RASL-Seq: A Gene Expression Platform to Identify Toxicity Mechanisms and Adaptive Responses

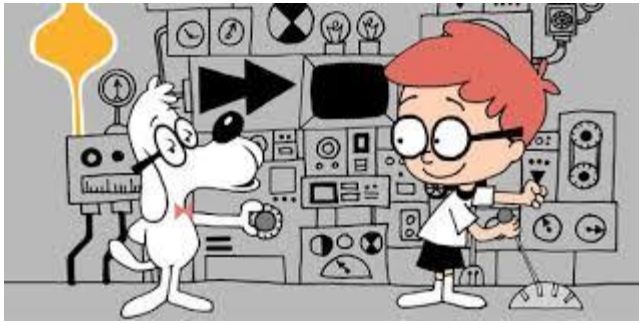
DAVID GERHOLD
NIH - NCATS
20 MARCH, 2014
EPA'S COMMUNITIES OF PRACTICE

NCATS

NIH National Center for Advancing Translational Sciences

National Heart Lung and Blood Institute

Summary



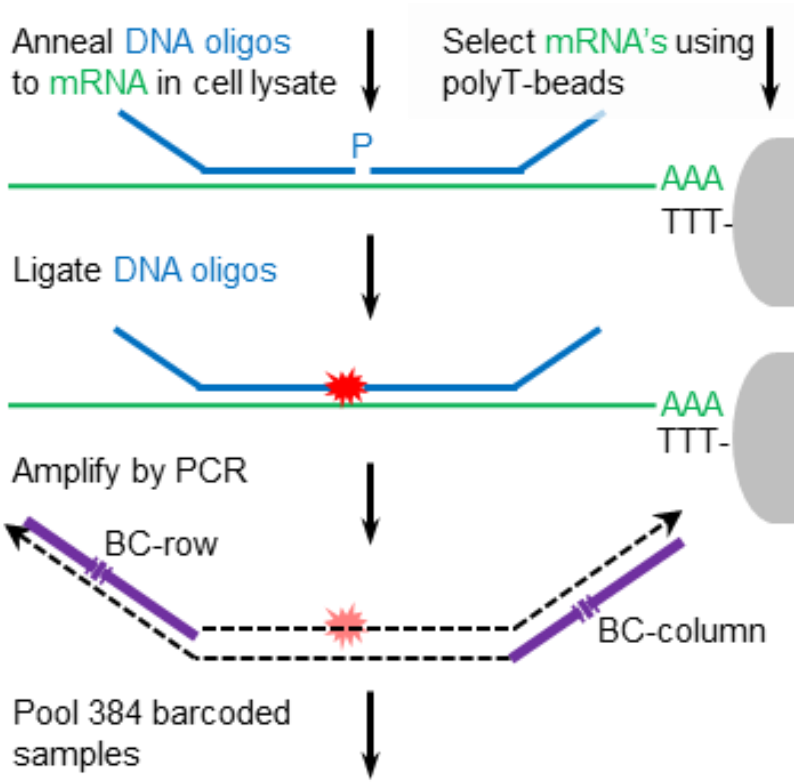
- RASL-Seq Selected, Results
- RASL-Seq platform works well
- Industrialization proceeding
- Selecting 1,400 genes to represent the genome(s)
- Goal: HT Gene Expression Core Facility
- How to apply it? Elucidate modes and mechanisms of toxicities or AEs. ‘Secondary screening’
- How will we use these results for risk assessment?

Goals

- A technology that will quantify mRNA responses in hundreds-thousands of genes
- Throughput > Thousands of samples to address: many compounds x multiple cell lines x multiple concentrations x multiple time points
- >> Gene Expression Core Facility at NCATS
- Low variance, intra- & inter-experiment, will enable data interpretation and construction of a rich reference database
- Data analysis pipeline

NIH National Center for Advancing Translational Sciences

RASL-Seq Schematic



RASL-Seq Characteristics

RASL-Seq has emerged from evaluation of six technologies based on:

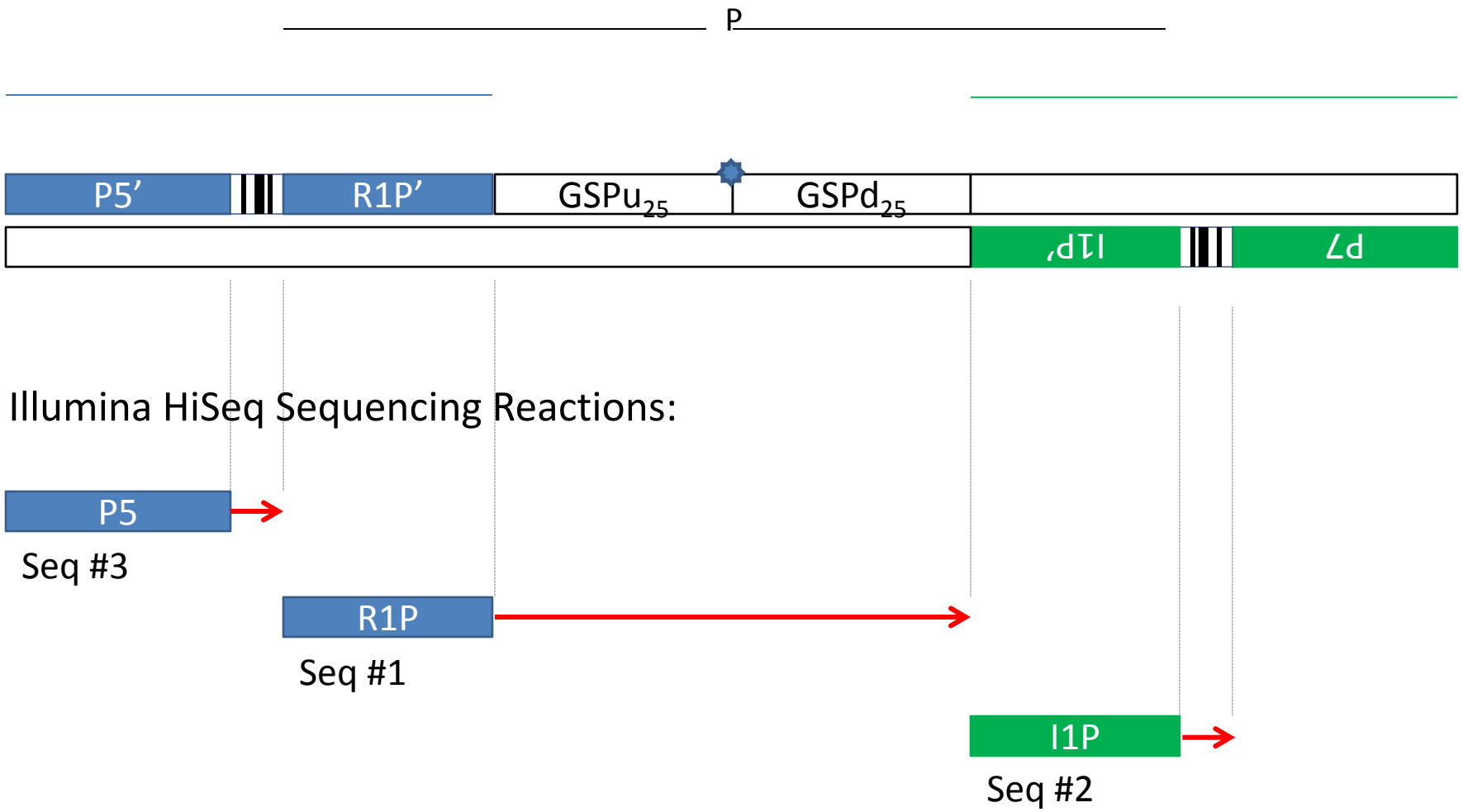
- **Multiplex:** ≥1,000 genes/sample, and 384 samples/sequencing reaction
- **Throughput:** (384 samples/run x ~10 runs/week)
- **Economy**: roughly $10.43/sample or $4,000/384 sample-run, including NextGen sequencing
- **Accuracy:** uses 3 redundant assays/gene
- **Reproducible:** avoids chaotic cDNA synthesis step. Intra-experiment $R^2 \geq 0.99$
- **Gene-Specific**: % mispairing events (after removing 5 promiscuous assays, of 360) is 0.26%

RASL-Seq invented by Xiang-Dong Fu Lab, Stanford U. Li H et al. PNAS 2012;109:4609-4614

Collaboration between BioSpyder (J. Yeakley & J. McComb) and NCATS to optimize RASL-Seq. Includes SBIR grant from NCATS.

NIH National Center for Advancing Translational Sciences

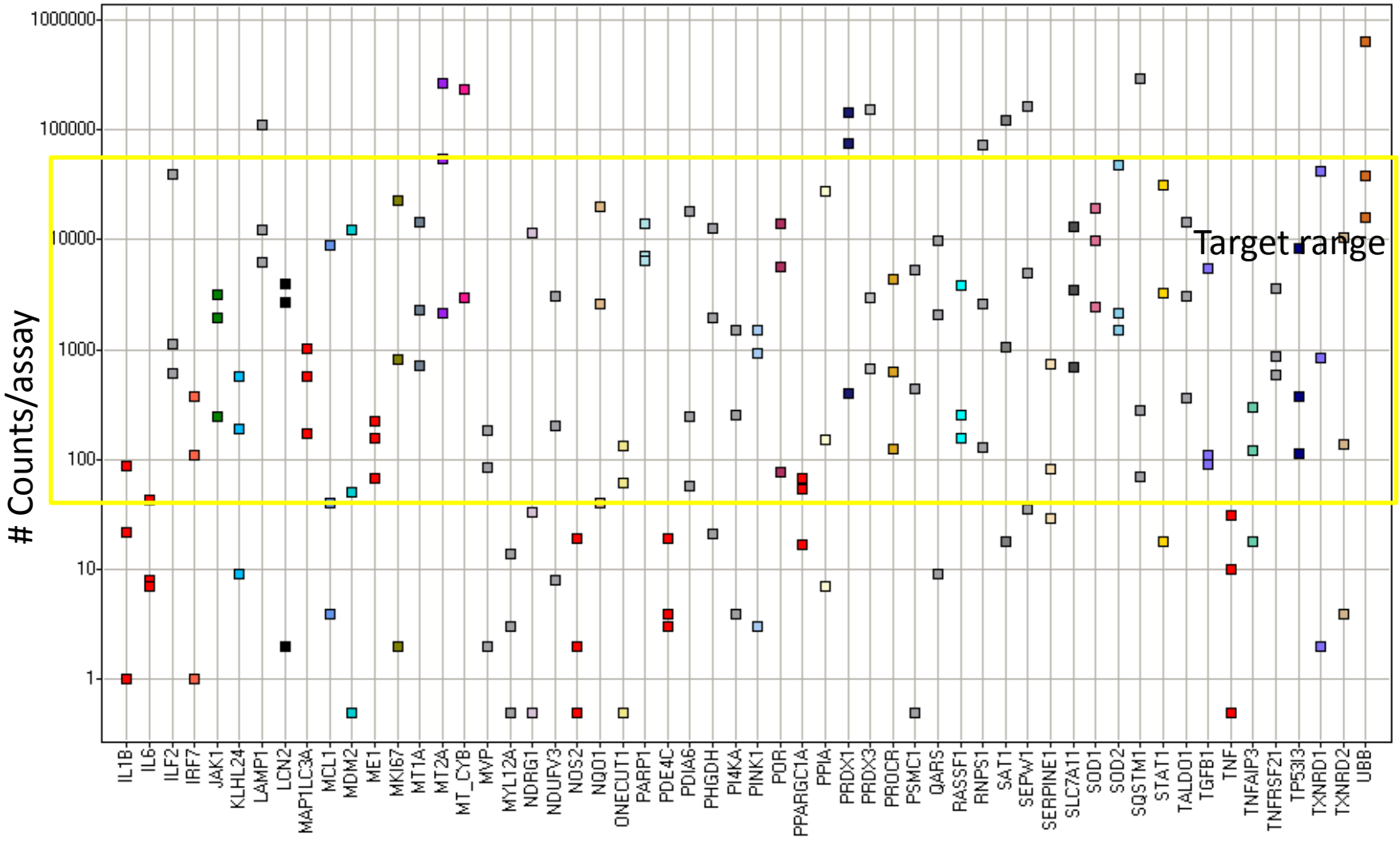
RASL Product Sequencing



3 RASL-Seq assays/gene vary widely in # counts

⇒ Remove 'greedy' assays

⇒ ignore assays that give low counts



RASL-Seq Progress March 2014

1st Gene set 120 genes (x 3 assays/gene) =360 assays. 13 'greedy' assays removed, and 5 non-specific (mispairing) assays removed.

- All steps performed by hand, using 384 samples, in microplates.
- Biomek FX robot installed 18 March 2014
- Good yields from PCR step in all five 'runs'
- 1-of-3 sequencing primers has been problematic. Two datasets now have resolved that issue for MiSeq instrument (24 or 50 samples ⇨ 11 million high-quality seq). Excellent data!
- Testing on HiSeq2000 instrument set for April 2014. Expect 150 million sequences/lane. 150 million sequences/384 samples = 390,000 sequences/sample.

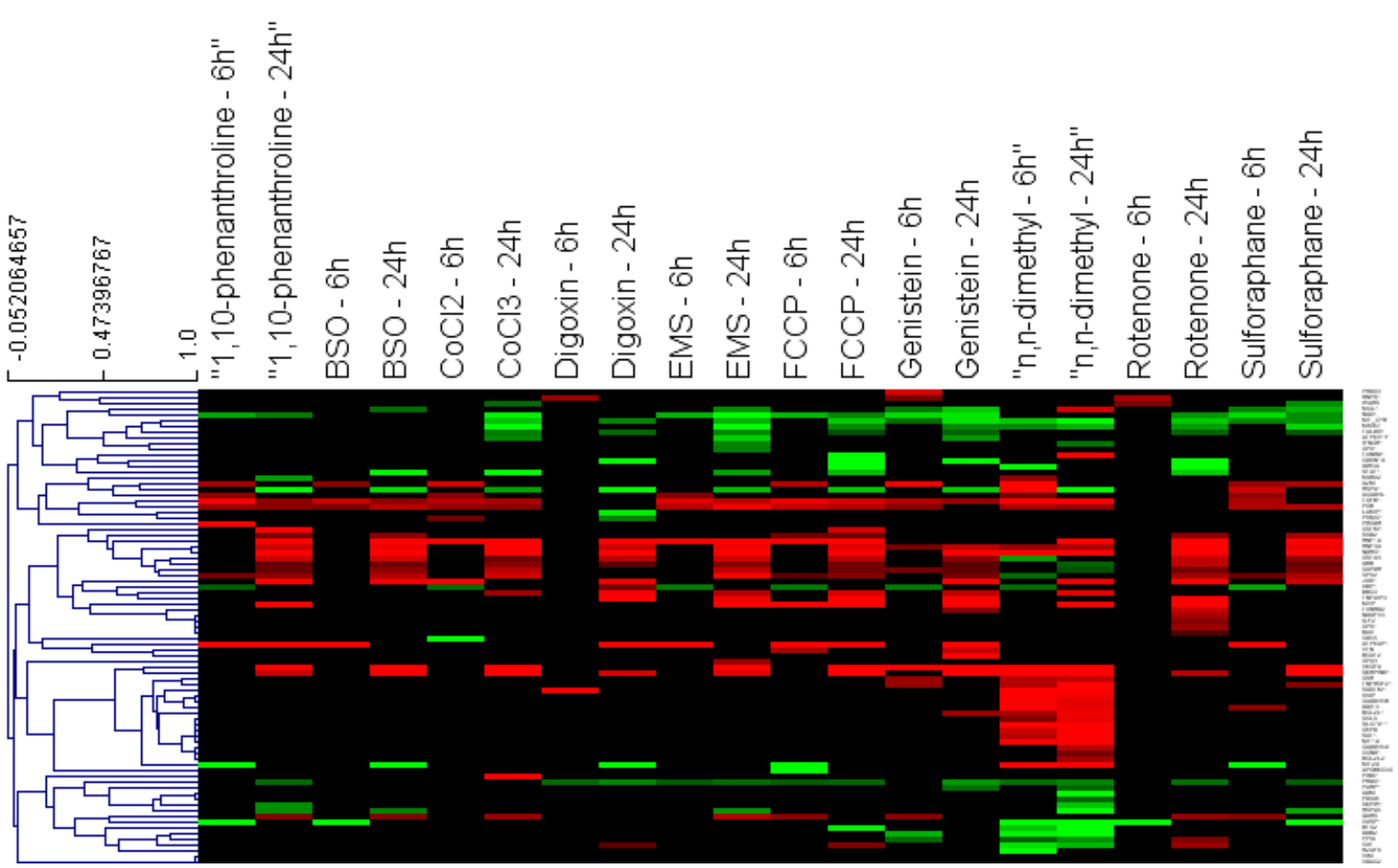
2nd Gene set: 320 genes (x 3 assays/gene) 960 assays being designed

3rd Gene set: ~1,000 genes being selected by working group.

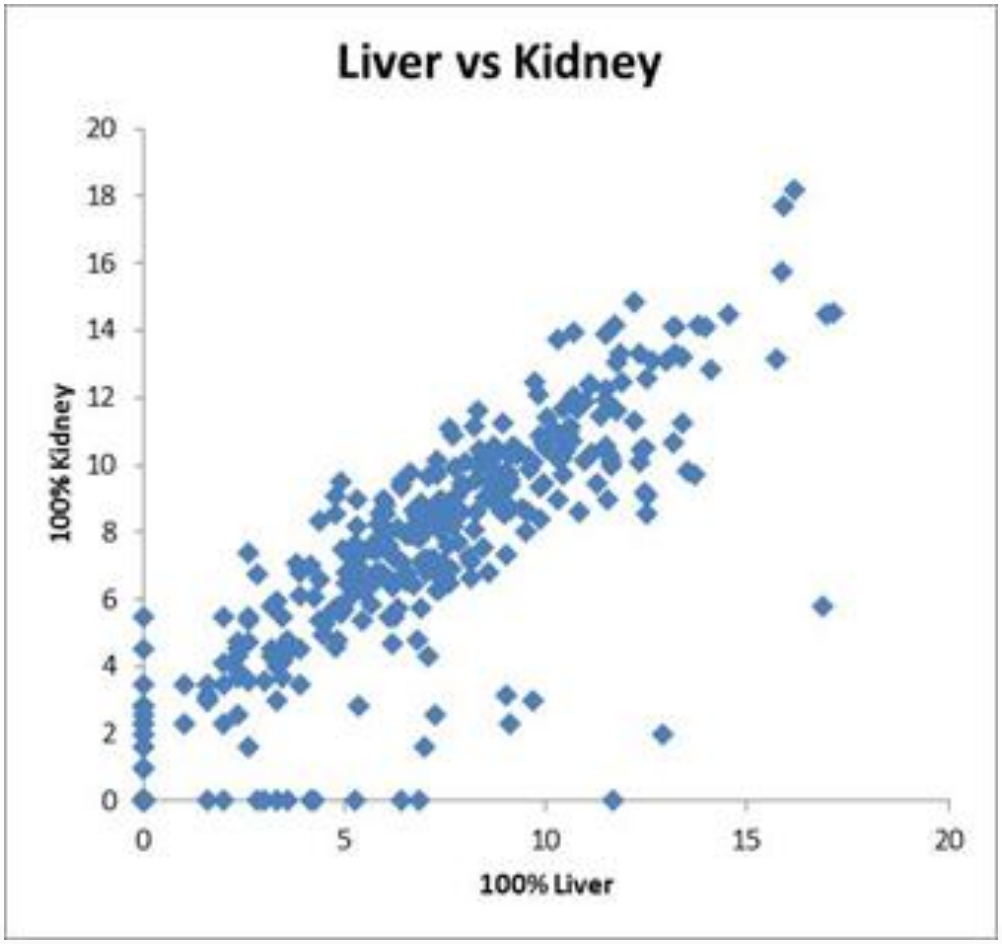
NIH
National Center for Advancing Translational Sciences

RASL-Seq Results

50-of-384 samples > 5 million reads on MiSeq

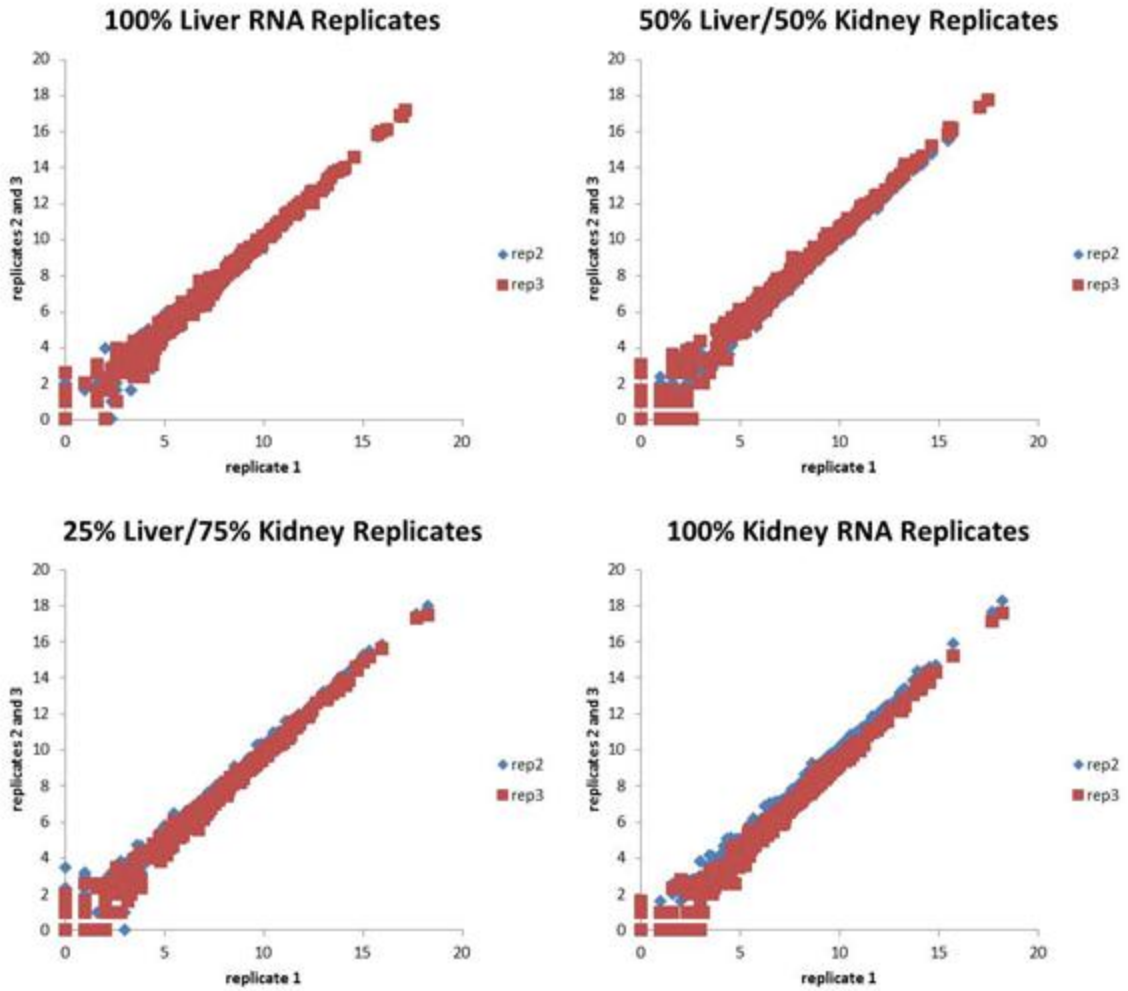


RASL-Seq Describes Differential Expression in Liver vs Kidney for First 120 Genes



Data courtesy of J. Yeakley, BioSpyder, 14 March 2014

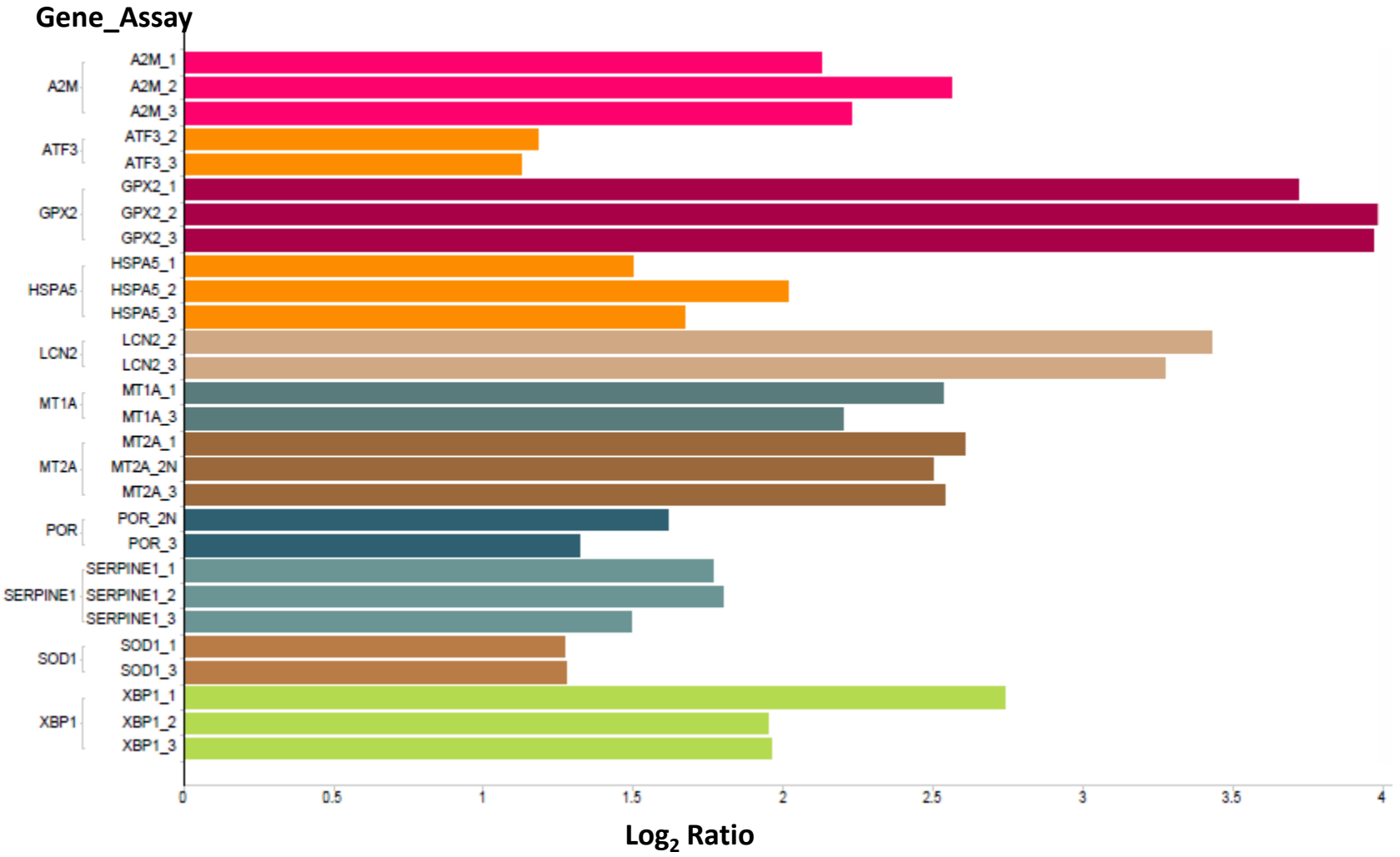
Intra-Experiment Variability is Low in RASL-Seq



The twelve replicate pairwise R^2 values ranged from 0.990 to 0.995

Data courtesy of J. Yeakley, BioSpyder
14 March 2014

Redundant RASL-Seq Assays Give Similar T/C Ratios



Tox21 Gene Set Selection

Select core set of 1000 genes to assay on all cell lines & samples.

Working Group established to select ~1,000 genes for humans…then other species. Led by Rick Paules

- to: identify ~ 1000 genes that will optimally represent the gene expression responses of the entire genome to diverse chemical and biological challenges.
- We reviewed costs. Selecting new genes and assays will cost ~$90/gene

NIH National Center for Advancing Translational Sciences

Typical RASL-Seq Application

Example:

- 60 compounds selected from primary mitochondrial membrane permeability screen Tox21
- Treat 3 models: (*e.g.* HepG2, LUHMES dopa-neurons, cardiomyocytes) x 2 concentrations x 2 time pts x 60 cpds = 720 treatments (2 x 384-well plates x 3 reps?).

Interpretation:

- Do cpds fall into groups that imply mechanisms?
- Do gene expression changes suggest mechanisms?
- Do cellular models react differently to same cpds?

Data Analysis & Interpretation

Data analysis pipeline:

- Filter to remove low quality sequences
- Deconvolute pooled sequences to 384 samples *via* barcodes for rows & columns
- Count matches to target genes (*e.g.* >45/50nt). Adapt NCATS extant pipeline for RNAseq data
- Perform statistical tests for significance for each assay (treated vs controls). Use median if 2-3 assays/gene?

John Braisted's scripts.
Can be automated

Data interpretation:

- Associate cpds that cause similar gene expression profiles. Does cpd profile resemble a reference perturbation profile?
  - ➢ Implies similar mode or mechanism of toxicity.
- Do early-responding genes *describe* the mode, *e.g.* DNA repair genes or E.R. stress genes?
- Clarify hypotheses by referring to RNAseq baseline data for each cell line. *E.g.* Is the hypothesized receptor/transcription factor/pathway transcribed in these cells?
- Do cellular models react differently to same cpds?

NIH National Center for Advancing Translational Sciences

How Would a Baseline RNAseq Dataset for Each Cell Line Inform RASL-Seq Data? Example:

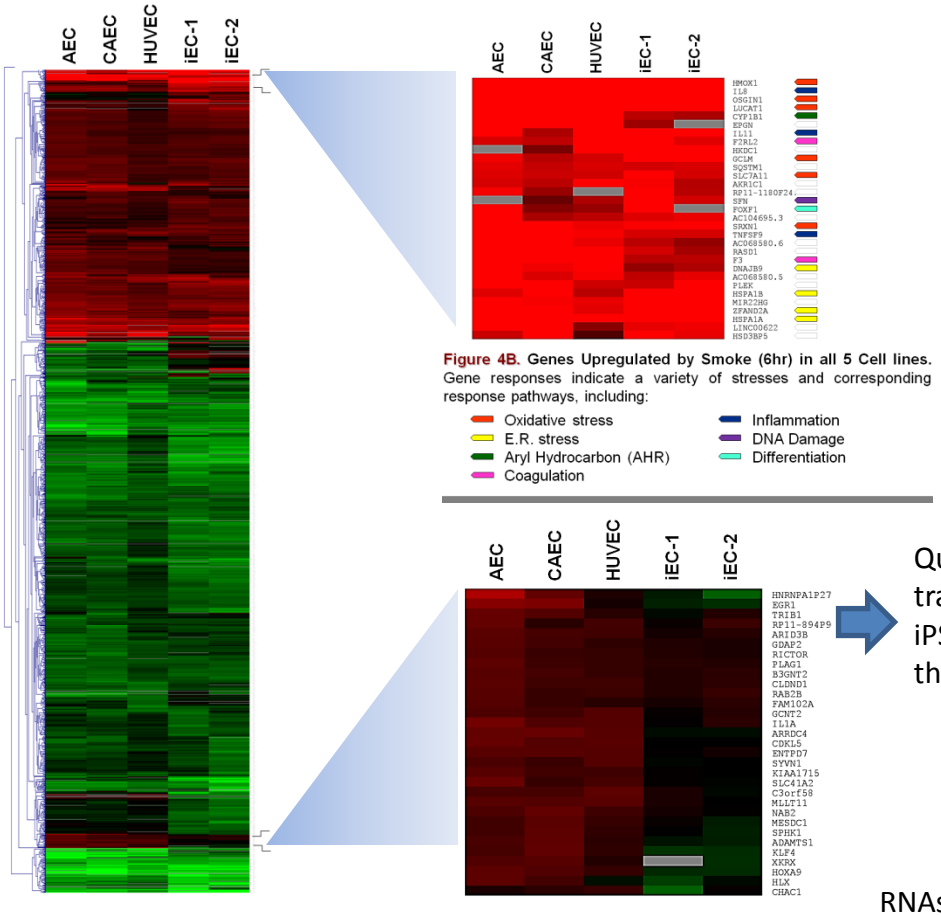


Figure 4B. Genes Upregulated by Smoke (6hr) in all 5 Cell lines. Gene responses indicate a variety of stresses and corresponding response pathways, including:

- Oxidative stress
- E.R. stress
- Aryl Hydrocarbon (AHR)
- Coagulation
- Inflammation
- DNA Damage
- Differentiation

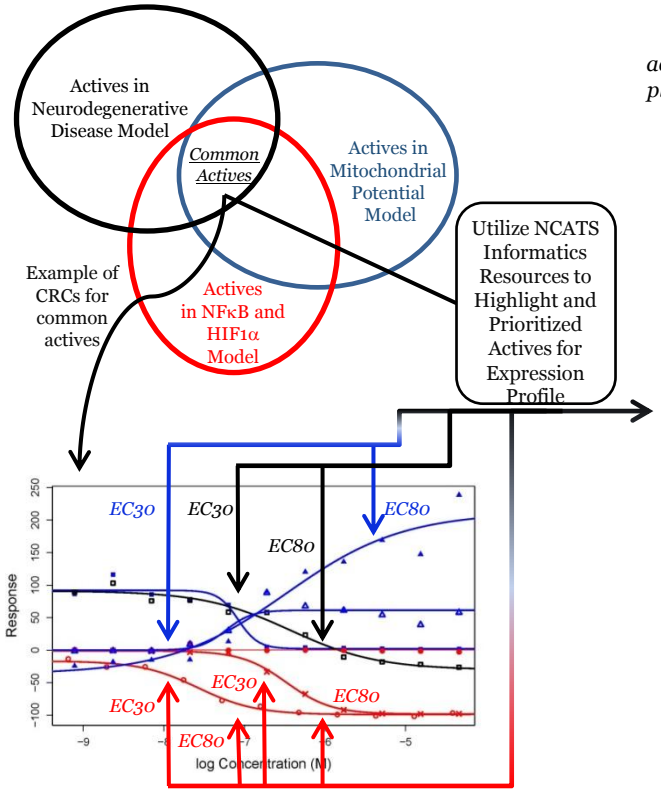
RNAseq (mRNA) Responses to Tobacco Smoke (6hr) in iPS-derived Endothelial Cells (iECs), and Primary Endothelial Cells

Question: Is a developmental transcription factor missing from the iPS-derived cells that could explain these nonresponsive genes?

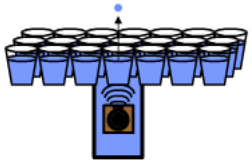
Examine reference RNAseq data

RNAseq data suggest GATA6 is expressed only in the primary cells. Testable Hypothesis

Disease Cell Models	*General Phenotype Cell Models*	*Pathway Informing Cell Models*
1) Cancer • Multiple Myeloma • Viability & Apoptosis • Rhabdomyosarcoma and Neuroblastoma • Viability & Apoptosis	**1)** Autophagy • ~~HEK 293A line~~	**1)** NFκB • Me-180 line
2) Diabetes • INS-1E pancreatic β-cells • ~~Insulin secretions~~ In	**2)** Mitochondrial Potential • Parkin line	**2)** ~~STAT3~~ • ~~HEPG2 line~~
3) Neurodegenerative Diseases • Hela line w/GFP-parkin • ~~Parkin Translocation~~	**3)** DNA Damage/Repair • P53 HCT-116 line	**3)** HIF1α • ME-180 line
4) Innate Immunity • A549 line w/GFP-RSV • RSV spread • Hela line w/GFP-Vaccinia • Vaccinia spread	**4)** TNFα secretion • THP-1 line **5)** ROS Induction/Mitigation • HEPG2 line **6)** Type 1 interferon response • *h*Fibroblast line **7)** Lipid Droplet accumulation • 3TL1 line	**4)** ~~NRF2~~ • HEPG2 line **5)** AP1 • ME-180 line **6)** Estrogen Receptor • HEK293 line **7)** Androgen Receptor • HEK293 line **8)** CREB • HEK293 line

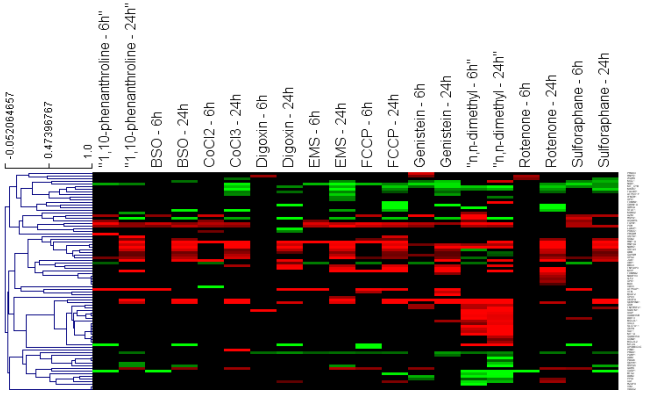


Macros written for EDC acoustic dispenser allow facile plating of each small molecule perturbagen at custom concentrations



Established protocols allow rapid RASL-Seq gene capture on 1000 genes per sample in 384-well plate format

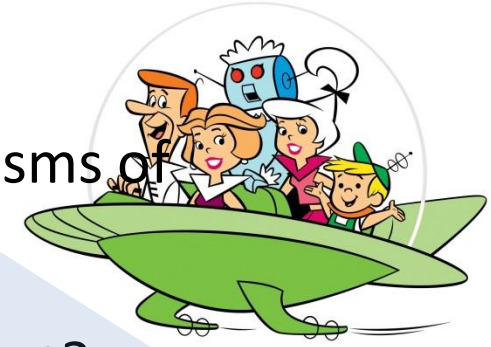
Comparative analysis of expression signatures from selected, active and mirrored inactive perturbagens



National Center for Advancing Translational Sciences

Summary

- RASL-Seq Selected, Results
- RASL-Seq platform works well
- Industrialization proceeding
- Selecting 1,400 genes to represent the genome(s)
- Goal: HT Gene Expression Core Facility
- How to apply it? Elucidate modes and mechanisms of toxicities or AEs. ‘Secondary screening’
- How will we use these results for risk assessment?



Collaborators

"Vieles ist bekannt, aber leider in verschiedenen Kopfen", W. Kollath
("much is known, but unfortunately in different heads")

NCATS

John Braisted
David Kuo
Pei-Hsuan Chu
David Gerhold
Anton Simeonov

Tox21 – NTP & EPA

Ray Tice
Rick Paules

BioSpyder Corp.

JoAnn Yeakley
Joel McComb
Bruce Seligmann

NHLBI

National Heart Lung and Blood Institute

Manfred Boehm
Avram Walts

NHLBI Seq Core

Jun Zhu
Yan Luo
Poching Liu

NIH National Center for Advancing Translational Sciences