Carcicast: Developing a Carcinogenicity Testing Toolbox

Nicole C. Kleinstreuer, PhD
Deputy Director, NICEATM
PI, Computational Toxicology Group, DIR/BCBB
National Institute of Environmental Health Sciences

29th April, 2019

Converging on Cancer Workshop
• National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)
• Division of Intramural Research, Biostatistics and Computational Biology Branch
Disclaimer: Thoughts and examples provided herein are my own opinions. Mention of trade names or technologies does not constitute endorsement.
Constructing “CarciCast”*

- Expert-driven approach
  - Identify assays/biomarkers that map to key characteristics of carcinogens/hallmarks of cancer

- Semi-supervised systematic review
  - Broad keyword search for all relevant scientific literature, abstract screening and tagging

- Applying HTS data
  - Prioritize environmental chemicals based on bioactivity against targets that map to cancer hallmark pathways
  - Construct QSAR models for key characteristics

*Naming Credit: Keith Houck, EPA/NCCT
Expert-driven Approach

- Understanding the relationship between hallmarks (HM: biology) and key characteristics (KC: chemistry)
- Which are measurable, and in what platforms?
- KC: Chemical properties with associated targeted assays
- HM: Biological properties requiring integrated models
No Clear One-to-one Relationship

Carcinogens induce one or more KC’s at one or more points in the process (i.e. initiation/promotion)

Tumors acquire one or more HMs at various points in the carcinogenic process

M. Fielden, Amgen
Converging Effects

Carcinogen

KC 1 Electrophilic
KC 2 Genotoxic

KC 3 Alters DNA repair or induces genomic instability

KC 5 Oxidative stress
KC 6 Inflammation

DNA Damage

Initiate and/or augment the Hallmarks

M. Fielden, Amgen
Building the CarciCast Toolbox

Key Characteristics of Carcinogens can be used to define a toolbox to assess new/untested chemicals or agents for carcinogenic hazards

- Manuscript in preparation – (Fielden et al. 2019) The Key Characteristics of Carcinogens: Relationship to the Hallmarks of Cancer and Assays and Biomarkers to Measure Them

M. Fielden, Amgen
Challenges

- Lack of well established or “gold standard” assays:
  - Eg. KC 6: Induces chronic inflammation, KC 5: Induces oxidative stress

- How to identify & characterize the most appropriate assays?
  - Endpoints specific and relevant to carcinogenic process?
  - Rationale for concentration/dose selection

- How to integrate results from multiple KC’s?

- How to relate in vitro results to realistic in vivo exposures?

- In vivo biomarkers of the KC’s in animals/humans needed to understand in vitro-in vivo translation and risk assessment
Measuring the Hallmarks: Complex Systems Models

• Ideal Characteristics:
  – Human-relevant platform
  – Ability to measure interdependent biological responses
  – Provide insight into tumor “tipping point”
    • Temporal and biological
  – Query impact of dose, frequency, repetition, duration, and multiplicity of exposures
  – Represent genetic differences in susceptibility, resilience, and resistance
3D models: Repurposing Drug Development Platforms

3D human breast tumor models have been bioprinted with defined multi-cellular composition and architecture.

Puls et al. Nat Sci Rep 2018
# BioMAP Oncology Panels

## Human Primary Cells + Microenvironment

### Human Biology Modeled by Oncology Panels

<table>
<thead>
<tr>
<th>Panel</th>
<th>System</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorectal Cancer (CRC) Panel</td>
<td>Colorectal Cancer - Stro</td>
<td>The Colorectal Cancer - Stro (StroHT29) system models the host stromal-tumor microenvironment by capturing the complex interactions between tumor cells, the host stromal network, and infiltrating immune cells recruited into the tumor mass.</td>
</tr>
<tr>
<td>Colorectal Cancer (CRC) Panel</td>
<td>Colorectal Cancer - Vasc</td>
<td>The Colorectal Cancer - Vasc (VascHT29) system models host vascular-tumor microenvironment by capturing the complex interactions between tumor cells, the host vascular network, and infiltrating immune cells associated with angiogenesis.</td>
</tr>
<tr>
<td>Non-Small Cell Lung Cancer (NSCLC) Panel</td>
<td>Lung Cancer - Stro</td>
<td>The Lung Cancer - Stro (StroNSCLC) host-NSCLC tumor microenvironment model system consists of human primary fibroblasts co-cultured with a NSCLC cell line, NCI-H1299, and human peripheral blood mononuclear cells. These conditions model the host stromal-tumor microenvironment by capturing the complex interactions between tumor cells, the host stromal network, and infiltrating immune cells recruited into the tumor mass.</td>
</tr>
<tr>
<td>Non-Small Cell Lung Cancer (NSCLC) Panel</td>
<td>Lung Cancer - Vasc</td>
<td>The Lung Cancer - Vasc (VascNSCLC) host-NSCLC tumor microenvironment model system consists of human primary vascular endothelial cells co-cultured with a NSCLC cell line, NCI-H1299, and human peripheral blood mononuclear cells. These conditions model the host vascular-tumor microenvironment by capturing the complex interactions between tumor cells, the host vascular network, and infiltrating immune cells associated with angiogenesis.</td>
</tr>
</tbody>
</table>
## Transformics Assay

<table>
<thead>
<tr>
<th>3-MCA</th>
<th>Main molecular endpoints</th>
<th>Transformation phenotypic endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.04 µg/ml</td>
<td>24h</td>
<td>72h</td>
</tr>
<tr>
<td>Cell adhesion mechanisms</td>
<td>Cell cycle regulation</td>
<td>Apoptosis regulation</td>
</tr>
<tr>
<td>4 µg/ml</td>
<td>Immune response</td>
<td>Cytoskeleton remodeling</td>
</tr>
<tr>
<td>Immune response</td>
<td>Apoptosis regulation</td>
<td>Nos statistically significant gene modulation</td>
</tr>
</tbody>
</table>

Mascolo et al. 2018, *Carcinogenesis*

- Combining the cell transformation assay with transcriptomics
- Identify dose- and time-dependent signals, discriminate adaptive from adverse responses
Multiplexed HM-related Endpoints In Vitro

Wilde et al. Arch Tox 2018

<table>
<thead>
<tr>
<th>Chemical</th>
<th>In vivo TD\textsubscript{50}</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCDD</td>
<td>0.000023 mg/kg/day</td>
<td>rat</td>
</tr>
<tr>
<td>MNU</td>
<td>0.0927 mg/kg/day</td>
<td>rat</td>
</tr>
<tr>
<td>MMS</td>
<td>32 mg/kg/day</td>
<td>mouse</td>
</tr>
<tr>
<td>MC</td>
<td>56 mg/kg/day</td>
<td>rat</td>
</tr>
<tr>
<td>Acetaldehyde</td>
<td>153 mg/kg/day</td>
<td>rat</td>
</tr>
<tr>
<td>DEHP</td>
<td>716 mg/kg/day</td>
<td>rat</td>
</tr>
<tr>
<td>H\textsubscript{2} O\textsubscript{2}</td>
<td>7540 mg/kg/day</td>
<td>mouse</td>
</tr>
<tr>
<td>NiCl\textsubscript{2}</td>
<td>Data unavailable</td>
<td></td>
</tr>
</tbody>
</table>

Legend:
- p21
- p-p53
- p53
- MN
- SGK1
- CHKA
- Cell area
- Nuclear area
- Seahorse
- G1
- S
- G2

Data unavailable for NiCl\textsubscript{2}
Transgenic Mouse Models

- Ex: STAT1 deficient mouse model (STAT1KO 129S6/SvEvTac-Stat1 tm1Rds, Borowsky lab)

- STAT1 deficiency is a germline mutation, the emergence of tumors requires secondary mutations and/or other adaptations within the microenvironment.

- Prolonged latency supports the “adaptive oncogenesis” theory: changes in the host microenvironment facilitate the expansion of preexisting mutant populations

- Models the most common category of human breast cancer: age related (post-menopausal) ERα+ luminal carcinoma

M. La Merrill, UC Davis
Semi-supervised systematic review

• Initial search strategy:
  Work with NTP Report on Carcinogens and Office of Health Assessment and Translation to identify keywords
  – 256 keywords mapped to HM of Cancer and KC of Carcinogens
  – 7 keywords for assays/biomarker, crossed with HM of Cancer and KC of Carcinogens

• Recruit participants to screen and tag abstracts
  – Metadata: KC, HM, Organism, Publication type, Study type
  – Mesh terms automatically tracked for PubMed articles
Initial corpus

- PubMed + Scopus database
- Literature from last 10 years

TOTAL:
- 32,605 PubMed
- 35,171 Scopus

Without replicates:
57,036 publications

https://sysrev.com/
Sysrev: Semi-automated review platform

https://sysrev.com/

• Freely available website
• Abstract screening and annotating
• Intuitive user interface
• Including mobile/tablet access
• Uses machine learning to rank the corpus

Register
https://sysrev.com/register

Project name
Hallmark and key characteristics mapping

Project link
https://sysrev.com/p/3588
Machine Learning: Inclusion/Exclusion Models

Neural network model produces a predicted score for each article
(0 not relevant, 1 relevant)

https://sysrev.com/p/3588/
Systematic review: an initiative to map cancer hallmarks and key characteristics

Alexandre Borrel\textsuperscript{1,}, Amy Wang\textsuperscript{2,}, Lara Handler\textsuperscript{3,}, N. Kleinstreuer\textsuperscript{1,4}

\textsuperscript{1}NEIHS/DOR/RC3B, RTP, NC; \textsuperscript{2}NEIHS/DNTP/RoC, RTP, NC; \textsuperscript{3}LS, RTP, NC; \textsuperscript{4}NEIHS/DNTP/NEAATM, RTP, NC

**Introduction**

Carcinogenesis is a multi-step process in which normal cells are transformed into cancer cells by acquiring various properties that allow them to form tumors or malignant cancers. These acquired properties of cancer cells that distinguish them from normal cells have been classified as a series of ten Cancer Hallmarks. (Hanahan 2011). From a chemical perspective, a set of key characteristics commonly exhibited by established human carcinogens have been defined and applied by the International Agency for Research on Cancer (IARC) (Smith 2016).

**2. Systematic review strategy**

- Define an initial publication corpus
- Manual reviewing (abstract screening)
- Meta-Analysis
- Final relevant corpus

**3. Initial corpus**

With the NTP Report on Carcinogens and Office of Health Assessment and Translation, we identified 258 keywords mapped to NTP of Cancer and KC of Carcinogens crossed with 7 keywords for assayed biomarkers. To identify the most relevant cutting-edge technologies (or those that are still in widespread use), only publications after 2008 were included, and book chapters/dissertation/theses were excluded. Keyword publication searches were performed on both PubMed and Scopus databases.

**5. Abstract screening**

Syrev provides a web interface to label abstracts. Seven labels can be specified for each abstract as well as biomarker and technology annotation. Each article need two positive reviews to be included in the corpus. In case of conflict a third reviewer need to resolve it.

**4. Syrev platform**

https://sysrev.com

**7. Preliminary results**

As of early April 2019, 2,234 publications have been reviewed. Only 21% of the publication reviewed are included in the final corpus, and cover every KC, and NTPs in varying degrees.

**8. Support**

We provide reviewer support for abstract screening with a Standard Operating Procedure available on the platform. Periodic webinars help standardize reviews by providing tips and decision trees.

**6. Machine learning**

Using a neural network approach, for each publication Syrev computes an inclusion score based on keywords in publications previously included in the corpus. The model is continuously iterating and improving based on reviewer feedback. See below for a distribution for reviewed and unreviewed publications.

**Conclusion**

47 reviewers already joined the project

Join us now!
Applying HTS data

- We observe that some chemicals perturb multiple cancer hallmark pathways.

- Hypothesis: A chemical that perturbs many pathways related to cancer hallmark processes will be more likely to cause cancer in the lifetime of an animal than a chemical that perturbs few such pathways.
  
  - Chemicals can increase cancer risk through many different patterns of pathway perturbations.
  
  - Tox21/ToxCast assays provide decent HM coverage.
Initial Approach

- Link Tox21/Toxcast assays with genes, pathways, cancer hallmarks
  - Use published pathways and Gene Ontology keywords
- Calculate univariate associations
  - *In vitro* assay x *in vivo* cancer endpoints, odds-ratio (OR)
  - Multiple testing corrections with permutation tests
  - Keep associations with OR>2, Lower Confidence Interval > 1
- Rank chemicals by number of hits
- Forward validate with 60 chemicals not in signature development set

*Kleinstreuer et al. 2013, Tox Sci*
Most cancer-related endpoints are linked to hallmark processes or xenobiotic metabolizing enzymes.

External Validation: simple linear model separates Probable/Likely/Possible from Non-Carcinogens (p~0.01).
Current Approach: Biologically-based Bayesian Networks

- Bayesian Networks provide a probabilistic means to predict an outcome based on measured values.

- Can this approach be used to predict chemical carcinogenicity potential from hallmark-related *in vitro* and *in silico* assays?
Hallmark/Gene/Assay mapping

- Identify updated (2019) ToxCast/Tox21 assay targets (~350) mapped to hallmark-related genes
Borrel et al. Poster

Systematic review: an initiative to map cancer hallmarks and key characteristics
Alexandre Borrel1, Amy Wang2, Lara Handler3, N. Kleinsteuber1,4

1. INTRODUCTION
Carcinogenesis is a multistage process in which normal cells are transformed into cancer cells by acquiring various properties that allow them to form tumors or malignant cancers. These acquired properties of cancer cells that distinguish them from normal cells have been classified as a series of ten Cancer Hallmarks (Hanahan, 2011). From a chemical perspective, a set of key characteristics commonly exhibited by established human carcinogens have been defined and applied by the International Agency for Research on Cancer (IARC) (Eilmann, 2016).

2. SYSTEMATIC REVIEW STRATEGY

3. INITIAL CORPUS
With the NTP's (National Toxicology Program) support on Carcinogens and Office of Health Assessment and Translation, we identified 256 keywords mapped to 11 of Cancer and KC of Carcinogens covered with 7 keywords for assay/evaluations. To identify the most relevant cutting-edge technologies for those that are still in widespread use, only publications after 2009 were included, and book chapters/literature/theses were excluded. Keyword publication searches were performed on both PubMed and Scopus databases.

4. SYSTREV PLATFORM
Systrev is a semi-automated review platform. It provides easy access to detailed information and comprehensive screening, making it intuitive and user-friendly, including mobile/tablet access. It uses machine learning to refine the corpus.

5. ABSTRACT SCREENING
Systrev provides a web interface to label abstracts. Seven labels can be specified for each abstract as well as biomarker and technology annotation. Each article received two positive reviews to be included in the corpus. In cases of conflict, a third reviewer was needed to resolve it.

6. MACHINE LEARNING
Using a neural network approach, for each publication, Systrev computes an inclusion score based on the keywords in publications previously included in the corpus. The model is continuously learning and improving based on reviewer feedback. See below for a distribution of reviewed and unreviewed publications.

7. PRELIMINARY RESULTS
As of early April 2019, 2324 publications have been reviewed. Only 15% of the publications reviewed are included in the final corpus, and cover every KC and NTP to varying degrees.

8. SUPPORT
We provide reviewer support for abstract screening with a standard operating procedure available on the platform. Periodic webinars help standardize reviews by providing tips and decision trees.

9. CONCLUSION
42 reviewers joined the project.
QSAR models for Key Characteristics

- Use KC mapping of HTS assays to identify training set chemicals (active/inactive) for each KC-QSAR model
Ongoing Work

• Include ToxCast assays without specific gene targets (e.g. proliferation, mitochondrial function)

• Refine scoring metrics, investigate tissue-specific endpoints, feature selection algorithms to id minimum assay set, id targets missing from HTS, investigate mis-predicted chemicals

• Incorporate informative priors based on systematic literature review results into BN learning

• Combine with KC-QSARs and low-throughput complex mechanistic assays to form integrated testing strategies

• Ultimate goal: probabilistic chemical (complex mixture) screening for carcinogenicity using battery of *in vitro* and *in silico* tests.
Addressing Carcinogenic Risk Probabilistically

Risk

% Population

Prior

Posterior

+ Mechanistic Information for Complex Chemical Exposures (from HTS assays, complex organotypic systems, QSAR models, transgenic animals, etc.)
QUESTIONS?