Recommended Cancer Informatics Investment Areas

In addition to the priorities that were recommended to ICI to act upon immediately, several other efforts or areas to focus for impact were suggested by the Summit participants. This list is provided below as a resource to the Summit attendees and the ICI Community. This list is for information only and does not imply immediate support or interest by the ICI Fund, however, some items listed may be of interest in the future.

Utility of decision support:
Test utility of clinical decision support methods (example in https://academic.oup.com/jamia/article/25/5/458/4791826)

Empower patients:
Empower patients to share their genomic and clinical data through novel informatics techniques. See the Count Me In (https://joincountmein.org) and Make an Impact (https://www.mskcc.org/research-programs/molecular-oncology/make-impact) patient initiatives.

Computational and algorithmic methods:
Continue funding computational methods development--few venues exist to support such projects.

Response profiles:
Method development to leverage heterogeneous omics, drug response and clinical data, if available, for deciphering regulatory pathways to study: (i) inter-patient population subtypes and intra-tumor heterogeneity, and (ii) mechanisms of drug resistance and response in cancer.

Single cell profiles:
Support method development for single cell data analysis and single-cell modeling (including imaging data, e.g., multiplexed IHC). Such studies may have a direct impact on clinical trials for selection of combination therapies including immunotherapy.

Match data:
Support collaborative efforts to match different data modalities for analysis/modeling algorithms to address relevant translational questions.

WGS:
Decision support for clinical cancer whole genome sequencing, including the visualization and interpretation of complex structural variation.

Genome 3D:
Characterize the chromatin state and 3D folding of rearranged cancer genomes.

Images & transcripts:
Leveraging digital pathology AI approaches to define H&E correlates of spatial / single-cell transcriptomics-defined cell types to understand tumor heterogeneity, tumor microenvironment, and tumor-stromal interactions in tumor specimens.

Clinical data:
Work towards large-scale data deposition into shared infrastructure using clinically generated data across both academic and community institutions, addressing both technical and data sharing policy issues and exploring federated searches.

Non-invasive tracking:
Non-invasive assays and computational approaches for cancer risk assessment, early detection, monitoring, surveillance, and longitudinal analysis. How do we translate this to the clinic?

Continued on next page

Microbiome:
Microbiome data as part of new data infrastructure for cancer genomics data (2.0?)

Trial metadata:
Organizing (ongoing, enrolling, past) "precision" or "targeted" clinical trial metadata into a queryable form on the basis of drug target and genotypic / molecular entry.

Negative loci:
Identifying relevant tumor subtype-specific "pertinent negative" loci [information from the absence of mutations] for a given patient using reference (e.g. TCGA, ICGC, Hartwig) data.

Rich data in few samples:
Touching on both new data modalities (e.g., metabolomics) and new machine learning methodologies: analysis of heterogeneous, "medium-size" datasets with potentially many feature-rich data modalities over comparatively few samples.

Decision psychology:
Interface between clinical informatics and psychology. How do you get doctors to make the right decision informed by detailed data? How should information be presented?

Remove decision bias:
When trying to inform medical decisions using observational data, the relative size of treatment effect and bias is critical. How can we estimate and account for these biases properly so we don't end up advising docs to recapitulate biases in observational data but also as to when an RCT (randomized controlled trial) is essential (size of bias and effect are on same scale) and where RCT may be less essential (size of bias is tiny by comparison to effect size).

ML/AI platform:
Platform for multi-modal, multi-scale decision support: reproducible, robust, ability to share and validate ML models, visualization of ML models, interpretation of ML & AI.

ML data integration:
Development of ML methods for intermediate data fusion of multi-modal data e.g. clinical, genomic, imaging, for clinical decision support.

Better decision support:
Clinical decision support + Machine learning: why have some of the previous efforts failed? Can we do better?

Molecular & clinical data:
Connecting molecular data / analysis to patient data and making clinical statements. (1) Developing methods that do or support this (e.g. CybersortX for deconvoluting bulk data, a method called Comet for identifying markers for cell populations in RNA-seq, etc) and (2) establishing pipelines to apply methods to "zoom out" from the molecular data to utilizing larger sample sizes that can compare to the clinical setting.

Spatial molecular:
Methods for analyzing and interpreting spatial molecular assays, and for making them clinically practical.

Drug synthetic lethality:
Approaches for better synthetic lethality prediction and linking those predictions to drug databases, including finding "failed drugs" for use in combinations.