Using CRISPR Genomic Screening to Examine Gene-Environment Interactions

Chris Vulpe

Amin Sobh, Graduate Student
Mani Tagmount, Scientist
Alex Loguinov, Bioinformatics
Quan Lu, Collaborator, Harvard University
Luoping Zhang, Collaborator, UC Berkeley
Martyn Smith, Collaborator, UC Berkeley
Silvia Balbo, Collaborator, Univ. of Minnesota
Functional Toxicology to reveal Cellular Adverse Outcome Pathways

Molecular Initiating Event

Toxicant Import

Toxicant Transport

Metabolism

Interacting Cellular Component 1

Interacting Cellular Component 2

Target 1

Adverse Outcome 1

Damage Repair

Metabolite Transport

Export

Molecular Initiating Event

Everything Else

Cellular Stress Response

Interacting Cellular Component 3

Target 2

Adverse Outcome 2

Generic Cell
Functional Toxicology to reveal Cellular Adverse Outcome Pathways

Molecular Initiating Event

Toxicant

Import

Transport

Metabolism

Metabolite

Export

Everything Else

Knock out a gene involved in response to toxicant

Cellular Stress Response

Damage Repair

Increased

Adverse Outcome 1

Adverse Outcome 2

Target 1

Target 2

Interacting Cellular Component 1

Interacting Cellular Component 2

Interacting Cellular Component 3

Increased

Knock out a gene involved in response to toxicant
Targeted CRISPR vs Genome Wide CRISPR

Gene of Interest

Targeting sgRNA to gene of interest

Introduce into cell/organism

Screen for KO

Isolate KO cell/organism

Assess Function in cell/organism

Gene 1 KO

Multiple genes of interest

Targeting sgRNA to genes of interest

(CRISPR KO library)

Introduce into cell

One per cell

Gene 1 KO

Gene 2 KO

Gene 3 KO

Generate pool of mutants

Screen for sensitivity to toxicant to identify and RANK the important genes

Gene n KO
Genome Wide CRISPR in Toxicology

Each KO is individually flagged with a unique molecular barcode so they can be tracked.

Comparing the growth of each KO in the toxicant to growth without the toxicant.

- **Gene 1 KO**: Grows okay
- **Gene 2 KO**: Grows poorly
- **Gene 3 KO**: Grows well
- **Gene n KO**: Not involved

Collect and count flags:

- Grows okay
- Sensitive
- Grows poorly
- Resistant
- Grows well
Key concepts/confusions in genome wide CRISPR screening

- "In vitro" - Using cell lines with all the accompanying issues and caveats –
  - e.g. metabolism, immortalized cells, toxicokinetics

- Any or every gene can be targeted in your library BUT

- Only a single gene is inactivated (KO) in each cell

- A pool (library) of individual mutant cells each containing a KO of single gene represents all genes

- The gene on each chromosome are KOd but the mutations are different on each chromosome

- Each cell with a KO is TAGGED/FLAGGED with unique DNA barcode (sgRNA) so you can see it in a crowd (pool)

- Generally measuring growth advantage or disadvantage of mutant cells in response to environmental exposure such as toxicant
<table>
<thead>
<tr>
<th>Cell line</th>
<th># Genes</th>
<th>Toxicant</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL-60 (Human AML Leukemia)</td>
<td>7114</td>
<td>6-TG, etoposide</td>
<td>Science 2014, 343, 80–84</td>
</tr>
<tr>
<td>A375 (melanoma)</td>
<td>18000</td>
<td>BRAF inhibitor</td>
<td>Nature 2015, 517, 583–588</td>
</tr>
<tr>
<td>HUES62 (ES cell)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K562</td>
<td>16000</td>
<td>DPT</td>
<td>Cell 2014, 159, 647–661</td>
</tr>
<tr>
<td>Human red blood cell leukemia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HepG2</td>
<td>18080</td>
<td>Triclosan</td>
<td>EST, 2016; 50(19): 10682-92</td>
</tr>
</tbody>
</table>
## Acetaldehyde and Arsenic Trioxide Toxicity

### Acetaldehyde
- Primary oxidative metabolite of ethanol
- Genotoxic
- Group 1 carcinogen (IARC)
- Likely underlies alcohol-associated cancers
- Mechanisms of toxicity are poorly understood

### Arsenic Trioxide
- Arsenic Trioxide used in blood cancer treatment – metabolized to usual As metabolites
- Arsenics - IARC type I carcinogen
- Drinking water exposure
- Mechanisms still controversial (depending on who you ask)

An abundance of mechanisms – there are too many mechanisms and it is unclear which are the most important

Can whole genome CRISPR give us some insights into the cellular mechanisms and maybe their relative importance?
What did we do?

Collect the cells that survive
Count molecular BARCODES from Toxicant/Control (SEQUENCING)

Sensitive Mutants
Relative Molecular Barcode Abundance
- Toxicant + Toxicant

Some mutant CRISPR KO mutants will depleted in presence of toxicant as compared to its relative abundance in control

Resistant Mutants
Relative Molecular Barcode Abundance
- Toxicant + Toxicant

Some mutant CRISPR KO mutants will enriched in presence of toxicant as compared to its frequency in control

K562 (human PRE-RED BLOOD CANCER cell line)

Gene 1 KO
Gene 2 KO
Gene 3 KO
Gene n KO

~18,000 genes

Gecko A Knock out Library

1 µM Arsenic Trioxide Control
2.5 mM Acetaldehyde Control

Lentivirus 0.3 MOI Puromycin

Use Dose that decreases growth by ~50% Of NORMAL cell

7 days/7 doublings

Collect the cells that survive
Count molecular BARCODES from Toxicant/Control (SEQUENCING)
Primary Screen

Each point represents a sgRNA – 3 sgRNA per gene

CRISPR KO screen GECKO A library, 3sgRNA gene, ~IC30-50, 7 days

Candidates for 2° screen
1, 2, or 3 sgRNA
FDR Cut-off: 0.1
Resistant mutants: 74
Sensitive mutants: 26

Analysis similar to RNA-seq – multiple sgRNA per gene is new twist

AsO₃ exposed vs control

Measure of mutant cell abundance in pool
## Arsenic Trioxide whole genome CRISPR screen

**CRISPR KO screen** GECKO A library, 3sgRNA gene, ~IC30-50, 7 days

### Top 10 - Whole Genome Screen Candidates

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene name</th>
<th>logFC</th>
<th>P Value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEAP1</td>
<td>kelch-like ECH-associated protein 1</td>
<td>2.05</td>
<td>3.13E-596.87E-55</td>
<td></td>
</tr>
<tr>
<td>SEPHS2</td>
<td>selenophosphate synthetase 2</td>
<td>1.77</td>
<td>1.88E-232.06E-19</td>
<td></td>
</tr>
<tr>
<td>EEFSEC</td>
<td>eukaryotic elongation factor, selenocysteine-tRNA-specific</td>
<td>1.25</td>
<td>1.09E-177.97E-14</td>
<td></td>
</tr>
<tr>
<td>PSTK</td>
<td>phosphoseryl-tRNA kinase</td>
<td>1.49</td>
<td>3.23E-171.77E-13</td>
<td></td>
</tr>
<tr>
<td>KRT73</td>
<td>keratin 73</td>
<td>-2.5</td>
<td>2.88E-151.26E-11</td>
<td></td>
</tr>
<tr>
<td>ARID1B</td>
<td>AT rich interactive domain 1B (SWI-like)</td>
<td>1.42</td>
<td>5.44E-131.99E-09</td>
<td></td>
</tr>
<tr>
<td>TXNDC17</td>
<td>thioredoxin domain containing 17</td>
<td>0.9</td>
<td>3.20E-101.00E-06</td>
<td></td>
</tr>
<tr>
<td>SLC6A12</td>
<td>solute carrier family 6 (neurotransmitter transporter), member 12</td>
<td>0.92</td>
<td>8.66E-102.37E-06</td>
<td></td>
</tr>
<tr>
<td>DCLRE1A</td>
<td>DNA cross-link repair 1A</td>
<td>-1.1</td>
<td>5.52E-091.34E-05</td>
<td></td>
</tr>
<tr>
<td>DLGAP5</td>
<td>discs, large (Drosophila) homolog-associated protein 5</td>
<td>-1.1</td>
<td>2.91E-086.38E-05</td>
<td></td>
</tr>
</tbody>
</table>

**FDR** – False Discovery Rate    **Log FC** – relative abundance in treated vs control
Arsenic Trioxide confirmatory CRISPR screen

<table>
<thead>
<tr>
<th>Gene</th>
<th>sgRNA</th>
<th>FDR</th>
<th>Log FC</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEAP1</td>
<td>8/8</td>
<td>0.000354</td>
<td>3.6</td>
</tr>
<tr>
<td>TXNDC17</td>
<td>8/8</td>
<td>0.000354</td>
<td>1.4</td>
</tr>
<tr>
<td>PSTK</td>
<td>7/7</td>
<td>0.000354</td>
<td>1.6</td>
</tr>
<tr>
<td>GFI1B</td>
<td>7/7</td>
<td>0.000354</td>
<td>1.1</td>
</tr>
<tr>
<td>SLC30A1</td>
<td>7/7</td>
<td>0.000354</td>
<td>1</td>
</tr>
<tr>
<td>FLCN</td>
<td>7/7</td>
<td>0.000354</td>
<td>1.3</td>
</tr>
<tr>
<td>EED</td>
<td>7/7</td>
<td>0.000354</td>
<td>0.7</td>
</tr>
<tr>
<td>RRAGC</td>
<td>8/8</td>
<td>0.000354</td>
<td>1</td>
</tr>
<tr>
<td>EEFSEC</td>
<td>6/7</td>
<td>0.000354</td>
<td>1.6</td>
</tr>
<tr>
<td>C15orf41</td>
<td>7/7</td>
<td>0.000354</td>
<td>0.6</td>
</tr>
<tr>
<td>SET</td>
<td>7/8</td>
<td>0.000354</td>
<td>0.8</td>
</tr>
<tr>
<td>SEPHS2</td>
<td>6/7</td>
<td>0.000354</td>
<td>1.4</td>
</tr>
<tr>
<td>SEPSECS</td>
<td>7/8</td>
<td>0.000354</td>
<td>0.7</td>
</tr>
<tr>
<td>DPH6</td>
<td>6/7</td>
<td>0.000354</td>
<td>0.8</td>
</tr>
<tr>
<td>NAA38</td>
<td>8/8</td>
<td>0.000928</td>
<td>0.7</td>
</tr>
</tbody>
</table>

sgRNA – the number of sgRNA for each gene

8/8 means 8 sgRNA out of 8 tested showed effect

Log FC – average relative abundance in treated vs control

FDR – False Discovery Rate

Gene sgRNA FDR Log FC
ABCC1 8/8 0.000619 -2.1
MTPN 7/7 0.000619 -0.7
NCAPD3 6/7 0.000619 -0.7
DEPDC5 7/7 0.000619 -0.4
UBE2H 7/8 0.000619 -0.6
NPRL2 6/6 0.000619 -0.3
CNOT2 7/7 0.000619 -0.6
NDE1 7/8 0.000619 -0.7

Resistant
- 100 Arsenic candidates
- 74 resistant
- 26 sensitive
- 2874 total genes
ATO Toxicity: Reactive Oxygen Species

Nrf2 primary anti-oxidant transcription factor – KEAP1 is REPRESSOR of NRF1

Increased resistance with KEAP1 KO

KEAP1 Inhibits Nrf2

ATO ?

ROS

We knew this already!

Suggests it is very important in ACUTE short term toxicity

Te-Chang Lee, I-Ching Ho, Wen-Jen Lu, and Jin-ding Huang. Enhanced Expression of Multidrug resistance-associated Protein 2 and Reduced Expression of Aquaglyceroporin 3 in an Arsenic-resistant Human Cell Line J. Biol. Chem. 2006 281: 18401-

Michael W. Carew, Elaine M. Leslie; Selenium-dependent and -independent transport of arsenic by the human multidrug resistance protein 2 (MRP2/ABCC2): implications for the mutual detoxification of arsenic and selenium, Carcinogenesis, Volume 31, Issue 8, Pages 1450–1455,
**Selenocysteine Incorporation into Proteins Increases Susceptibility to Arsenic Trioxide**

Selenocysteine
The 21st amino acid

Selenium
Previously known As binds Se

What did we find?
Loss of any gene needed for SeCys Incorporation in proteins leads to RESISTANCE

THIS IS NEW AND KINDA UNEXPECTED

Many of these proteins involved in response to oxidative stress

**Resistance of PSTK−/− K562 cells to ATO validated by cell viability assay**

Se-As complex hypothesis

1938 – Moxon noted that Arsenic exposure can prevents Selenium poisoning (L.A. Moxon, Science, 88 (1938), p. 81)

Mutual antagonism (Se prevents As poisoning too) in multiple species including people (reviewed in Environment International 69 (2014) 148–158)

As-Se complexes formation may play a role

Inside the cell – glutathione complexes form

As + GSH \rightarrow (GS)_{2}As–OH

(GS)_{2}As–OH + Se \rightarrow [(GS)_{2}AsSe]–

[(GS)_{2}AsSe]– \rightarrow Less toxic and possibly exported

The Seleno Bis(S-glutathionyl) Arsinium Ion Is Assembled in Erythrocyte Lysate
Chemical Research in Toxicology 2006 19 (4), 601-607

Protective effect of selenium (Na_{2}SeO_{3}) pre-treatment on ATO cytotoxicity in K562 cells.
Arsenic Trioxide Summary

- Oxidative stress is (the) major player in acute arsenic toxicity – Nrf2 implicated
- Disrupting Selenocysteine incorporation increases ATO resistance - Mechanism unclear – several hypotheses
- Arsenic transport by AQP2 and MRP1
- Additional genes implicated in both sensitivity/resistance
Blocking DNA Repair Increases Susceptibility to Acetaldehyde

Validated Acetaldehyde Susceptibility Candidates

<table>
<thead>
<tr>
<th>Gene</th>
<th>sgRNA sequence</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>OVCA2</td>
<td>GCCGAGCTCGTGTGCCTCAG</td>
<td>4.78E-12</td>
</tr>
<tr>
<td>OVCA2</td>
<td>GACACCAAGAGGATAACCGG</td>
<td>8.91E-10</td>
</tr>
<tr>
<td>OVCA2</td>
<td>TGCTTCACCGAAGGATCTGC</td>
<td>8.91E-10</td>
</tr>
<tr>
<td>HELQ</td>
<td>TGCTGGAATAGATACTATTG</td>
<td>1.39E-08</td>
</tr>
<tr>
<td>HELQ</td>
<td>GGAGTTGCTATACACACAG</td>
<td>1.20E-07</td>
</tr>
<tr>
<td>OVCA2</td>
<td>CCGGGGCTCCGTAAGAACAC</td>
<td>9.18E-07</td>
</tr>
<tr>
<td>HELQ</td>
<td>GTTGACAGCAGAAGCTGAGAA</td>
<td>9.22E-07</td>
</tr>
<tr>
<td>HELQ</td>
<td>TCTGATCAGTGGGTAGCAGA</td>
<td>9.73E-07</td>
</tr>
<tr>
<td>OVCA2</td>
<td>TTCCAATGCGGAGAAGACCT</td>
<td>9.73E-07</td>
</tr>
<tr>
<td>OVCA2</td>
<td>GGGCTTCCGTAAGAAGACC</td>
<td>1.66E-06</td>
</tr>
<tr>
<td>HELQ</td>
<td>TGAAGTATATATCCTCAATCA</td>
<td>4.16E-06</td>
</tr>
<tr>
<td>ERCC8</td>
<td>GCAAGATATATGCTAGACAC</td>
<td>0.000198</td>
</tr>
<tr>
<td>ERCC5</td>
<td>TTAATGGCTGAAAGAGGCTG</td>
<td>0.0003243</td>
</tr>
<tr>
<td>ERCC8</td>
<td>CAGGGGTATCCTGATGAC</td>
<td>0.0003705</td>
</tr>
<tr>
<td>PPP4R2</td>
<td>CATGACAGAAGAACATGACCC</td>
<td>0.0003731</td>
</tr>
<tr>
<td>PPP4R2</td>
<td>TTCCAATGCGGAGAAGACCT</td>
<td>0.0005561</td>
</tr>
<tr>
<td>OVCA2</td>
<td>GAGGGCAGCAGATCGACCT</td>
<td>0.0008455</td>
</tr>
<tr>
<td>HELQ</td>
<td>ACCAATGCTACCAGTGAC</td>
<td>0.0022578</td>
</tr>
<tr>
<td>OVCA2</td>
<td>CAAGTGCCGAGCGGTCTTTCC</td>
<td>0.0025663</td>
</tr>
<tr>
<td>ERCC8</td>
<td>TGTTAAGCGAGGCTGTTCCAT</td>
<td>0.0028369</td>
</tr>
<tr>
<td>NANS</td>
<td>GAGATCGGCGGACAACCCA</td>
<td>0.020249</td>
</tr>
<tr>
<td>NANS</td>
<td>TAGTGAGCAGTTCAACCTCT</td>
<td>0.0524015</td>
</tr>
<tr>
<td>FBXO40</td>
<td>AACCTCCGGCTTTAAGGGCA</td>
<td>0.0648084</td>
</tr>
<tr>
<td>UVSSA</td>
<td>AATGGAATCTGGGCTTGAGG</td>
<td>0.0769292</td>
</tr>
<tr>
<td>HELQ</td>
<td>CTTATCTTTACCTCGACGG</td>
<td>0.0776389</td>
</tr>
</tbody>
</table>

ERCC5 – XPG – 3’ endonuclease
ERCC8 – CSA – TC-NER
UVSSA – TC – NER pathway

HELQ – SS DNA Helicase – ICL repair
PPP4R2 – PPP4 subunit – DSB repair

OVCA2 – Ovarian cancer associated

We did not find Brca1, Brca2 or any FA genes
sgRNA targeting ineffective?
Not stringent enough screen?
Involved at lower/higher doses?

Increased Susceptibility to Acetaldehyde Toxicity

Something OLD

Modified from: Hira et al., Blood, 2016
Transcription coupled repair

Potential Role of **OVCA2** in DNA Repair

**OVCA2**
- Ovarian tumor suppressor candidate 2
- Strongest candidate in our screen
- Validated with 8 sgRNAs out of 8 in a secondary screen
- Loss-of-function increases sensitivity to Acetaldehyde
- Predicted hydrolase (esterase) activity
- Downregulated in multiple cancer types
- Yeast homolog (*FSH1*) is essential for growth in ethanol media
- What the heck is it?

Something NEW, and blue of course
OVCA2 KO increases acetaldehyde sensitivity and DNA adduct levels

Increased sensitivity of OVCA2−/− K562 cells to acetaldehyde. Cell viability in the presence of 5 mM acetaldehyde (24 hr) is represented as percentage of untreated controls. NTC corresponds to non-targeting sgRNA control. OVCA1-1 and OVCA-2 are 2 different sgRNAs targeting OVCA2. * p-value < 0.05.

Increased accumulation of the acetaldehyde-induced DNA adduct N2-Ethyl-2-deoxyguanosine (N2EtdG) in OVCA2−/− K562 cells. NTC corresponds to non-targeting sgRNA control. OVCA1-1 and OVCA-2 are 2 different sgRNAs targeting OVCA2.
What did we learn?

- Oxidative stress is (the) major player in acute arsenic trioxide toxicity
- Selenium metabolism is important in acute arsenic toxicity
- DNA damage is important in acute acetaldehyde toxicity
- Unexpected role for OVCA2 – a new DNA repair gene, DNA damage tolerance?

Implications?

- Help fill in adverse outcome pathway for Arsenic

As → ROS → BAD THINGS

- Suggest selenium deficiency could decrease acute effects and selenium sufficiency could increase acute effects – Public Health Implications?
- Acetaldehyde – confirm genotoxic mechanism – suggest DNA damage also important for Acute toxicity
- OVCA2 is tumor suppressor gene—maybe role in DNA repair explains why
ToxCrispr

- Quan Lu and Luoping Zhang

- 3675 Toxicology-related genes
  - S1500+ gene set prioritized by NIEHS/NTP/Tox21 program
  - 647 Environmental Genome Project (EGP) genes
  - Selected toxicant response-focused genes,

- Subset CRISPR library for probing toxicology mechanisms

- Can use less cells – need only 7.5 million vs 30 million for whole genome

- Enable more rapid

Revigo scatter plot of enriched GO terms

http://revigo.irb.hr/revigo.jsp
Acknowledgments

Amin Sobh
Graduate Student

Mani Tagmount
Scientist

Luoping Zhang

Martyn Smith

Quan Lu
Harvard

Alessia Stornetta

Silvia Balbo
Univ. of Minnesota

Alex Loguinov
Bioinformatics

UC BERKELEY SUPERFUND RESEARCH PROGRAM
SCIENCE FOR A SAFER WORLD

UF Center for Environmental & Human Toxicology
College of Veterinary Medicine

UF UNIVERSITY OF FLORIDA

Srp
Superfund Research Program

NIEHS