U.S. DEPARTMENT OF ENERGY | Office of Science

Addressing the NASEM Committee on: A Long-Term Strategy for Low-Dose Radiation Research in the United States

Todd Anderson, Ph.D.

Director, Biological Systems Science Division, Department of Energy, Office of Biological & Environmental Research

January 24, 2022

Current Funding Situation

Activities, if any, within BER to prepare for the low-dose radiation program

Available funds currently and plans for funding projects and with what mechanisms

- Office of Science, and the Federal Government are operating under a Continuing Resolution (CR) through February 18th
- There is the potential for a Full-Year Continuing Resolution

At this moment, SC is awaiting the final status of the FY 2022 budget

Considerations for Restarting a Low Dose Radiation Research Program

Description of funding mechanisms to support research at national laboratories and universities and internationally; project review and selection process. Please direct the committee to public links/documents when available

Should a program be restarted within BER there are several funding mechanisms that could be considered including:

- Solicitation of proposals from the Academic and DOE National Laboratory research communities.
  - International partners not excluded; International Lead PIs, not impossible, but likely difficult.
- Proposals could take the form of a combination of:
  - Individual PI efforts
  - Small teams of researchers, or;
  - Large, multidisciplinary, team-oriented (virtual) research centers
- Proposals would be peer-reviewed via the SC Peer review policies relying on researchers in the radiation biology community both in academia and at the DOE National Labs.
- There will need to be some coordination among Federal agencies involved in low dose research including those identified in the recent OSTP/NSTC report.

Update RadBIO-AI Project

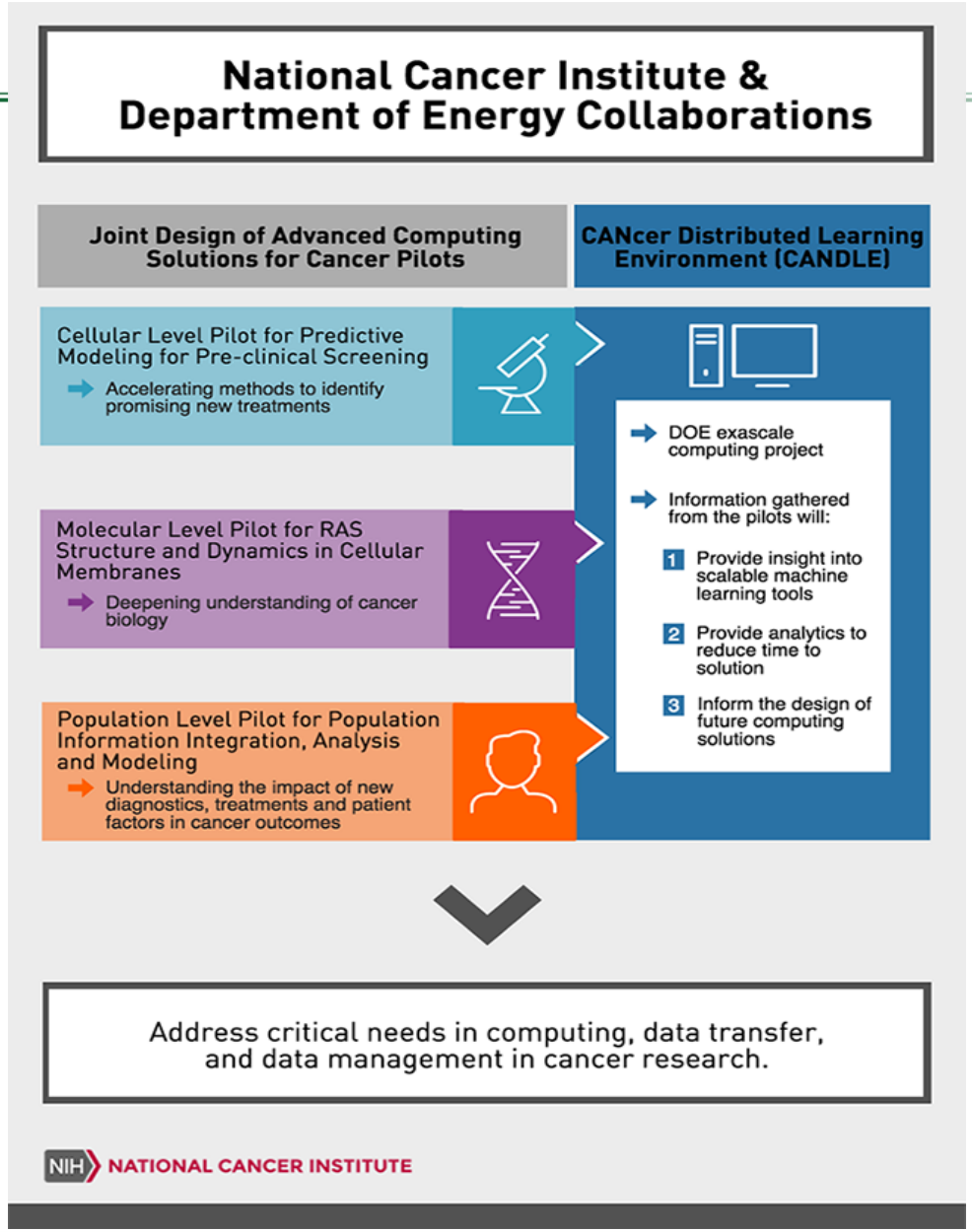
Progress with advancing low-dose research through projects currently funded by the low-dose program (e.g., RadBio-AI); plan for continuing support for these projects.

Project led by Argonne National Laboratory in collaboration with:

Brookhaven National Laboratory

Oak Ridge National Laboratory

Supplement to an existing DOE-NCI Effort under the Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) Pilots



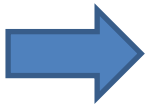
The RadBIO-AI Project is a Supplement to CANDLE

CANcer Distributive Learning Environment (CANDLE)

- An open source, collaboratively developed software platform that provides deep learning methodologies for accelerating cancer research

  - Identification of key molecular interactions, based on molecular dynamic simulations of proteins, specifically RAS.
  - Predictions of tumor response to drug treatments, based on molecular features of tumor cells and drug descriptors.
  - Better characterization of cancer patient trajectories and outcomes using a growing compendium of clinical information.

Using AI and ML concepts to as tools link molecular scale interactions with cancer drug treatments/responses to patient outcomes



Extending this effort to the low dose radiation research

One of DOE's Exascale Computing Projects

Large Language Models (LLMs) – Training AI Models

The RadBIO-AI project is adapting LLMs to low dose radiation research

- Using GTP-J an open-source version of OpenAI's GTP-3 (Generative Pre-Trained Transformer-3) code
- One of the largest transformer codes in use with over 6 billion parameters

What does that mean?

- The code is trained to recognize English language
  - Access to many Academic literature datasets including PubMed Central, ArXiv, USPTO, Freelaw, others
  - Access to internet resources OpenWebTech, Wikipedia, Pile-CC, others
  - Access to literary sources, books etc.
  - Can be adapted as a scientific tool with sufficient training

The RadBIO-AI project is building the training datasets and apply them to low dose research

Update RadBIO-AI Project – Campaign 1

Campaign 1 is aimed at the questions related to discovering signatures of radiation damage, characterizing these signatures, and producing a set of predictive models that can be used to detect these signatures in a sample.

Gene Expression and signatures

- Developed new methods for encoding datasets from the GEO and GDC repository that contain low-dose experiments from many different gene expression platforms
- Enables LLMs to train on data that is generalizable across text, gene expression and genomics data
- Tentative results successfully predict exposure and dose level with integrated expression data from multiple platforms
- This encoding process is a key step for integrating data from many disparate sources into a common body of training data
- An additional 400 datasets are in the process of encoding to enable joint training with biomedical literature

- **NCBI GEO**
- **Genomic Data Commons (GDC)**
- **Literature (Campaign 3)**

Jain V, Das B. **Global transcriptome profile reveals abundance of DNA damage response and repair genes in individuals from high level natural radiation areas of Kerala coast.** PLoS One. 2017 Nov 21;12(11):e0187274. doi: 10.1371/journal.pone.0187274. PMID: 29161272; PMCID: PMC5697823.

Campaign 1 Continued

Integrating population-level epidemiological datasets with environmental exposure data in the context of the low-dose scientific literature

- Selected exemplar datasets from CEDR and RERF and have built models to explore feature importance (fields in the data that contribute most to model performance)
- Using this feature importance map to suggest input representation needed to jointly train with molecular data and background textual data
- Exploring use of learning prompts to drive generative LLMs to answer queries that relate molecular, population level data and background knowledge
- Demonstrated a based GPT-J model able to answer queries that relate to gene expression changes due to low-dose radiation.
- Validation of GPT-J model results is ongoing

- **Comprehensive Epidemiologic Data Resource (CEDR)**
- **Radiation Effects Research Foundation (RERF)**
- **Literature (Campaign 3)**

Extracting molecular-level signatures of low-dose radiation

- Successfully demonstrated iterative prompting in GPT-J to relate background knowledge of radiation biology to cancer
- Experimenting with alternative representations of mutations/variations with those used in DNAbert to be used for tuning GPT models on mutation datasets
- Constructing test datasets and systematic queries for evaluating training strategies on low-dose related biomedical literature
- LLMs are evaluated differently than classical ML models and need careful validation of responses to queries

- **UK Biobank**
- **RERF public datasets**
- **Literature (Campaign 3)**

Davidson, P.R., Sherborne, A.L., Taylor, B. *et al.* **A pooled mutational analysis identifies ionizing radiation-associated mutational signatures conserved between mouse and human malignancies.** *Sci Rep* **7,** 7645 (2017). https://doi.org/10.1038/s41598-017-07888-0

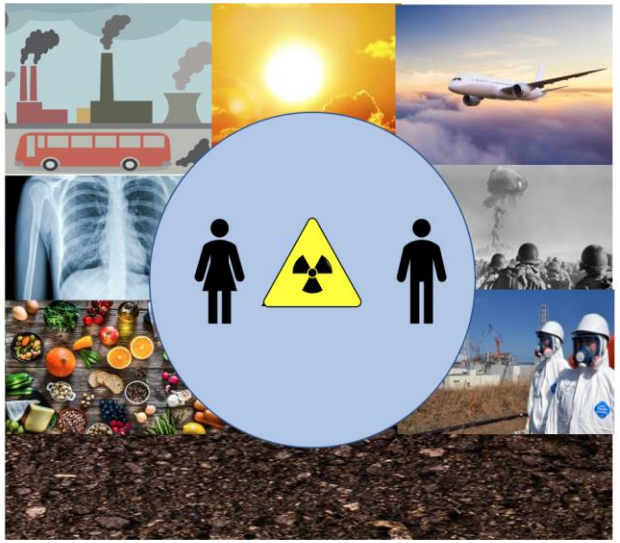
Update RadBIO-AI Project – Campaign 2

Campaign 2 is aimed at building a framework for capturing longitudinal radiation exposure profiles and using these profiles to estimate cancer risk.

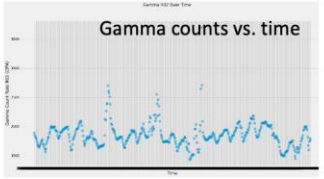
- Builds on elements of Campaign 1
- Capture and model longitudinal environmental exposure.
- Estimate how environmental factors change the risk of cancer throughout an individual's lifetime?
- Extract terms from medical records that indicate a patient was exposed to radiation?

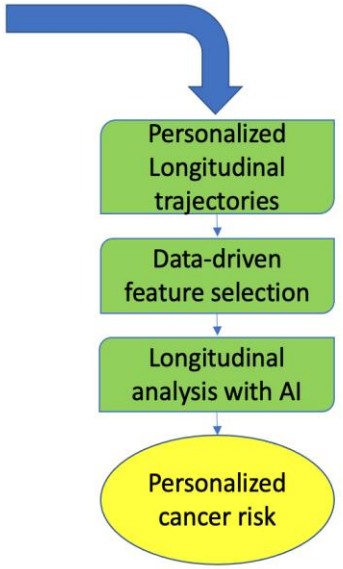
Dataset
Radnet (background gamma counts)
Radon zone (EPA)
Nuclear power plant
I-131 exposure (NCI)
Per capita air travel
Air quality
Cancer Incidence

Source of human radiation exposure



Changes with time and location





Additional Component: *Modeling Biological Signatures at Sub-cellular Scales*

Simulate effects of radiation exposure to sub-cellular organelles.

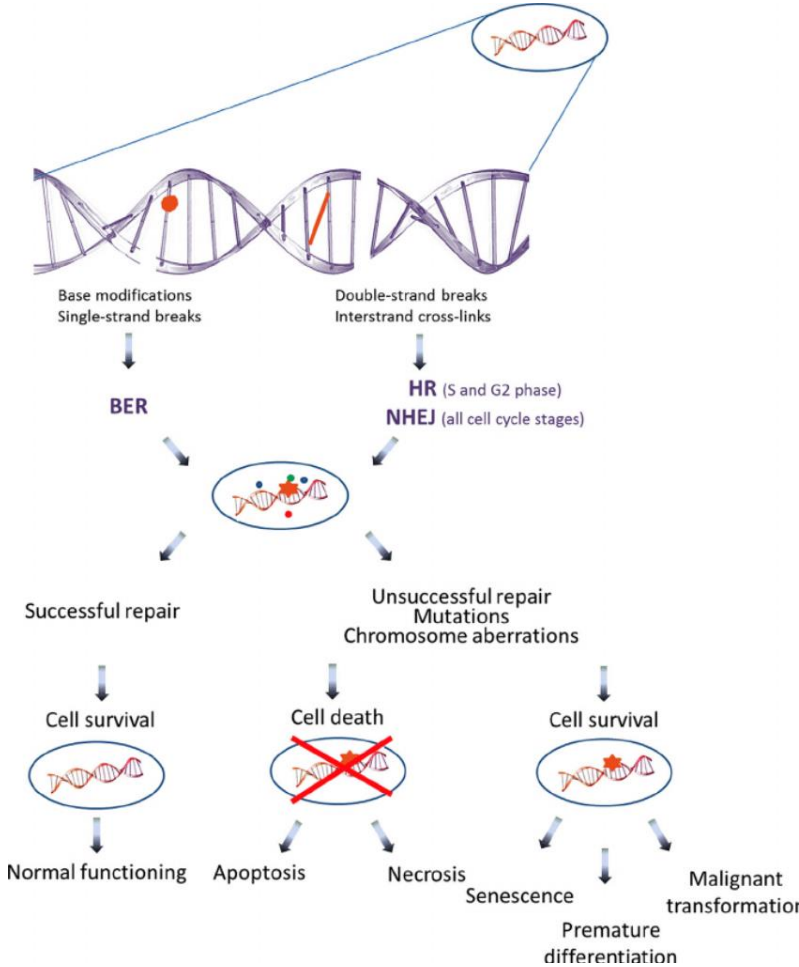
Simulation kit: Topas-nbio (extension of TOol for PArticle Simulations)

- TOPAS is a toolkit for the simulation of particle interactions in matter
- TOPAS-nBio is an extension to TOPAS for radiobiology applications
- Can simulate the physical, chemical, and biological interactions resulting from irradiation of cells and cell nuclei

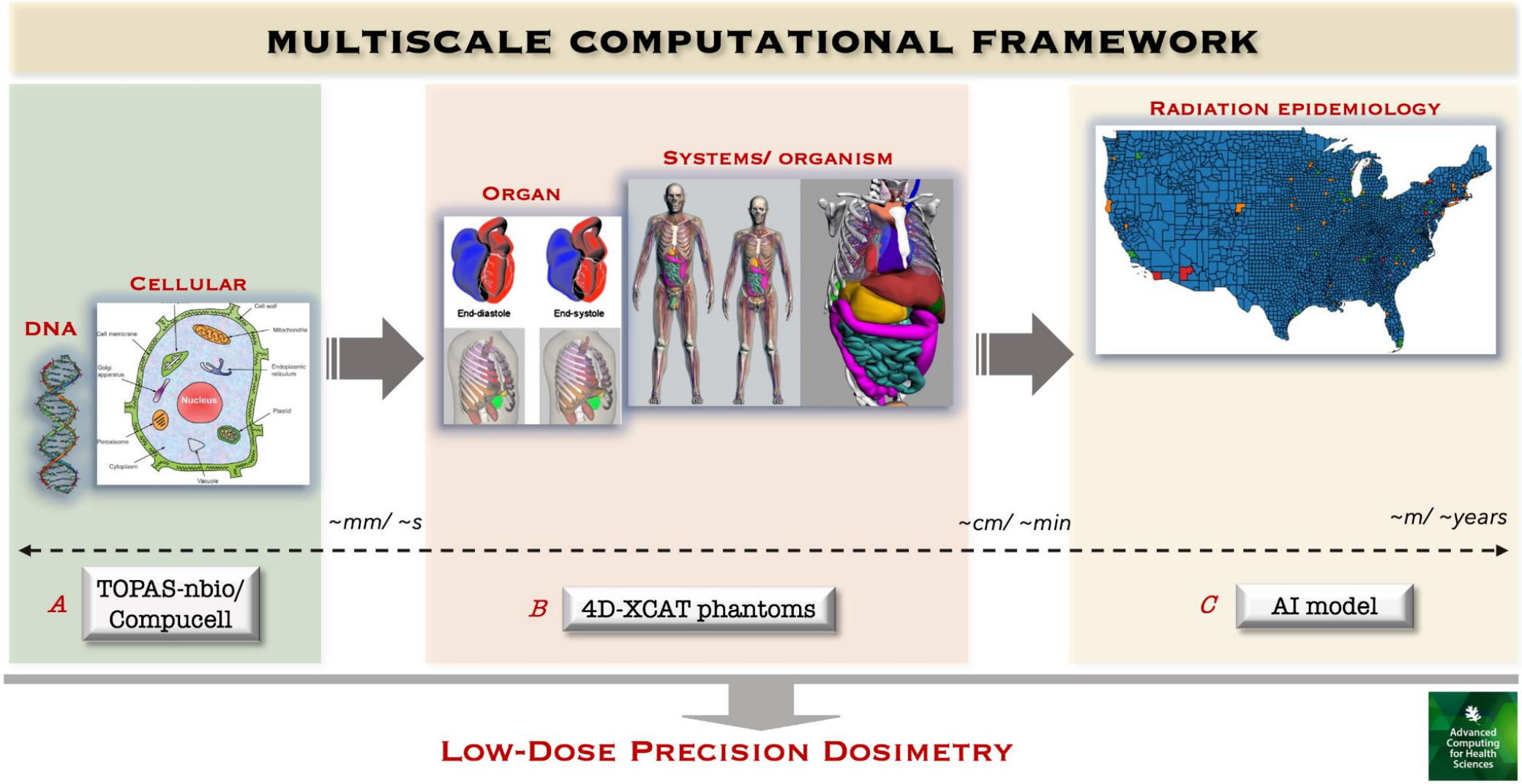
Provides a modeling component to help link exposure data to physical, chemical, biological processes within the sub-cellular environment.

Allows the computational exploration of radiation effects within cells through the generation of digital twins.

Complementary to and can be informed by single cell experimental analyses.



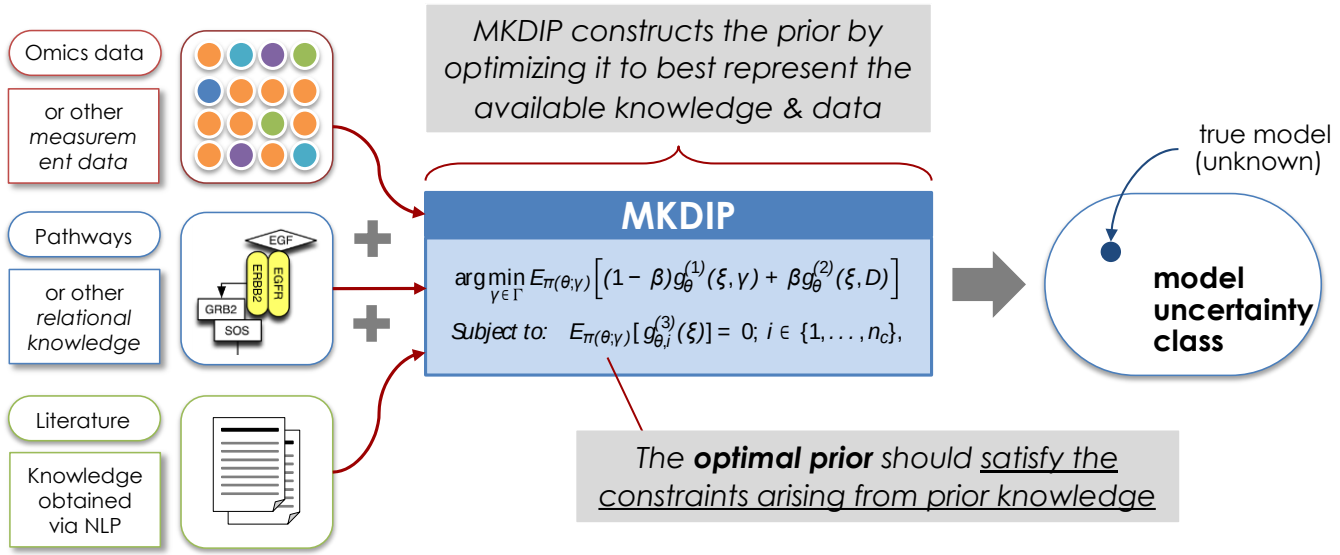
Risk Assessment across Scales



Update RadBIO-AI Project – Campaign 3

Campaign 3 is aimed at extracting associations from the scientific and biomedical literature that can be used to advance our understanding of low-dose radiation biology, create novel hypotheses for computational testing, and to create Bayesian priors that may improve the predictive capacity of our machine learning methods.

- Use machine learning/AI to extract from the biomedical literature the identification of radiation induced cellular pathway perturbations?
- Use the biomedical literature to generate priors that improve our machine learning methods?



Use Natural Language Processing (NLP) to search the literature, identify datasets, add the relevant datasets to the existing knowledgebase and improve predictions. Links back and informs Campaign 1.

Comments on NASEM Presentations

- Excellent presentations
- Good background on which to construct a strategic plan for low dose research.

A few things that caught my eye……

8/20/21 Excerpts from presentations

Brenner - Focus on the "Big Picture"

- Test the underlying assumptions of the LNT.
- Evaluate variances in radiation sensitivity.
- Find "unique fingerprints" of radiation exposure

Cuccinotta - Information on the interagency connections with NASA, Japan and EU programs

Fornace – myriad signaling pathways affected by LDR, myriad impacts on metabolism

Yuan - Acute LDR-induced adaptive response is protective

Chronic LDR-induced adaptive response may be detrimental due to prolonged p53 downregulation

Comments on NASEM Presentations Cont'd

9/24/21 Excerpts from presentations

Ansari - Need for a goal-oriented program

Berrington de Gonzalez – several low dose epidemiology projects in progress

The REB priority questions align with LDR research goals.

Costes – NASA GeneLab – a resource for low dose data

- visualization of transcriptomic data

11/16/21 Excerpts from presentations

Bernstein - WECARE

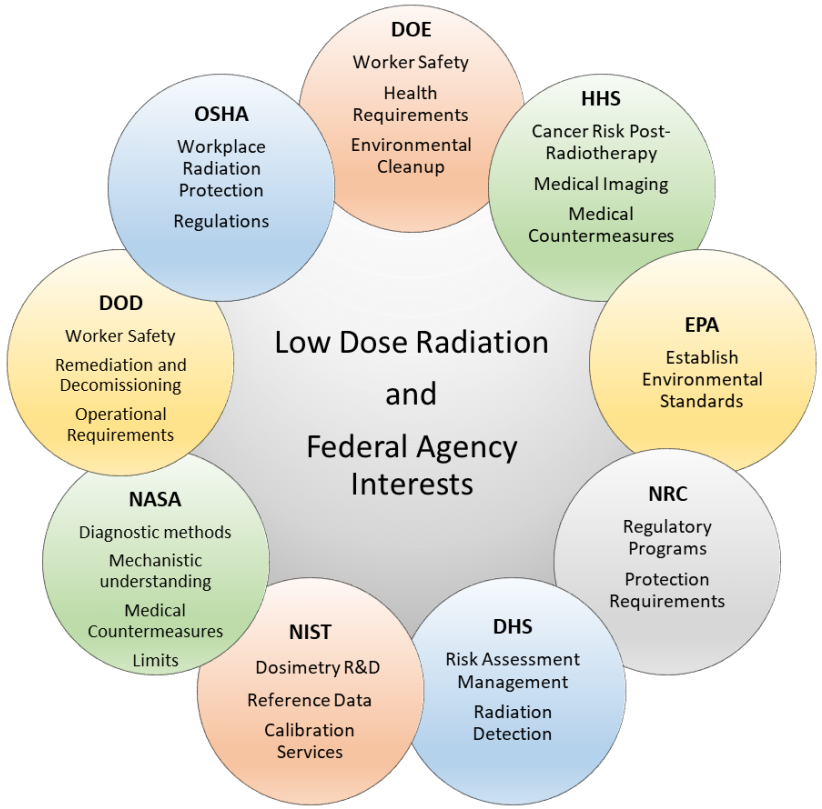
- Coupling epidemiology, genetics/genomics and sample preservation
- "Universal Digital Cohort and Biobank"

Bolch – dosimetry modeling

OSTP/NSTC Report

OSTP/NSTC Report recommended:

- Coordinated research among Federal Agencies Engaged in Low Dose Research
- CIRRPC-style Low Dose Radiation Research Steering Committee (LDRRSC)



A major concern of the Sub-Working Group was how a CIRRPC-style steering would actually operate:

- How would such a group function without independent funding?
- What authority would such as group have?
- What level of coordination is effective?
  - Management-level versus researcher level?

Additional Concerns

Some Considerations to standing up a new or renewed program in Low Dose Radiation Research

- Clearly defined goals on a 5-yr and 10-yr horizon.
- Technology has changed dramatically in the past 5-8 years. What new approaches to low dose radiation research should be pursued that take advantage new technologies including, for example:
  - High throughput omics analyses and single cell gene expression capabilities
  - Imaging technologies such as cryo-EM.
  - Advances in computational structural biology
  - Computational techniques, including big data approaches
  - Reanalyzing/sequencing tissue collections – is this possible/valuable?
- Thoughts on how to coordinate with other agency researchers.
  - The Committee's perspective on what are the best mechanisms for doing so?

Thank you

https://science.osti.gov/ber

https://www.energy.gov/science/ber/biological-and-environmental-research