**MicroLab Breakout Group Notes**

**Combined Summaries of Action Ideas and Additional Information**

1. **Summary of Action Ideas**

**FULL GROUP CHAT**

* **Continue to discuss using dynamical models (particularly ODEs) to fit to observations**, etc., then running many scenarios on them. It's something we've been doing with agent-based models on HPC. But to be truly powerful, need to combine these with ML to help bridge some of the key remaining gaps.
* **Circulate to this interest group some ways that we envision the data need**s and what could be used for development of digital twin in a clinical sense, and then see what others in the group envision, and home in on a set of working questions.
* **Work on Phenotype Prediction Within This Group (offers tremendous scope)**
	+ Develop studies that would project treatment effects on a stream of data based on considerations like PKPD, tumor localization, target protein interactions with small molecules, etc.
* **Counterfactual Digital Twin**

Could offer “counterfactual digital twin” as a more precise distinction from other uses, that meets goal of precision medicine: given an individual’s data to date, which (set of) treatment(s) are “right” for that individual at that time? whether ML, or ODE dynamical models, or other modeling approaches are used.

* **Find a focused clinical target area is key to homing in on a tractable problem**. To engage clinicians effectively, the selection of use case is important because each cancer site requires specialty treatment and has different clinical questions.
	+ Choose a clinical target area that already includes some “success stories” retrospectively, such as melanoma or non-small-cell lung cancer (immunotherapy innovations) and/or glioblastoma.
* **Explore computational experimentation with digital twins.** Leverage expertise (e.g., [CISNET](https://cisnet.cancer.gov/)) to address the vast amount of work required (e.g., large-scale uncertainty quantification) to develop the calibrated digital twins.
	+ Determine what types of models would be considered digital twins, e.g., mechanistic models, microsimulation, and potentially natural history models

* **Createan email list (listserv)** to circulate ideas through the group.

**DIGITAL TWIN**

* **Concrete definition needed for “twin,” i.e. what might be a possible quantitative metric to define similarity**
* Primary goal should be **eventual digital testing on the twin** for genuine treatment decision enabling.
* **Oncology physicians and researchers** – pull from the NCI community – want to engage them from the beginning to get a physician’s voice.
	+ Reach out to community meetings such as [ASCO](https://www.asco.org/) (great place to engage clinicians and get them involved)
* **Potential funding resource**: Philanthropic groups – [Breast Cancer Research Foundation](https://www.bcrf.org/)
* **Industry could build a data commons or sandbox and collect data across universities and begin modeling process**
* **Defining a tractable problem would give those interested in participating a way of having something concrete to explore**, and a way to get some footing in some actual work.
	+ A tractable problem that could be modularly expanded would be helpful. Current efforts include building tumor metastasis dissemination models, based on longitudinal mouse models. Would be interesting to improve and scale that to human scales, with joint expertise, then start plugging in better details.
* **Envision a stream of data where there would be interest in projecting treatment effects based on considerations like PKPD, tumor localization, target protein interactions with small molecules, etc**. There is tremendous scope in this group for really working on phenotype prediction.
	+ **Propose circulating to this interest group** some ways that we envision the data needs and what could be used for development of digital twin in a clinical sense, and then see what others in the group envision, and home in on a set of working questions.
		- Selection of use case is important because each cancer site requires specialty treatment and has different clinical questions.
	+ **Propose picking a clinical target area that already includes some “success stories” retrospectively,** such as melanoma or non-small-cell lung cancer (immunotherapy innovations) and/or glioblastoma?

**SYNTHETIC DATA**

* **Utilize pre-existing synthetic data systems like** [**MDClone**](https://www.mdclone.com/) **to deliver synthetic datasets**.
* **Assemble an email list for future scoping exercises and actions.**
* **Short-term Goal**: Provide data sets to students/collaborators for use
* **Long-term Goal**: Determine a viable research area for synthetic data development in oncology research.
* **Develop methods that will consume a corpus of real data of various cancer domains (pathology, radiation etc.) that can generate synthetic data.** These methods can then be distributed to various institutions.

**MACHINE LEARNING FOR HYPOTHESIS GENERATION**

* **Need well-defined use cases for machine learning for hypotheses generation; use cases to map the right data sets and address the challenge**.
* **Need ML people working with clinicians to better understand the questions**.
* From an engineering perspective, **propose using a tool approach: run different statistical models against data and literature so models could be populated and updated as code is written or updated**. Hypotheses generation could be accelerated when given information about underlying data.
* **Need to do the following outside of clinical trials:**
	+ **Study cancer treatment response**. Need to study cancer response across ALL patients. Inclusion and exclusion criteria are not applicable.
	+ **Mine and harness electronic medical record (EMR) data**
	+ **Develop capability to validate the models**, for example, a correlative validation.
* **Targeted Therapy**: once the genomics is understood, develop a prediction, then target therapy for the prediction.
* **Precompetitive curation/labeling/annotation and data labeling for reuse is needed**. Need to facilitate labeling around annotation or curation; seed points can be generalized across applications. Important to make the data more statistically revant and match the right metadata sets for ML applications.

**ADAPTIVE TREATMENTS**

* **Find a suitable dataset at high resolution for a minimal model** including only mutations whose significance is known (which is less than 1/10 of 1%).
	+ Identify funding for such datasets and get a data set against which to validate mathematical models.
* **Implant small sensors into tumors** to better understand what is going on in the micro environment to inform adaptive treatments.
* **Develop data sets that capture tumor changes, and work on bolstering a good training set.**
	+ Look at genetic signatures over time, observing the genetic evolution to get a phenomic description of tumors over time.
	+ Come up with model on that level and look for patterns of genetic evolution to find universal characteristics.
* **Collecting data from true biological systems**
	+ A preclinical model of 2-d or 3-D cell culture of biopsy derived cells may be a reasonable place to start to investigate the effect of different treatment sequences.
	+ Microfluidics could be used to apply different treatment sequences to cell culture.
* **Using Patient Derived Xenografts (PDX) rather than cell lines**
	+ Look at how a curable model is different from a treatment resistant model. Sequence patient tumor, create landscape, use on mouse and cure mouse. Could then look at cases in which tumor is cured and one which there is therapeutic resistance.
* **Instrument a mouse with high tech sensors/nanowires to sense pH changes, etc.**
	+ Have a cured model and resisted/Ras-mutated model from which to collect data. Model could be recapitulated, reproduced, done in replicates providing statistical info of robustness in findings.
	+ Barcoding technique could be applied into cell lines then transfected into mouse models.
* **Could use two different technologies – PDX would give more realism, and cell-line dataset would provide more sub-clone resolution**.
	+ Would need more varied data sources, and to aggregate existing data lakes
* **With the PDX approach could use the resulting data to feed Machine Learning models**. Could also model patient tumors with 3D culture systems plated out on wide array of possible combinations and compare them to each other and to mouse models.
1. **Additional Resources from the Participants**

**FULL GROUP CHAT**

* MDClone is a great example of service-oriented synthetic data in development <https://www.mdclone.com/>
* ASCO’S CancerLinQ <https://cancerlinq.org/>
* Cancer Intervention and Surveillance Modeling Network (CISNET) <https://cisnet.cancer.gov/>
* Melanoma Research Foundation (MRF) Breakthrough Consortium (Accelerating Research and Treatment Development <https://melanoma.org/research-science/scientific-initiatives/mrf-breakthrough-consortium-mrfbc/>
* Potential Funding:
	+ <https://grants.nih.gov/grants/guide/pa-files/par-16-349.html>
	+ NCI Informatics Technology for Cancer Research (ITCR) <https://itcr.cancer.gov/funding-opportunities>
* Society for Simulation in Health Care (SSIH)- Healthcare Systems Modeling & Simulation Affinity Group

<https://www.ssih.org/Interest-Groups/Healthcare-Systems-Modeling-Simulation>

SSIH webinars online to educate our clinicians <https://www.youtube.com/channel/UCMWVW9plawga7UtWnISrszg/videos>

**DIGITAL TWIN**

* Roswell Park, experience in phenotypic projection <https://www.roswellpark.org/>
* American Society of Clinical Oncology <https://www.asco.org/>
* NIH Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) Initiative <https://datascience.nih.gov/strides>
* Melanoma Research Foundation – Accelerating Research and Treatment Development
	+ <https://melanoma.org/research-science/scientific-initiatives/mrf-breakthrough-consortium-mrfbc/>
* **Potential funding resources**:
	+ <https://grants.nih.gov/grants/guide/pa-files/par-16-349.html>
	+ NCI Informatics Technology for Cancer Research (ITCR) <https://itcr.cancer.gov/funding-opportunities>
* Society for Simulation in Healthcare - Healthcare Systems Modeling & Simulation Affinity Group
	+ <https://www.ssih.org/Interest-Groups/Healthcare-Systems-Modeling-Simulation>
* Healthcare Systems Modeling and Simulation Webinar <https://www.youtube.com/channel/UCMWVW9plawga7UtWnISrszg/videos>
* Cancer Intervention and Surveillance Modeling Network <https://cisnet.cancer.gov/>
* Breast Cancer Research Foundation <https://www.bcrf.org/>
* **Interagency Modeling and Analysis Group (IMAG)** <https://www.imagwiki.nibib.nih.gov/>
* **Presentation on Industrial Digital Twins**
	+ <https://www.imagwiki.nibib.nih.gov/sites/default/files/GE%20Digital%20Twin%20Overview%20and%20Tutorial_RRI%20v2.pdf>

**SYNTHETIC DATA**

* **Highlights applications of synthetic data** to DoD, DHS, and other national/international efforts (2 slides) <http://people.virginia.edu/~ss7rs/synthetic_population_tutorial_2/slides.php>
* This **online, open provincial EHR platform** allows e-health solutions to be tested in a virtual EHR environment, giving innovators a space to prototype new ideas. <https://www.innovation-lab.ca/>

**MACHINE LEARNING FOR HYPOTHESIS GENERATION**

* “No Free Lunch Theorems for Optimization.” IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, VOL. 1, NO. 1, APRIL 1997 <https://ti.arc.nasa.gov/m/profile/dhw/papers/78.pdf>

* “Cancer subtype identification using somatic mutation data,” British Journal of Cancer 118, 1492–1501 (2018) <https://www.nature.com/articles/s41416-018-0109-7>
* NCI Informatics Technology for Cancer Research <https://itcr.cancer.gov/funding-opportunities>
* **NCI Advisory Board Ad Hoc Working Group on Data Science and Machine Learning Report**

<https://deainfo.nci.nih.gov/advisory/ncab/workgroup/DataScienceWG/WGJune2019recommendations.pdf>

* **Chan Zuckerberg Foundation, potential funding avenue** <https://chanzuckerberg.com/rfa/essential-open-source-software-for-science/>
* The Cancer Imaging Archive <https://www.cancerimagingarchive.net/>

**ADAPTIVE TREATMENTS**

* Beckman RA, Yeang CH. **Nonstandard Personalized Medicine Strategies for Cancer May Lead to Improved Patient Outcomes**. Per Med. 2014 Sep;11(7):705-719. doi:10.2217/pme.14.57. PubMed PMID: 29764056.<https://www.ncbi.nlm.nih.gov/pubmed/?term=beckman+pnas+2012>
* Beckman RA, Schemmann GS, Yeang CH. **Impact of genetic dynamics and single-cell heterogeneity on development of nonstandard personalized medicine strategies for cancer.** Proc Natl Acad Sci U S A. 2012 Sep 4;109(36):14586-91. doi:10.1073/pnas.1203559109. Epub 2012 Aug 13. PubMed PMID: 22891318; PubMed Central PMCID: PMC3437850. <https://www.pnas.org/content/109/36/14586>
* Yeang CH, Beckman RA. **Long range personalized cancer treatment strategies**

**incorporating evolutionary dynamics**. Biol Direct. 2016 Oct 22;11(1):56. PubMed PMID: 27770811; PubMed Central PMCID: PMC5075220. <https://www.ncbi.nlm.nih.gov/pubmed/27770811>

###