CA Regional Water Quality Control Board

Environ. Sci. Technol. 42: 6257-63, 2008

Does specificity hold with more chemicals?

Cu

Zn

2,4-DNT

WO_4

TNT

TNB

DX

Natalia Vinas-Reyero

Ed Perkins

US Army Engineering Research and Development Center

US Army ERDC

5000

cDN

norm

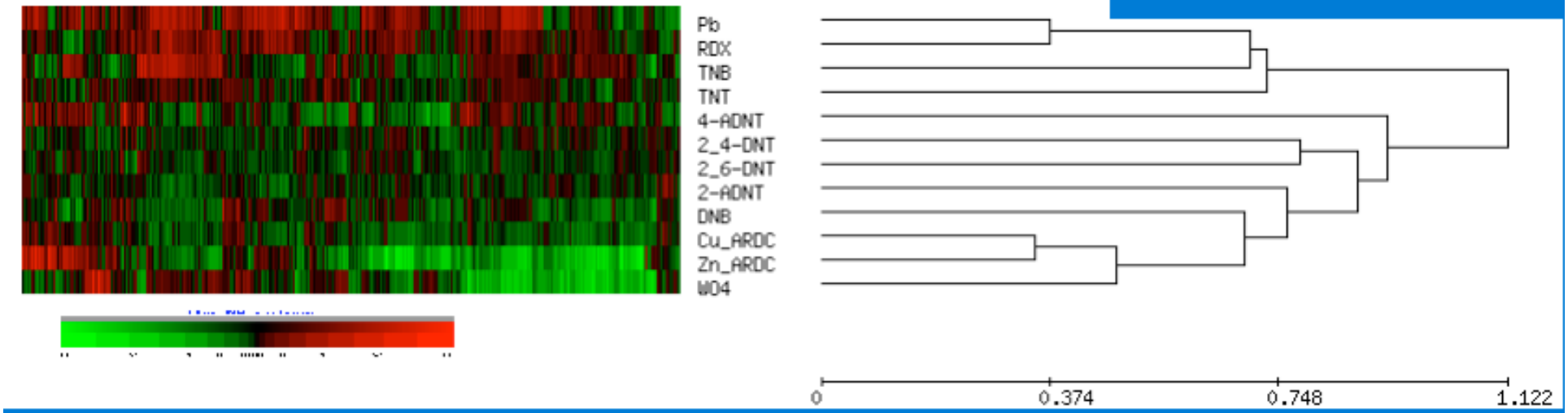
ins

ry

2000 Sequences Subtracted library, exposed to 12 chemicals

Environ Sci Technol 43: 4188-93, 2009

Specific expression profiles for each chemical

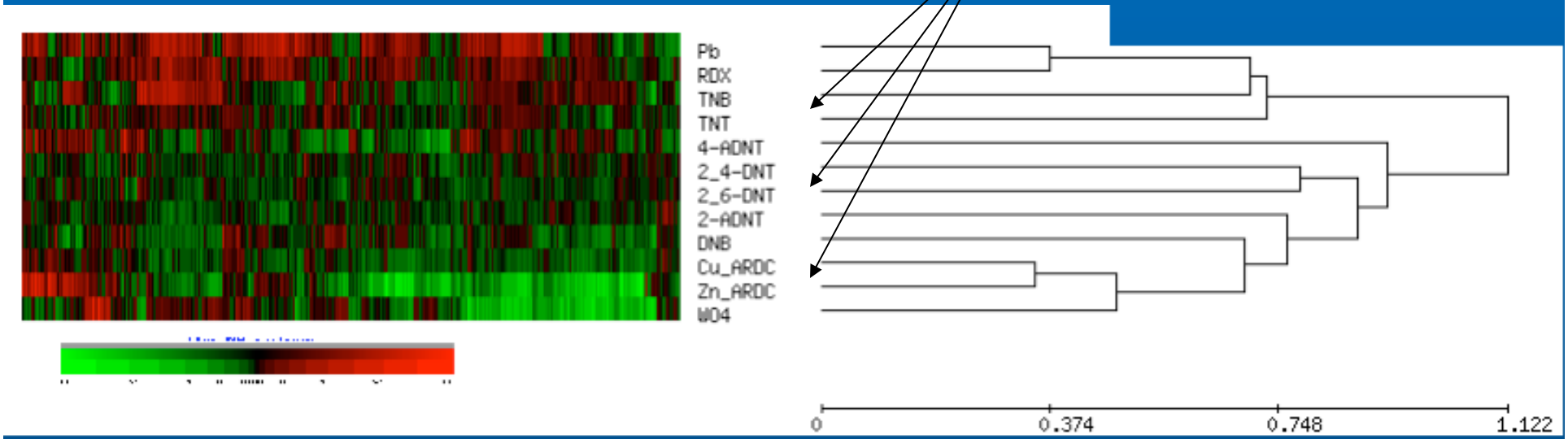


RDX specific biomarker suite

CANDIDATE BIOMARKERS	RDX	TNT	2-ADNT	4-ADNT	2,4-DNT	2,6-DNT	DNB	TNB	Cu	Zn	Pb	WO4
RDX												
no homology	-0.92	-0.53	-0.43	0.337	0.063	-0.02	-0.31	0.699	0.118	1.403	0.684	0.252
similar to D. pulex 94071, secreted protein	-0.74	-0.47	-0.24	0.406	-0.03	0.106	-0.1	0.618	0.065	1.193	0.624	0.221
AC004767: cuticle protein	-0.72	0.01	0.184	-0.59	0.111	0.063	-0.06	-0.01	-0.09	-0.02	-0.57	0.116
no homology	1.102	0.072	-0.24	-0.17	-0.3	-0.02	0.246	-0.13	-0.09	0.4	1.34	-0.08
XM_393544: myosin light chain	1.133	-0.18	-0.4	-0.51	-0.07	-0.11	0.02	-0.01	-0.25	0.078	1.578	0.195
similar to D. pulex 239793	1.046	0.287	-0.29	-0.36	-0.06	-0.14	-0.16	1.238	-0.59	-0.11	1.365	0.288
AY255624: actin	0.992	0.571	-	-	-0.22	-0.29	-0.08	-0.25	-0.36	-0.12	0.467	-0.1
no homology	0.966	-	0.042	-	-0.19	-0.15	0.267	-0.09	-0.17	0.212	1.137	-0.35
AY572863: actin	0.938	0.172	-0.5	-0.29	-0.23	-0.16	-0.11	0.086	-0.32	0.017	1.161	-0

Molecular Toxicity Identification Evaluation

Unknown chemical where does it fit?



- Improved toxicity identification
  - Sublethal effects, complex mixtures

Suite of QPCR biomarkers

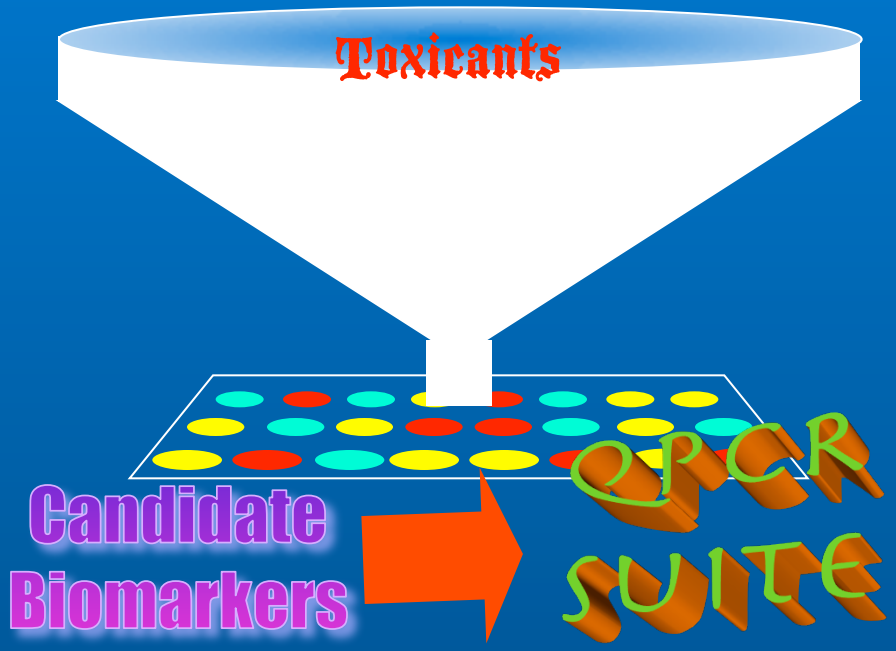
Expose to Individual Toxicants

Identify differentially expressed genes/toxicant

Select candidate QPCR biomarkers/toxicant

Independent exposures Validate QPCR biomarkers

Suite of Toxicant specific QPCR Biomarkers



Further Validation
Dose response
Mixtures
Abiotic/Biotic Factors
Field Studies

Environ Sci Technol 43: 4188-93, 2009

Sweet Suite

	4CA12	13CF5	BGBP	9AD9	PERITROPHIN	FERRITIN	12BG1
TNT			-2.56			2.2	
DNB		1.21	-1.84				
2,4 - DNT		2.25	2.01	1	2.03	0.81	
TNB	1.9		-1				-0.66
2,6 - DNT	2				-1	-2	1.26
NaWO4			1.64				-1.71
CuSO4			-1.34			0.71	-0.81
ZnCl2			1.52	1.58	1.22		
RDX	1.1	0.93					

UP

DOWN

NO CHANGE

7 QPCR assays in *D. magna* for genes responsive to 9 ORCs.

24 hr exposure, 1/10 LC50, 1 chemical/exposure

Individual QPCR biomarker are not toxicant specific

Pattern of "suite" of QPCR biomarkers is toxicant specific

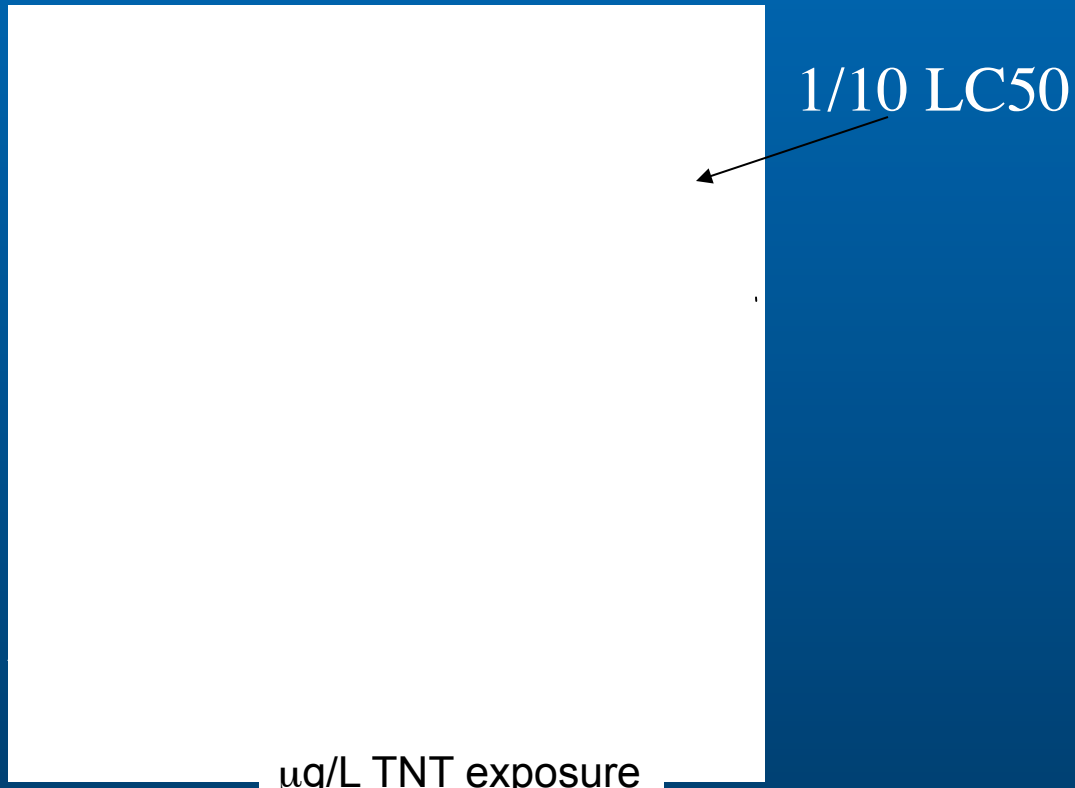
Environ Sci Technol 43: 4188-93, 2009

QPCR Suite - Dose Response

Compound	Single exposure (μg/L) ppb				
TNT	1,194	835.8	299	119.4	13.134
24DNT	2,000	1400	500	200	22
26DNT	1,600	1120	400	160	17.6
DNB	14,600	10220	3,650	1460	160.6
TNB	454	317.8	114	45.4	4.994
RDX	29,500	20650	7,375	2950	324.5

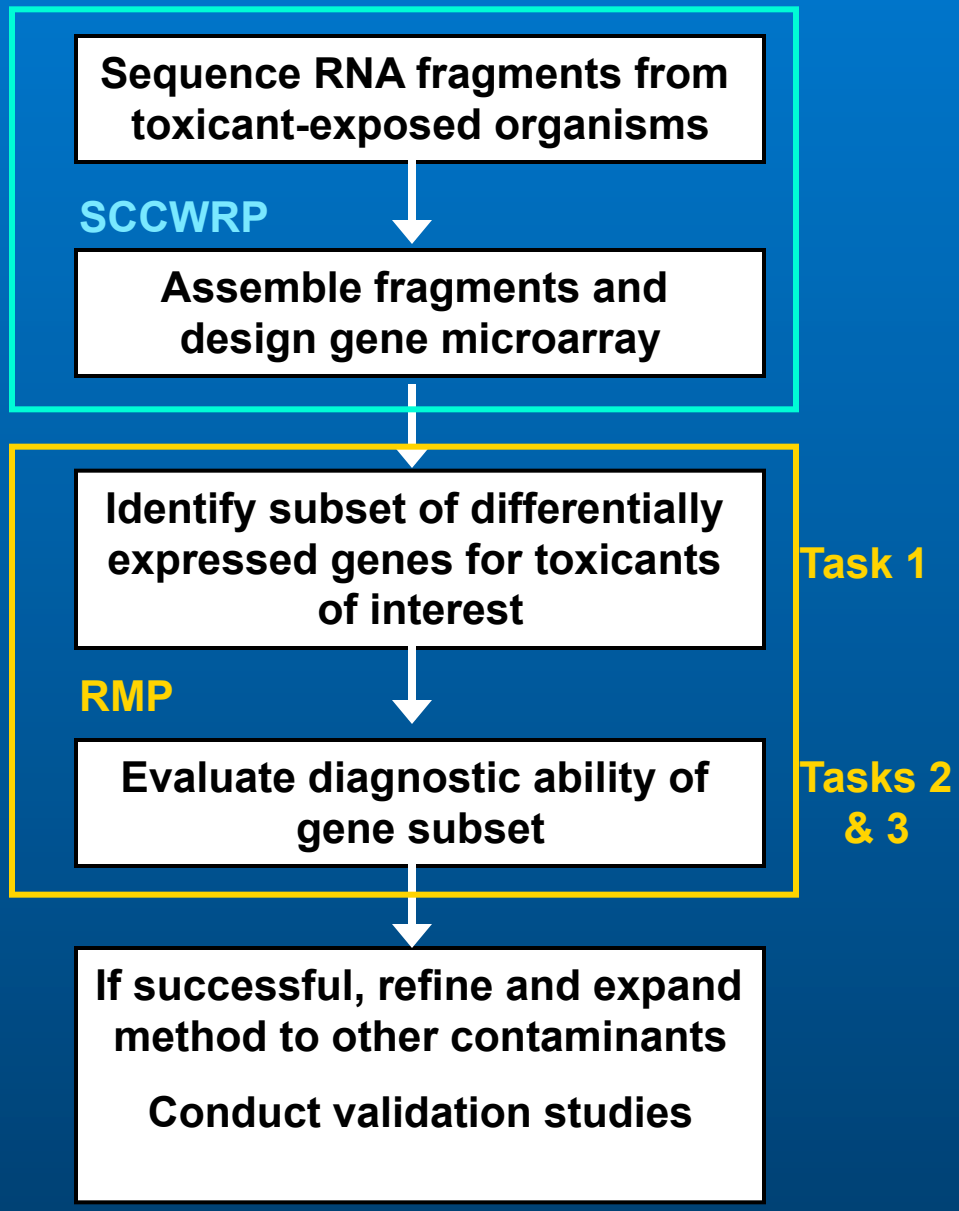
Quantitative relationship of expression to TNT (ppb)

1/10,000 LC50



Molecular TIE Development Plan

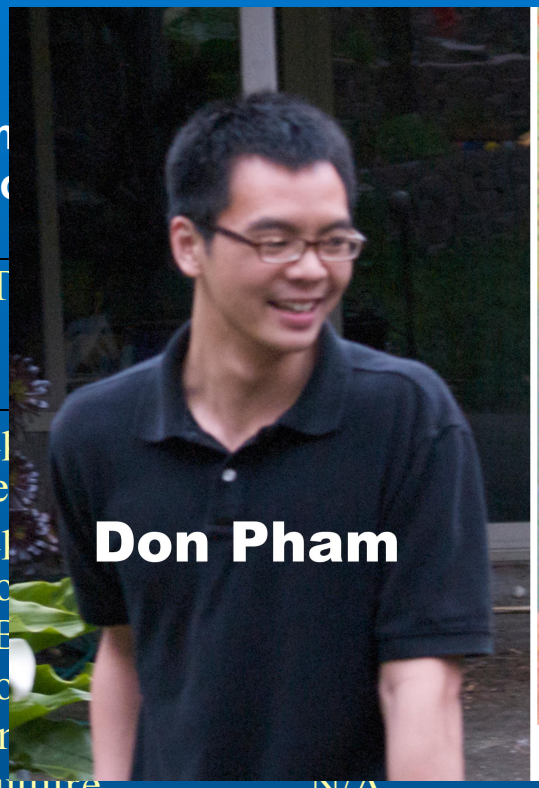
- Use amphipod *Eohaustorius estuarius*
  - Benchmark test species for most monitoring programs
- Build upon current SCCWRP research to develop preliminary microarray
  - Task 1: Identify diagnostic genes for selected contaminants
  - Task 2: Analyze lab and field samples to see if patterns can be detected
  - Task 3: Evaluate performance and potential of method



Project Update

- Microarray development finished
- RMP project in initial stages
  - Acquiring test samples
- Additional collaboration in development
  - Should accelerate research progress

Microarray Development



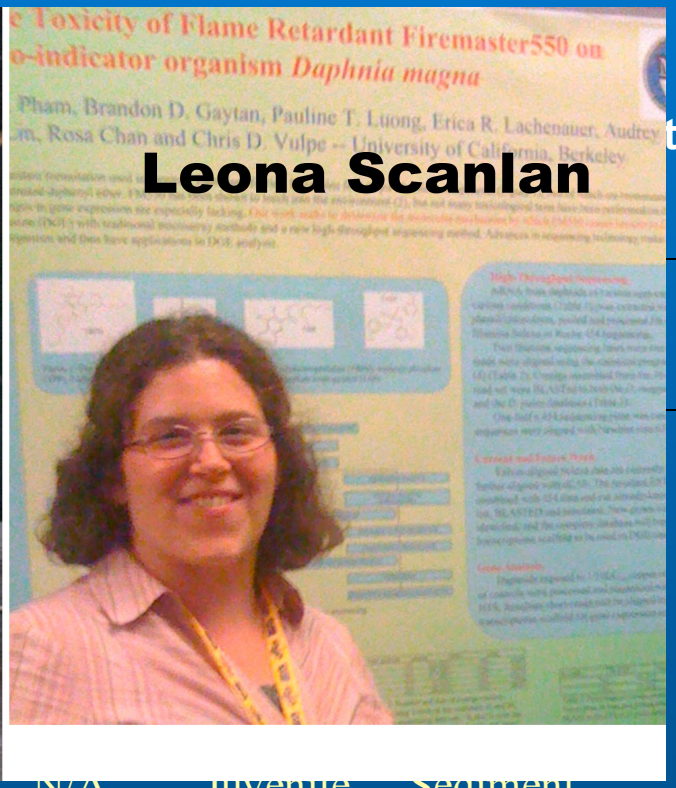


Table 1. Treatm **tive of different stresso**

Sample ID	T					Exposure Duration (days)
EE107-118	Fi					4
EE109-119	Pe					4
EE107-140	Fi					4
EE107-103	Co					4
EE108-23	PE					10
EE110-28	Co					10
EE110-10	Ar					10
Juv1-2	Culture	N/A	N/A	Juvenile	Sediment	N/A[1]
Juv2-2	Culture	N/A	N/A	Juvenile	Sediment	N/A
Int2	Culture	N/A	N/A	Adult	Sediment	N/A
Large2	Culture	N/A	N/A	Large Adult	Sediment	N/A

[1] Obtained from laboratory culture.

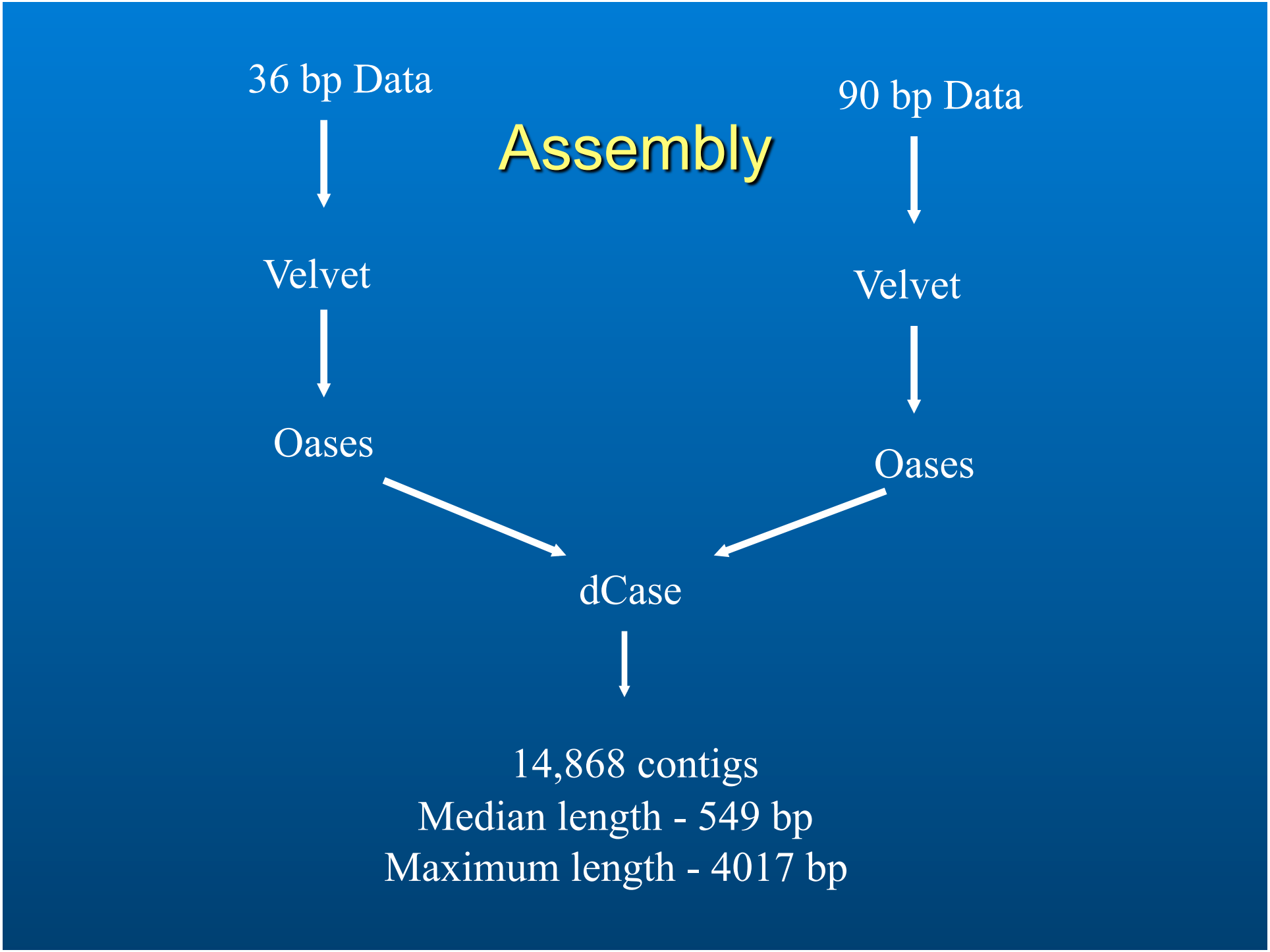
High throughput Transcriptome Sequencing

Illumina/Solexa Short read sequencing

# of base pair per read	# of sequence reads	Total Bases sequenced
36	14,699,407	529,178,562
90	16,727,943	1,595,504,870
		2,124,683,432

If assume 35K transcripts of 1K in size then 1X transciptome = 35,000,000

Then ~60X coverage - but assumes equal expression


36 bp Data
90 bp Data
Assembly
Velvet
Velvet
Oases
Oases
dCase
14,868 contigs
Median length - 549 bp
Maximum length - 4017 bp

Annotation

14,868 contigs → Blast2Go NCBI NR protein database → 3,698 annotated contigs

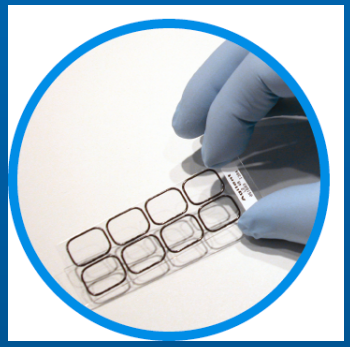
Sequence desc.	Seque	Hit desc.	E-Value
apoptotic chromatin condensation inducer-like protein	4017	gi|193674119|ref|XP_001949796.1|PREDICTED: similar to hook-like CG10473-PA [Acyrth	9.57E-34
tripartite motif protein	3343	gi|157134133|ref|XP_001663162.1|tripartite motif protein trim2,3 [Aedes aegypti] >gi|1	6.05E-127
collagen alpha-2	3210	gi|242016919|ref|XP_002428942.1|collagen alpha-2 precursor, putative [Pediculus huma	1.74E-123
scavenger receptor class a- lysyl oxidase (agap004118-pa)	3141	gi|149257729|ref|XP_001479723.1|PREDICTED: similar to hCG2040007 [Mus musculus]	1.22E-04
cg3808 cg3808-pa	3084	gi|91088207|ref|XP_973242.1|PREDICTED: similar to CG3808 CG3808-PA [Tribolium cast	3.96E-117
200 kda antigen	3038	gi|221505438|gb|EEE31083.1|conserved hypothetical protein [Toxoplasma gondii VEG]	1.21E-09
spliceosome associated protein	2969	gi|91076314|ref|XP_969681.1|PREDICTED: similar to spliceosome associated protein [Tri	6.54E-93
4snc-tudor domain	2816	gi|241835645|ref|XP_002415051.1|4SNc-Tudor domain protein, putative [Ixodes scapula	0
sallimus cg1915- isoform c	2806	gi|15425681|dbj|BAB64297.1|I-connectin [Procambarus clarkii]	1.62E-69
neuroglian cg1634- isoform a	2781	gi|189242457|ref|XP_970217.2|PREDICTED: similar to AGAP000720-PA [Tribolium castar	2.18E-58
26s proteasome regulatory subunit rpn1	2732	gi|91075936|ref|XP_967560.1|PREDICTED: similar to AGAP002481-PA [Tribolium castane	0
translation initiation factor subunit	2715	gi|110766548|ref|XP_623580.2|PREDICTED: similar to eIF3-S8 CG4954-PA [Apis mellifera	0
serine threonine protein kinase	2672	gi|158292024|ref|XP_313587.4|AGAP004315-PA [Anopheles gambiae str. PEST] >gi|157(	7.62E-77
ubiquitin-activating enzyme e1	2594	gi|270014908|gb|EFA11356.1|hypothetical protein TcasGA2_TC011512 [Tribolium castar	0
alpha-cop partial	2553	gi|115649185|ref|XP_001179078.1|PREDICTED: similar to alpha-cop protein, partial [Strc	1.03E-131
neutral alpha-glucosidase ab	2547	gi|260791718|ref|XP_002590875.1|hypothetical protein BRAFLDRAFT_115975 [Branchio	0
I-connectin [Procambarus clarkii]	2463	gi|15425681|dbj|BAB64297.1|I-connectin [Procambarus clarkii]	0
jumonji domain containing 1b	2449	gi|110758018|ref|XP_392473.3|PREDICTED: similar to jumonji domain containing 1B [Ap	0
vasa protein	2307	gi|201067640|gb|ACH92926.1|vasa protein [Parhyale hawaiensis]	0
alpha 2 macroglobulin	2294	gi|118076609|gb|ABK60046.1|alpha-2-macroglobulin [Macrobrachium rosenbergii]	0
rab gdp-dissociation inhibitor	2230	gi|91080775|ref|XP_968281.1|PREDICTED: similar to rab gdp-dissociation inhibitor [Tribo	0
vasa protein	2229	gi|201067640|gb|ACH92926.1|vasa protein [Parhyale hawaiensis]	0
clathrin heavy	2209	gi|193669177|ref|XP_001945333.1|PREDICTED: similar to AGAP003021-PA [Acyrthosiph	0
igf2 mrna binding protein	2174	gi|170036665|ref|XP_001846183.1|igf2 mRNA binding protein [Culex quinquefasciatus] >	7.31E-67

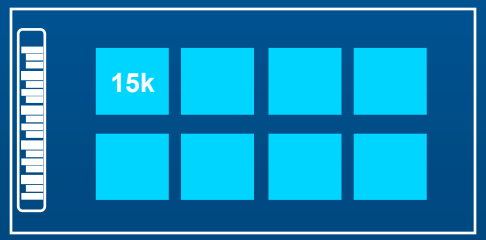
Example annotations

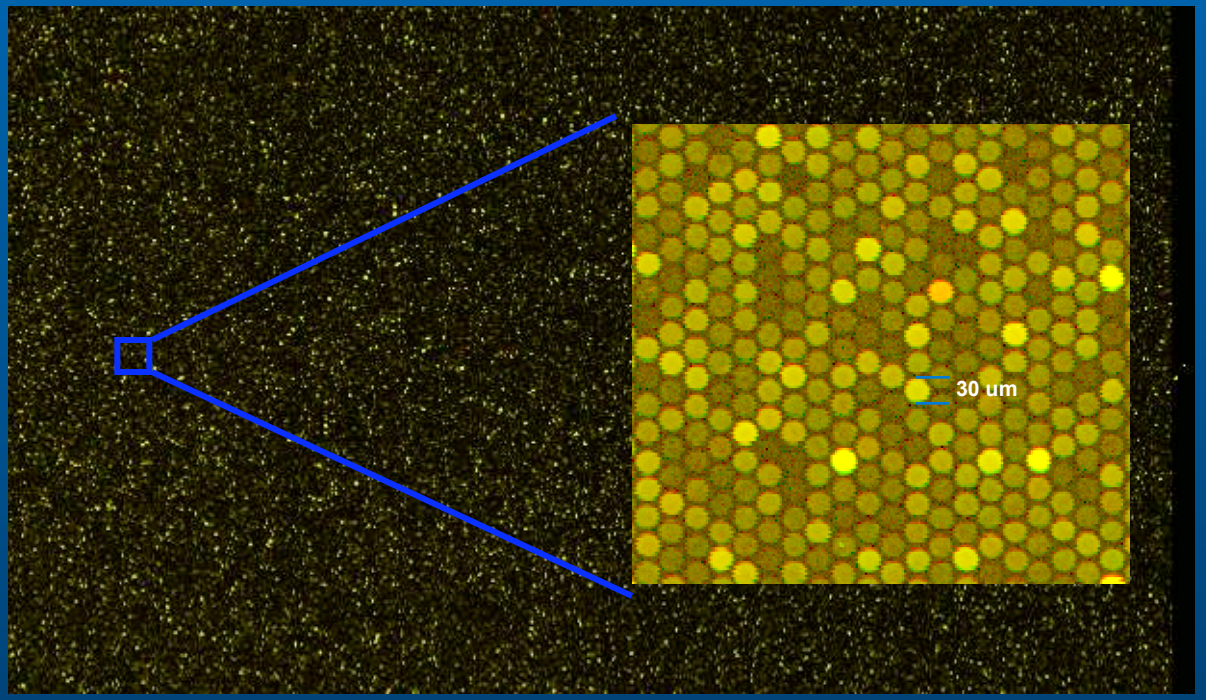
Microarray design

Agilent 8 X15K array

Contigs > 300 bp - 9681 contigs
Contigs < 300 bp with annotation - 289 contigs
eArray probe selection - 9970 probes, 536 Agilent controls,
4494 duplicates







Evaluation of Gene Expression for TIE

- Acquiring amphipod samples for use in Task 1 (calibration) and Task 2 (analysis)
- Water and sediment exposure to target contaminants (multiple concentrations)
  - Cyfluthrin (SCCWRP)
  - Chlordane (SCCWRP)
  - Pyrene (future, from MPSL)
- RMP 2009 field samples
- So Calif. field samples
  - Pyrethroid-impacted site w/ PAHs, metals, and CHCs

Work Plan

- **Task 1: Calibrate microarray for model toxicants**
  - **Identify diagnostic genes to differentiate exposure to:**
    - **Cyfluthrin, chlordane, pyrene**
    - **Sediment vs. water**
- **Task 2: Analyze samples representing exposure to target compounds and mixtures**
  - **Toxic and nontoxic dose levels (blind samples)**
  - **With other potential stressors (ammonia, salinity)**
  - **SF Bay and So Calif. field samples**

Work Plan

- **Task 3: Evaluate results to describe performance of gene expression approach**
  - **Variation among replicates**
  - **Accuracy in identifying model stressors**
  - **Sensitivity to confounding factors**
- **Final report will describe approach and provide recommendations for further development**
- **Project completion by December 2010**

New Collaboration

- Developing joint project with Hollings Marine Lab
  - Genomics core facility
- Task 1: Investigate effects of dose and duration
  - Needed to optimize test design
  - Identify subset of differentially expressed genes for each contaminant
- Task 2: Calibrate and test microarray for multiple contaminants
  - Planned overlap with RMP project (interlab comparison)
  - Information on additional contaminants
  - Alternative data analysis approaches

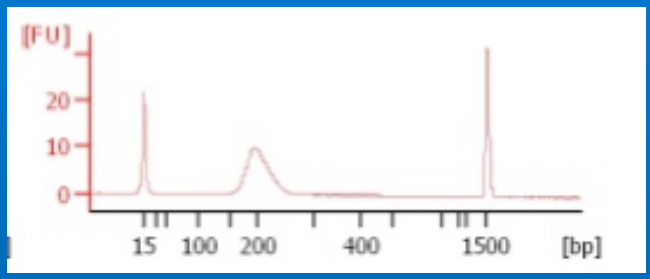
Summary

- On schedule
  - Too early to judge success
- Interest in program is growing
  - New collaboration with HML
  - Other agencies
- Additional development work required
  - This project will test concept
  - Further refinement and calibration needed to develop tool for use in programs

Solexa Genome Analyzer

- Sequences 36 - 90bp “reads”
- Similar to PCR
- Utilizes colored nucleotides

Sample Preparation



mRNA sample

Fragmented cDNA

Sequencing

Quality Control

cDNA with adapters

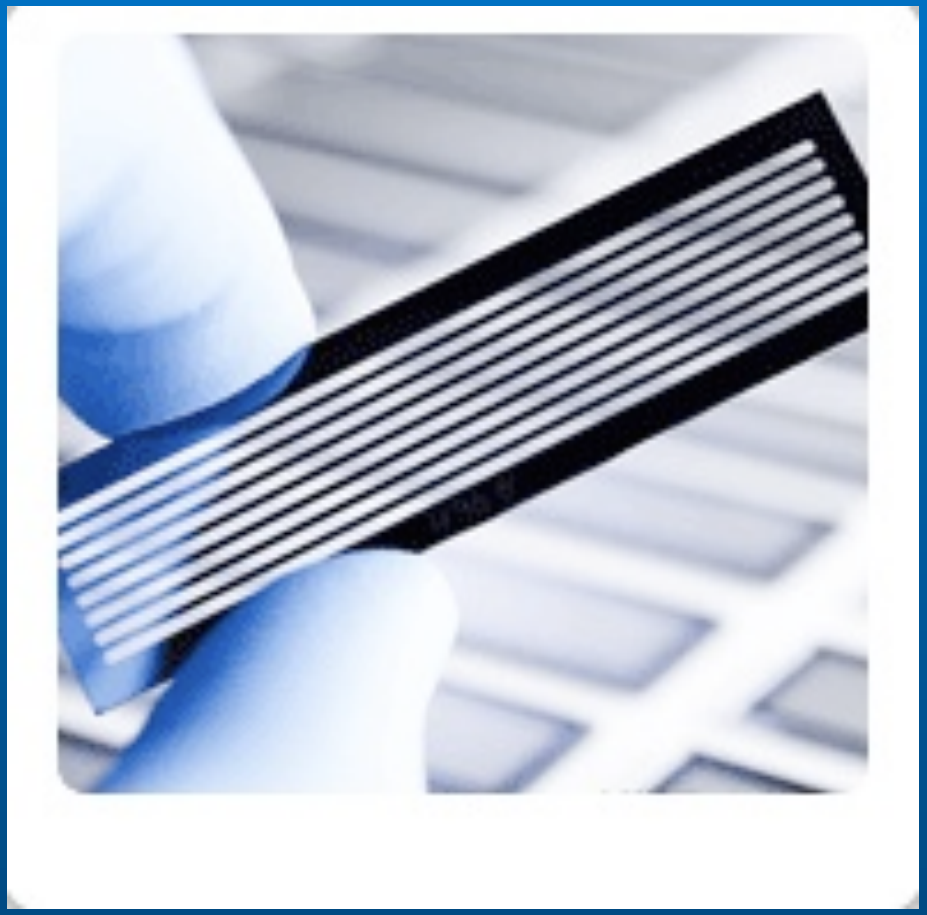
Amplification

Gel purification --200bp

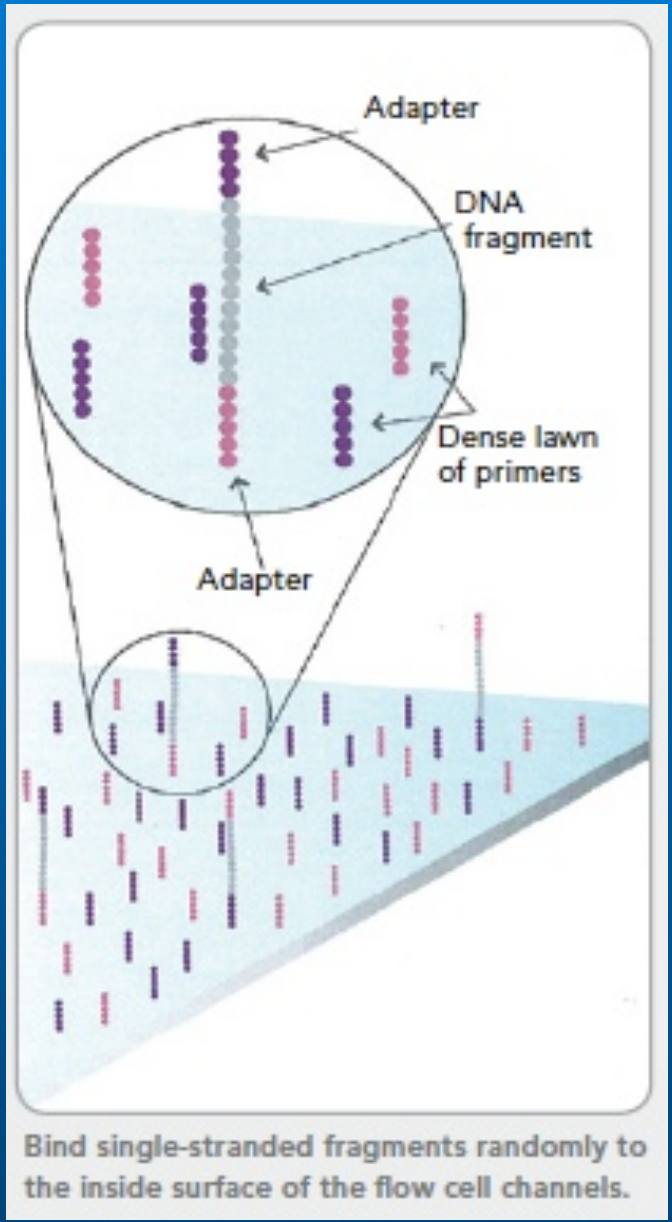
Sequencing

"Flow cell"

- 8 lanes or samples
- Solid phase amplification



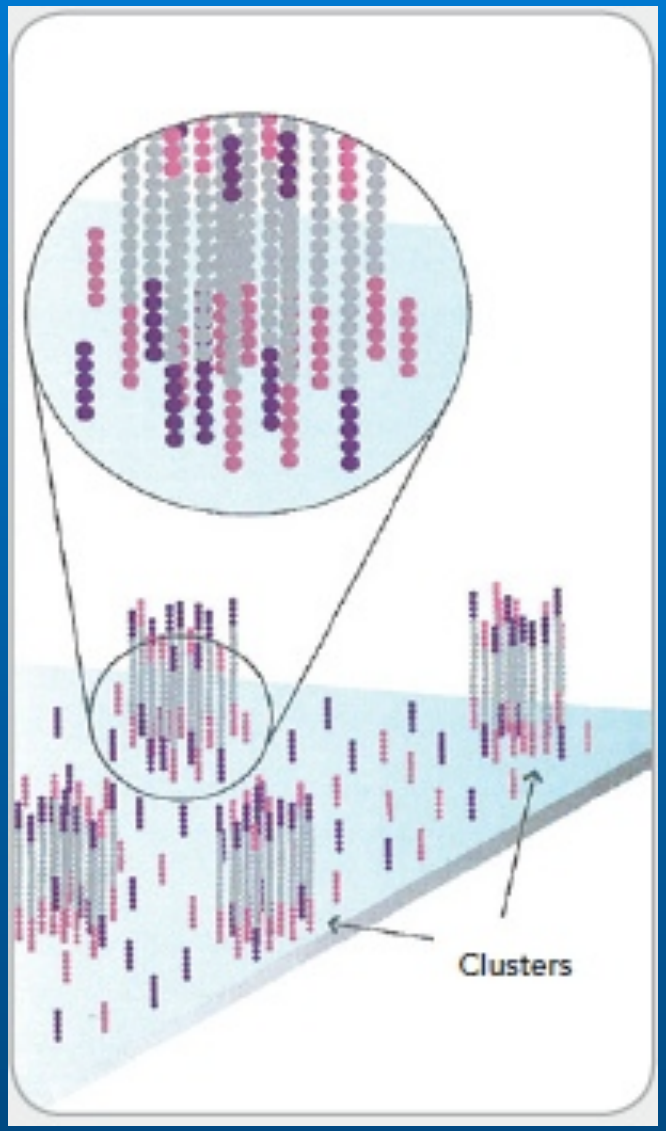
$5x10^6x36=180X10^6$ bases per flow lane



- Each molecule separated by 1uM
- "Lawn of primers"

Add enzymes, normal dNTPs

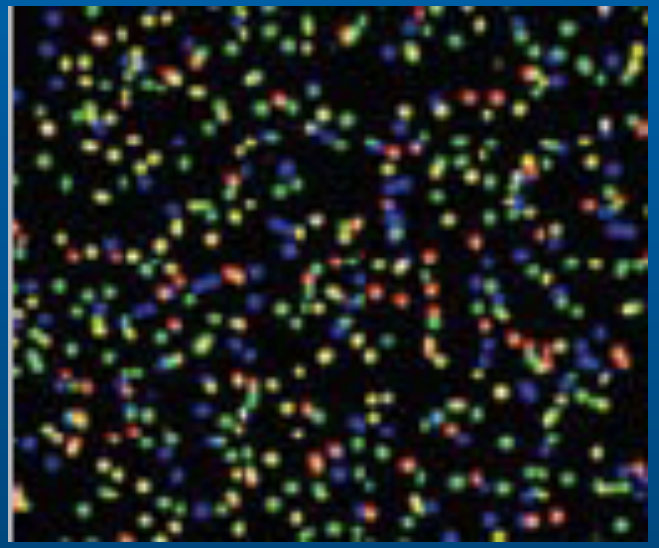
- Each molecule is amplified up to 1,000x



- Four proprietary fluorescently-labeled dNTPs
- Reversible termination modified
- One base added at a time

Colors!

- A photo is taken at end step
- Information is transmitted to a computer and transformed into base sequence
- Reaction is terminated and stop base is stripped off
- Reaction is repeated.



Alignment

- **Fragment short-reads, align according to overlap parameters**
- **Create “transcriptome”**