Envisioning Computational Innovations for Cancer Challenges (ECICC): Scoping Meeting March 6 – March 8, 2019

The ECICC meeting included generating ideas for challenges. The process required participants to use post-it notes to document the ideas. Throughout the meeting, the post-its were grouped according to teams/areas to address.

This document incorporates the post-it notes that participants created throughout the ECICC meeting.

Audio Files for the meeting were created and IM'd as follows:

LLNL-MI-770120: Audio through Thursday, March 7, 1:30p.m.

LLNL-MI-770119: Final Audio recording Thursday, March 7, 3:30p.m. – 6:00p.m.

Prepared by LLNL under Contract DE-AC52-07NA27344. LLNL-MI-770618

AGENDA

Envisioning Computational Innovations for Cancer Challenges: Scoping Meeting

March 6-7, 2019

Livermore Valley Open Campus

High Performance Computing Innovation Center, Building 6475

7000 East Ave - Livermore, CA 94550

March 6

8:00 am – 8:45 am **ARRIVAL AND CHECK-IN**

8:45 am – 9:45 am **WELCOME AND ORIENTATION**

Why Now?
Emily Greenspan, *NCI*
Carolyn Lauzon, *DOE SC*

HPC in Cancer Research Brief Overview
Amy Gryshuk, *LLNL*
Eric Stahlberg, *FNLCR*

Knowinnovation Introduction
Andy Burnett, *Knowinnovation (KI)*
Stavros Michailidis, *Knowinnovation (KI)*

9:45 am – 10:45 am **SESSION 1: WHAT'S POSSIBLE COLLABORATING ACROSS DISCIPLINES AND ORGANIZATIONS**

JDACS4C Panel:
Jessica Boten, *NCI*
Yvonne Evrard, *FNLCR*
Dwight Nissley, *FNLCR*
Rick Stevens, *ANL*
Fred Streitz, *LLNL*
Gina Tourassi, *ORNL*

Moderators:
Michael Cooke, *DOE SC*
Betsy Hsu, *NCI*

10:45 am – 11:30 am **A Coffee Break with a Purpose** - Generating ideas for lean-in challenge areas (part 1)

11:30 am – 12:30 pm **KEYNOTE PRESENTATION**

DOE Success Story: Leading with Science in a Computational Context

Peter Nugent, *LBL*

12:30 pm – 2:00 pm **Working Lunch** - Generating ideas for lean-in challenge areas (part 2)

2:00 pm – 3:00 pm **SESSION 2: CHALLENGE AREAS IN CANCER RESEARCH**

Panel:

Gregory Cooper, *University of Pittsburgh*
Tina Hernandez-Boussard, *Stanford University*
Paul Macklin, *Indiana University*
John Quackenbush, *Harvard University*
Amanda Randles, *Duke University*
William Richards*, Brigham and Women's Hospital*
Ilya Shmulevich, *Institute for Systems Biology*
Amber Simpson, *MSKCC*

Moderators:

Amy Gryshuk, *LLNL*
Roxanne Jensen, *NCI*

3:00 pm – 3:45 pm **SPEED NETWORKING** - Generating ideas for lean-in challenge areas (part 3)

3:45 pm – 4:00 pm **BREAK**

4:00 pm – 5:00 pm **SESSION 3: DOE CAPABILITIES AND RESEARCH**

DOE National Lab Panel:

Frank Alexander, *BNL*
Silvia Crivelli, LBL
John Feddema, *SNL*
Sarah Michalak, *LANL*
Ana Paula de Oliveira Sales, *LLNL*
Robert Rallo, *PNNL*
Rick Stevens, *ANL*
Gina Tourassi, *ORNL*

Moderators:

Carolyn Lauzon, *DOE SC*
Eric Stahlberg, *FNLCR*

5:00 pm – 5:30 pm **Closing plenary | ADJOURN DAY 1**

6:30 pm – 8:00 pm **Dinner in self organized groups**

March 7

8:00 am – 8:45 am	**ARRIVAL AND CHECK-IN**
8:45 am – 9:00 am	**RECAP DAY 1 | OVERNIGHT IDEAS**
9:00 am – 10:00 am	**KEYNOTE PRESENTATION** **Blue Sky Possibilities at the Intersection of Oncology and Computing** Warren Kibbe, *Duke University*
10:00 am – 11:00 am	**BREAK OUT GROUPS** - Generating ideas for lean-in challenge areas (part 4)
11:00 am – 12:00 pm	**SYNTHESIS OF LEAN-IN CHALLENGE AREAS**
12:00 pm – 1:00 pm	**LUNCH**
1:00 pm – 1:30 pm	**PRIORITIZATION OF LEAN-IN CHALLENGE AREAS & WRITING GROUP FORMATION**
1:30 pm – 3:30 pm	**WRITING GROUPS** For each lean-in challenge area, small writing teams will address: ● An introduction to the challenge ● Why is this a relevant and important challenge? ● Why is now the right time? ● How will it drive innovation in cancer research? What is the impact for cancer research? ● How will it drive innovation in high-performance computing (HPC)? What is the impact for HPC? ● What are the key and historical challenges? ● What cultural shifts are required?
3:30 pm – 5:30 pm	**WRITING GROUP PRESENTATIONS WITH FEEDBACK**
5:30 pm – 5:45 pm	**WRAP-UP**
6:00 pm	**ADJOURN DAY 2**

Wednesday, March 6, 2018

10:45 am – 11:30 am	**A Coffee Break with a Purpose** - Generating ideas for lean-in challenge areas (part 1)
11:30 am – 12:30 pm	**KEYNOTE PRESENTATION** **DOE Success Story: Leading with Science in a Computational Context** Peter Nugent, *LBL*
12:30 pm – 2:00 pm	**Working Lunch** - Generating ideas for lean-in challenge areas (part 2)
2:00 pm – 3:00 pm	**SESSION 2: CHALLENGE AREAS IN CANCER RESEARCH** **Panel:** Gregory Cooper, *University of Pittsburgh* Tina Hernandez-Boussard, *Stanford University* Paul Macklin, *Indiana University* John Quackenbush, *Harvard University* Amanda Randles, *Duke University* William Richards, *Brigham and Women's Hospital* Ilya Shmulevich, *Institute for Systems Biology* Amber Simpson, *MSKCC* **Moderators:** Amy Gryshuk, *LLNL* Roxanne Jensen, *NCI*
3:00 pm – 3:45 pm	**SPEED NETWORKING** - Generating ideas for lean-in challenge areas (part 3)

How might cancer cells/ Tumors inspire Computing Capabilities?

Consolidate drug response data or generate new data in an annotated, standardized way.

Be able to predict patient x response from a to be defined set of patient data

Reproducibly of Comp. Model

How to maximize utility of clinical data?

Bridging the Gaps between experimentist and CS:
1) reducing duplication of efforts
2) enabling knowledge sharing.

We need to re-architect how we capture, store and access patient data so it can be easily available for research efforts.

Want a Software package that helps you design drugs or small molecules. Preferably starting with a broad, non-target specific base, but that can later be target-specific.

Science needs to be in streamline fashion.

Integration of Mechanistic + Machine learning models

Lots of data... ...How to decide?

Cancer prevention: baseline description of health and definition of cancer-related deviation

Team Science needs not just experts in different domains... ... but also people who are not experts, rather they have a good understanding in several domains. These people are "Liaison" people.

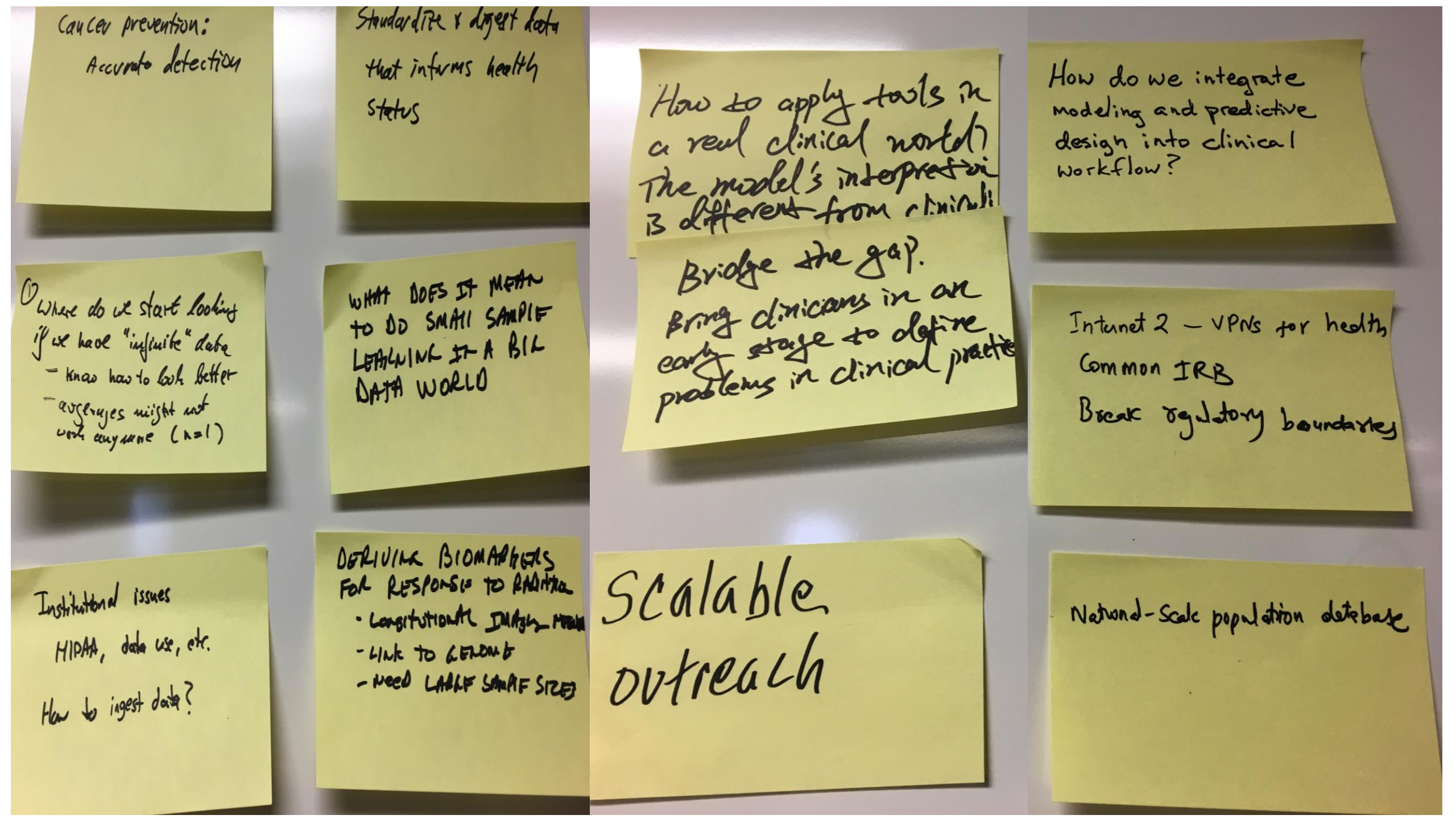
Need integration of multiscale ~~integration of~~ model, data, and expertise from cellular to tissue and population levels allowing more realistic prediction.

How to integrate different types of data

How does one provide basis of model for regulatory work.

Exposome: nutrition, environmental exposures, behaviors

a) How do we integrate models/prediction models.
⇒ Why does it work
⇒ How do we intervene



Cancer prevention:
Accurate detection

Standardize & digest data that informs health status

How to apply tools in a real clinical world? The model's interpretation is different from clinical

How do we integrate modeling and predictive design into clinical workflow?

Where do we start looking if we have "infinite" data
- know how to look better
- averages might not work anyone (n=1)

WHAT DOES IT MEAN TO DO SMALL SAMPLE LEARNING IN A BIG DATA WORLD

Bridge the gap.
Bring clinicians in an early stage to define problems in clinical practice

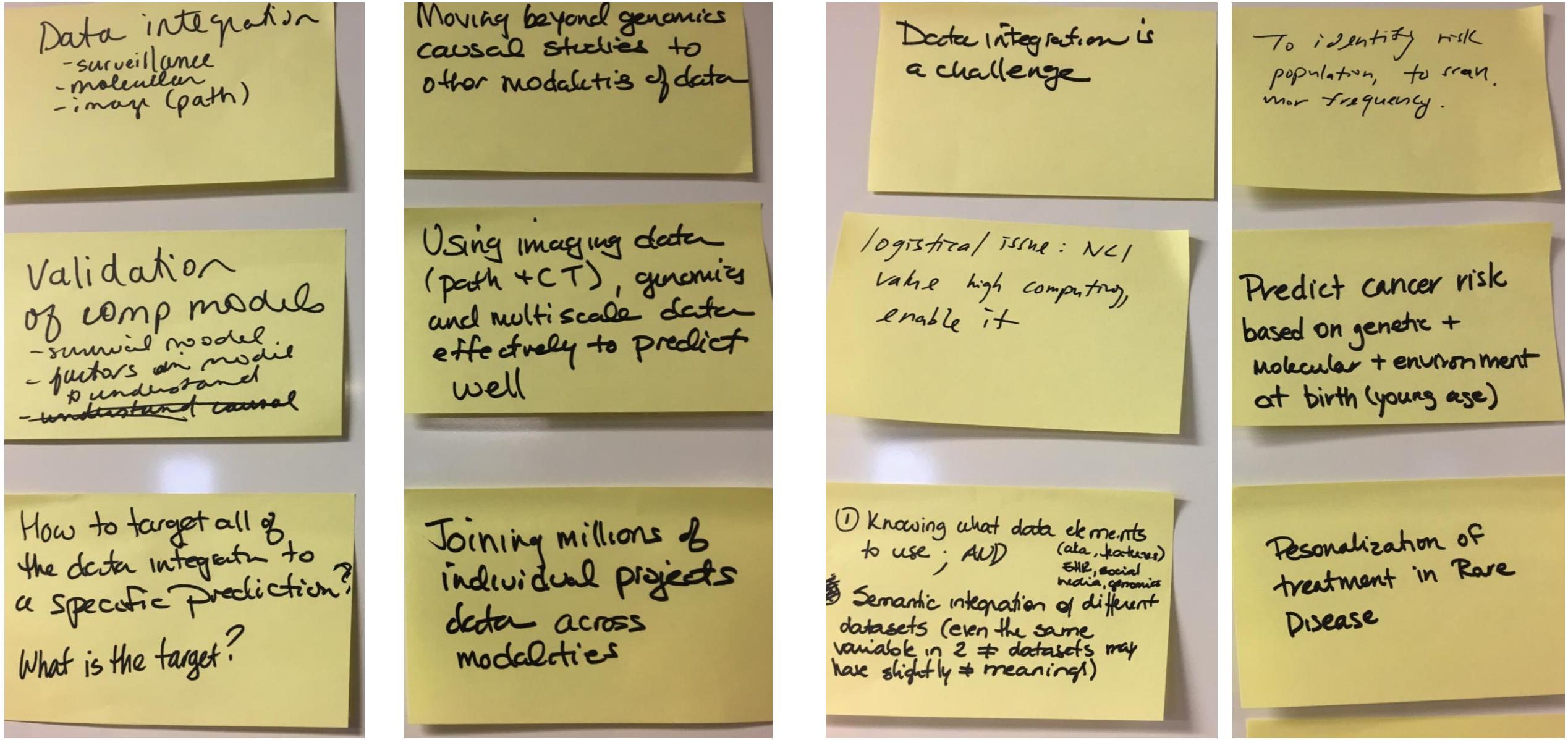
Internet 2 – VPNs for health
Common IRB
Break regulatory boundaries

Institutional issues
HIPAA, data use, etc.
How to ingest data?

DERIVING BIOMARKERS FOR RESPONSE TO RADIATION
- LONGITUDINAL IMAGING
- LINK TO GENOMICS
- NEED LARGE SAMPLE SIZES

Scalable outreach

National-Scale population database



Data integration
- surveillance
- molecular
- image (path)

Validation of comp models
- survival model
- factors in model to understand
- ~~understand causal~~

How to target all of the data integration to a specific prediction?

What is the target?

Moving beyond genomics causal studies to other modalities of data

Using imaging data (path + CT), genomics and multiscale data effectively to predict well

Joining millions of individual projects data across modalities

Data integration is a challenge

logistical issue: NCI value high computing, enable it

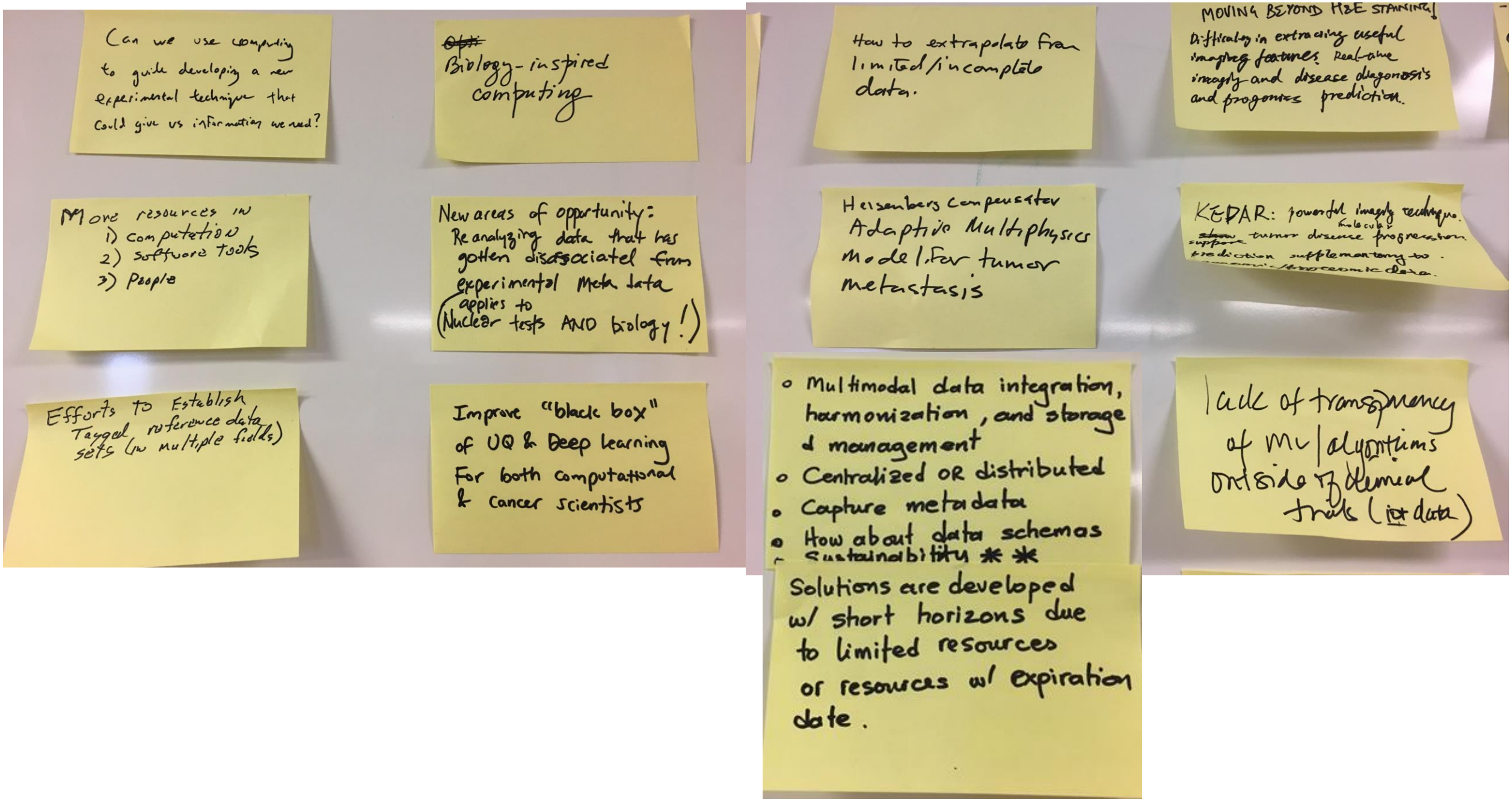
① Knowing what data elements to use; AUD (data, features) EHR, social media, genomics

② Semantic integration of different datasets (even the same variable in 2 ≠ datasets may have slightly ≠ meanings)

To identify risk population, to scan more frequency.

Predict cancer risk based on genetic + molecular + environment at birth (young age)

Personalization of treatment in Rare Disease



Can we use computing to guide developing a new experimental technique that could give us information we need?

~~Opti~~ Biology-inspired computing

More resources in
1) computation
2) software tools
3) People

New areas of opportunity:
Re analyzing data that has gotten disassociated from experimental Meta data
(applies to Nuclear tests AND biology!)

Efforts to Establish Tagged reference data sets (in multiple fields)

Improve "black box" of UQ & Deep learning For both computational & cancer scientists

How to extrapolate from limited/incomplete data.

MOVING BEYOND H&E STAINING
Difficulty in extracting useful imaging features. Real-time imaging and disease diagnosis's and progonosis prediction.

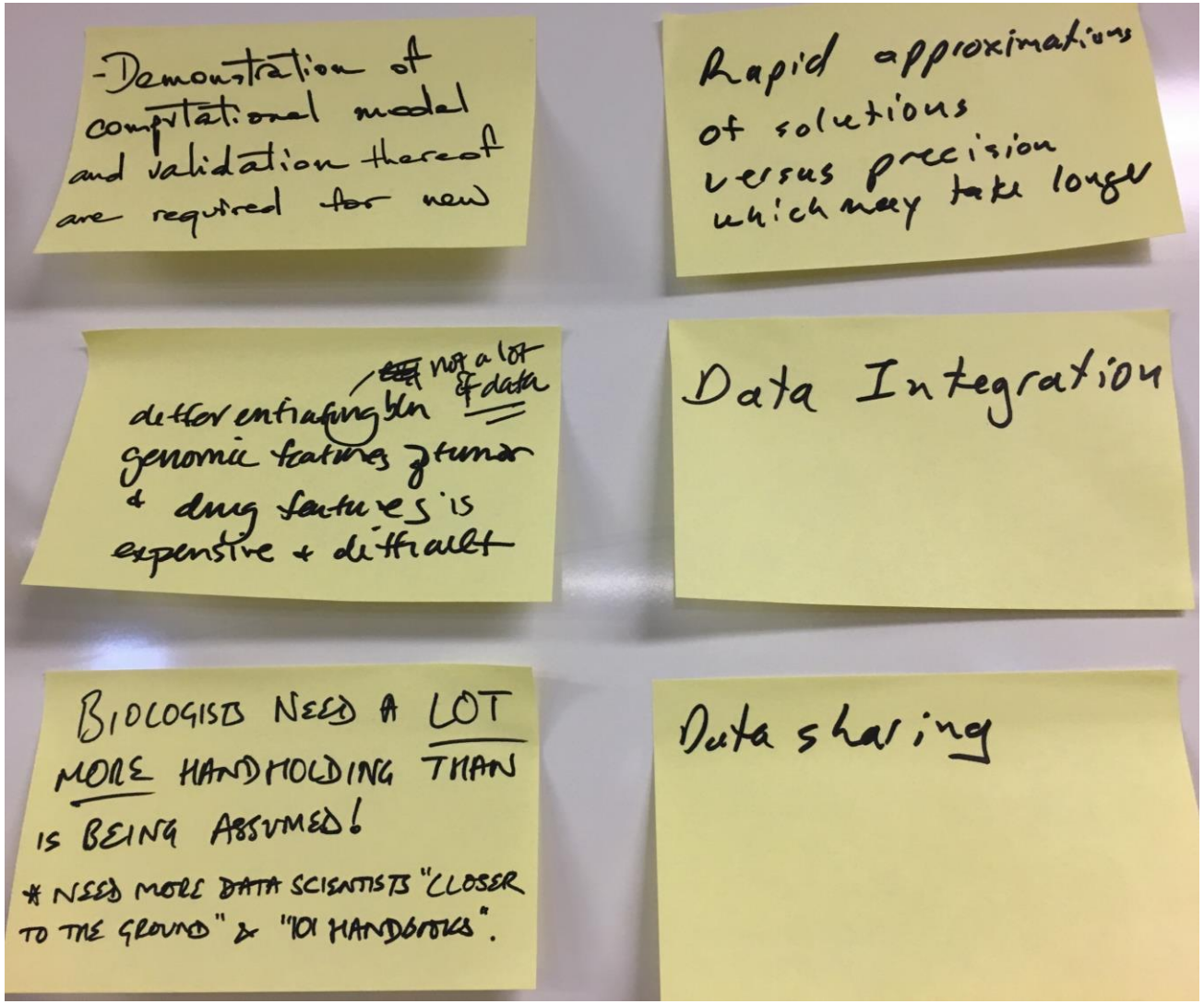
Heisenberg compensator
Adaptive Multiphysics Model for tumor metastasis

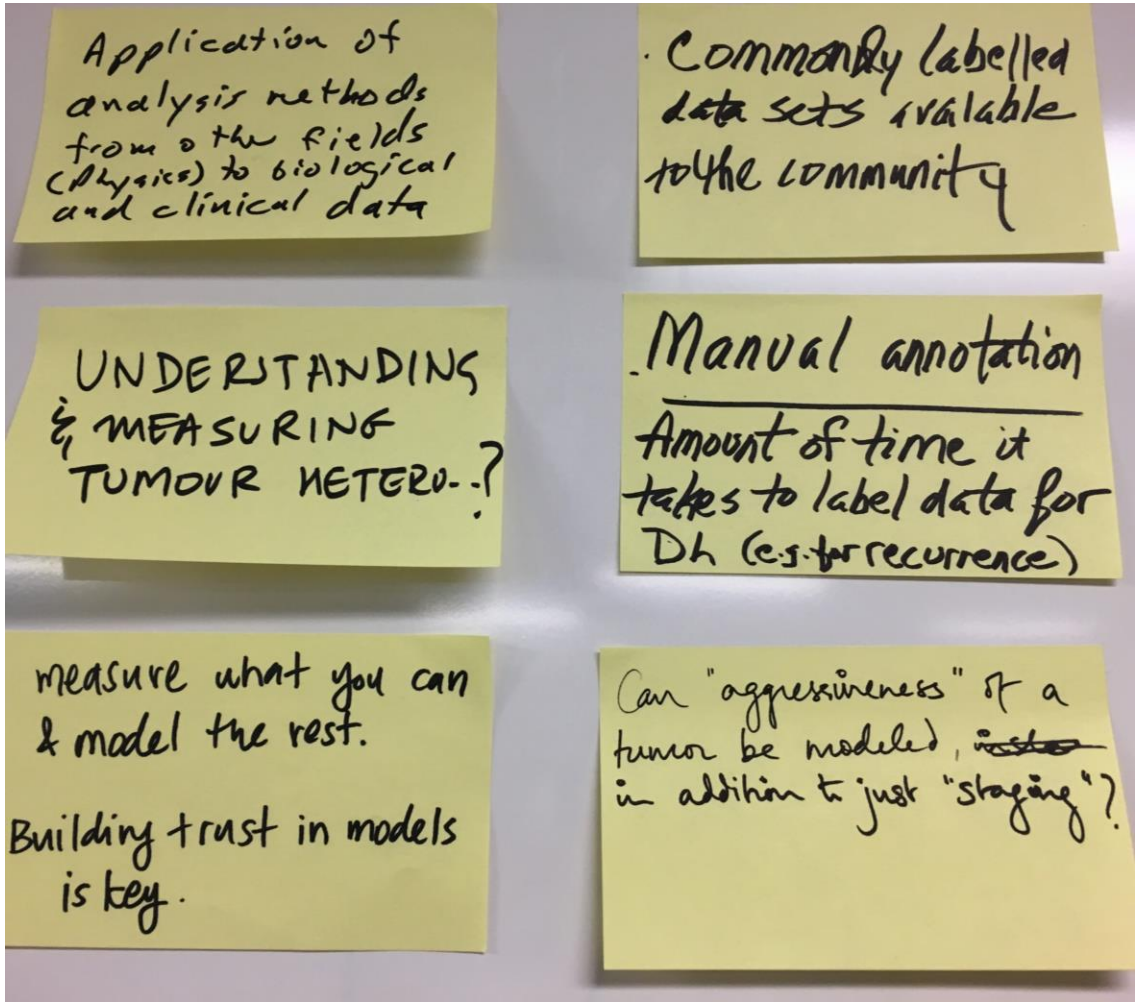
KEDAR: powerful imaging technique. Enable cube
~~show~~ tumor disease progression support prediction supplementary to genomic/proteomic data.

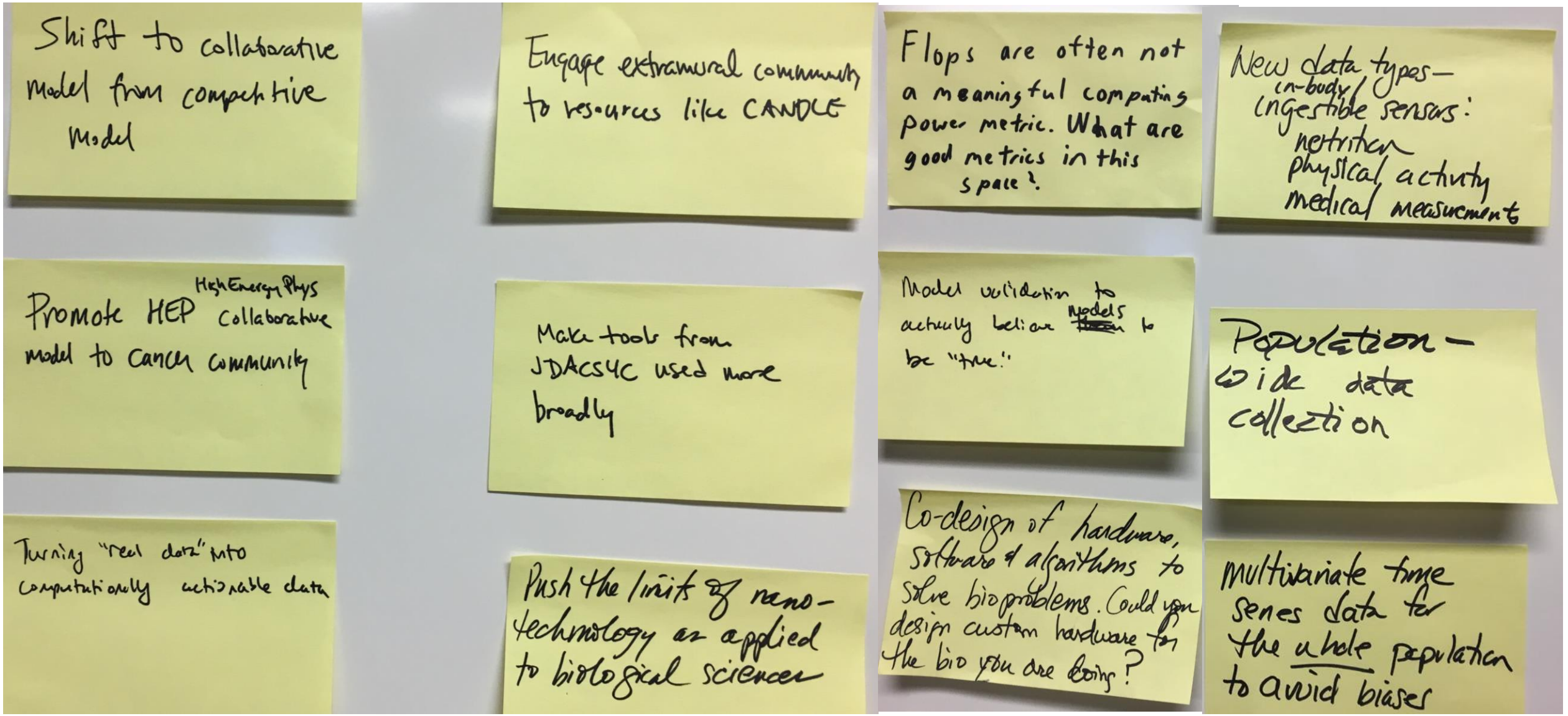
- Multimodal data integration, harmonization, and storage & management
- Centralized OR distributed
- Capture metadata
- How about data schemas
- Sustainability ** **

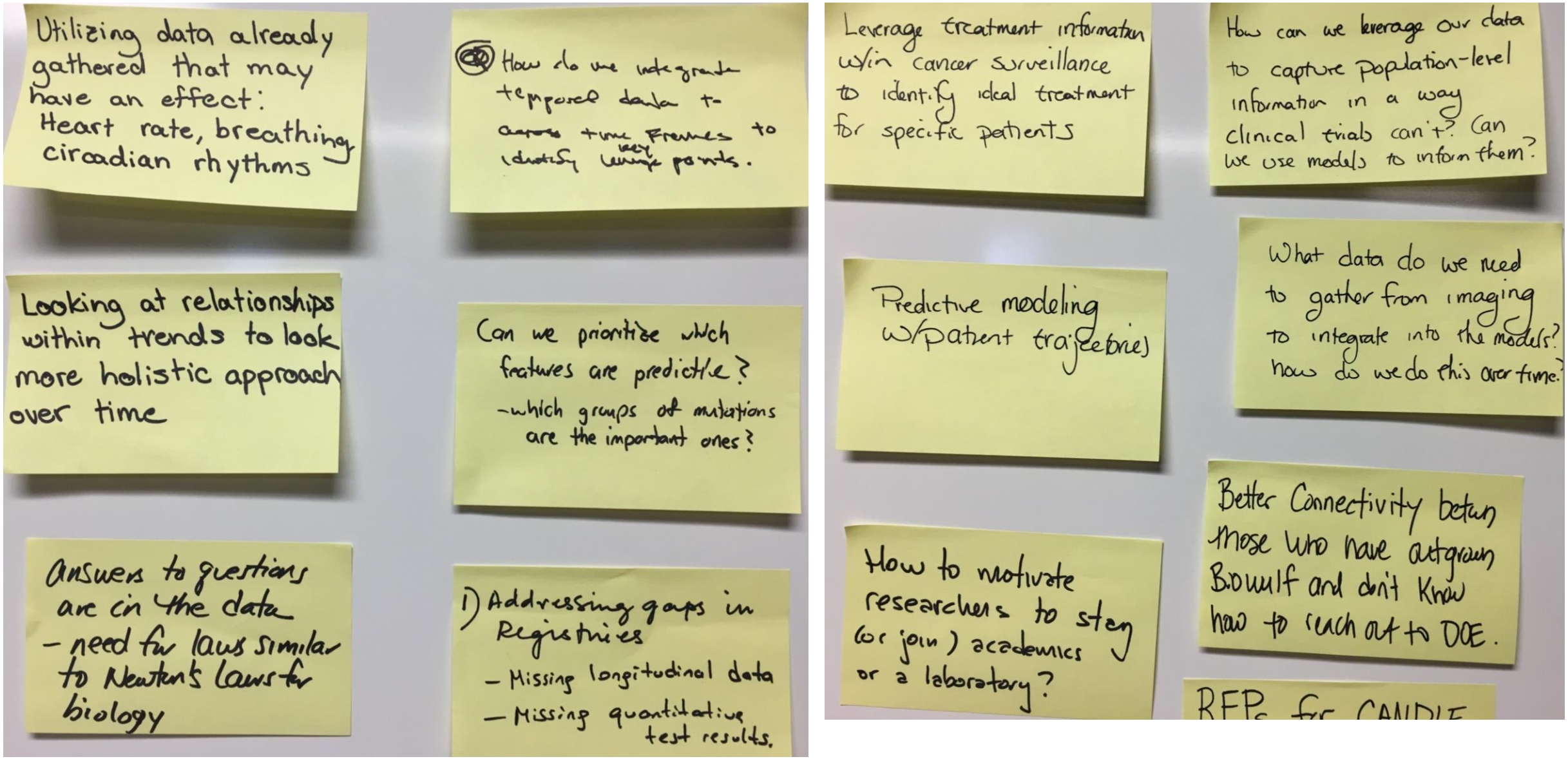
Lack of transparency of ML/algorithms outside of chemical track (ICT data)

Solutions are developed w/ short horizons due to limited resources or resources w/ expiration date.


-Demonstration of computational model and validation thereof are required for new
Rapid approximations of solutions versus precision which may take longer
detter entiating ~~btw~~ not a lot of data genomic features & tumor & drug features is expensive + difficult
Data Integration
BIOLOGISTS NEED A LOT MORE HANDHOLDING THAN IS BEING ASSUMED!
* NEED MORE DATA SCIENTISTS "CLOSER TO THE GROUND" & "101 HANDBOOKS".
Data sharing


Application of analysis methods from other fields (physics) to biological and clinical data
· Commonly labelled data sets available to the community
UNDERSTANDING & MEASURING TUMOUR HETERO-?
· Manual annotation
Amount of time it takes to label data for DL (e.g. for recurrence)
measure what you can & model the rest.
Building trust in models is key.
Can "aggressiveness" of a tumor be modeled, ~~instead~~ in addition to just "staging"?


Shift to collaborative model from competitive Model
Engage extramural community to resources like CANDLE
Flops are often not a meaningful computing power metric. What are good metrics in this space?
New data types— in-body/ ingestible sensors: nutrition physical activity medical measurement
Promote HEP (High Energy Phys) collaborative model to cancer community
Make tools from JDACS4C used more broadly
Model validation to actually believe models to be "true."
Population-wide data collection
Turning "real data" into computationally actionable data
Push the limits of nano-technology as applied to biological sciences
Co-design of hardware, software & algorithms to solve bio problems. Could you design custom hardware for the bio you are doing?
multivariate time series data for the whole population to avoid biases


Utilizing data already gathered that may have an effect: Heart rate, breathing circadian rhythms
How do we integrate temporal data to across time frames to identify key turning points.
Looking at relationships within trends to look more holistic approach over time
Can we prioritize which features are predictive?
-which groups of mutations are the important ones?
Answers to questions are in the data
- need for laws similar to Newton's laws for biology
1) Addressing gaps in Registries
- Missing longitudinal data
- Missing quantitative test results.
Leverage treatment information w/in cancer surveillance to identify ideal treatment for specific patients
How can we leverage our data to capture population-level information in a way clinical trials can't? Can we use models to inform them?
Predictive modeling w/patient trajectories
What data do we need to gather from imaging to integrate into the models? how do we do this over time?
How to motivate researchers to stay (or join) academics or a laboratory?
Better Connectivity betwn those who have outgrown Beowulf and don't know how to reach out to DOE.
RFPs for CANDLE

Virtual Drugs?

How does Biology community know what's now available?

Understand cell spatial interr to inform therapeutic mechanism and incorporate into ML modeling

Confidence intervals / uncertainty testing

Collect, right information How to use that information to feed into models and validate

-Access to data/Standardize
-Push more in community access dataset, data sharing
- In medical/clinical research communities

Defining higher level treatment strategies

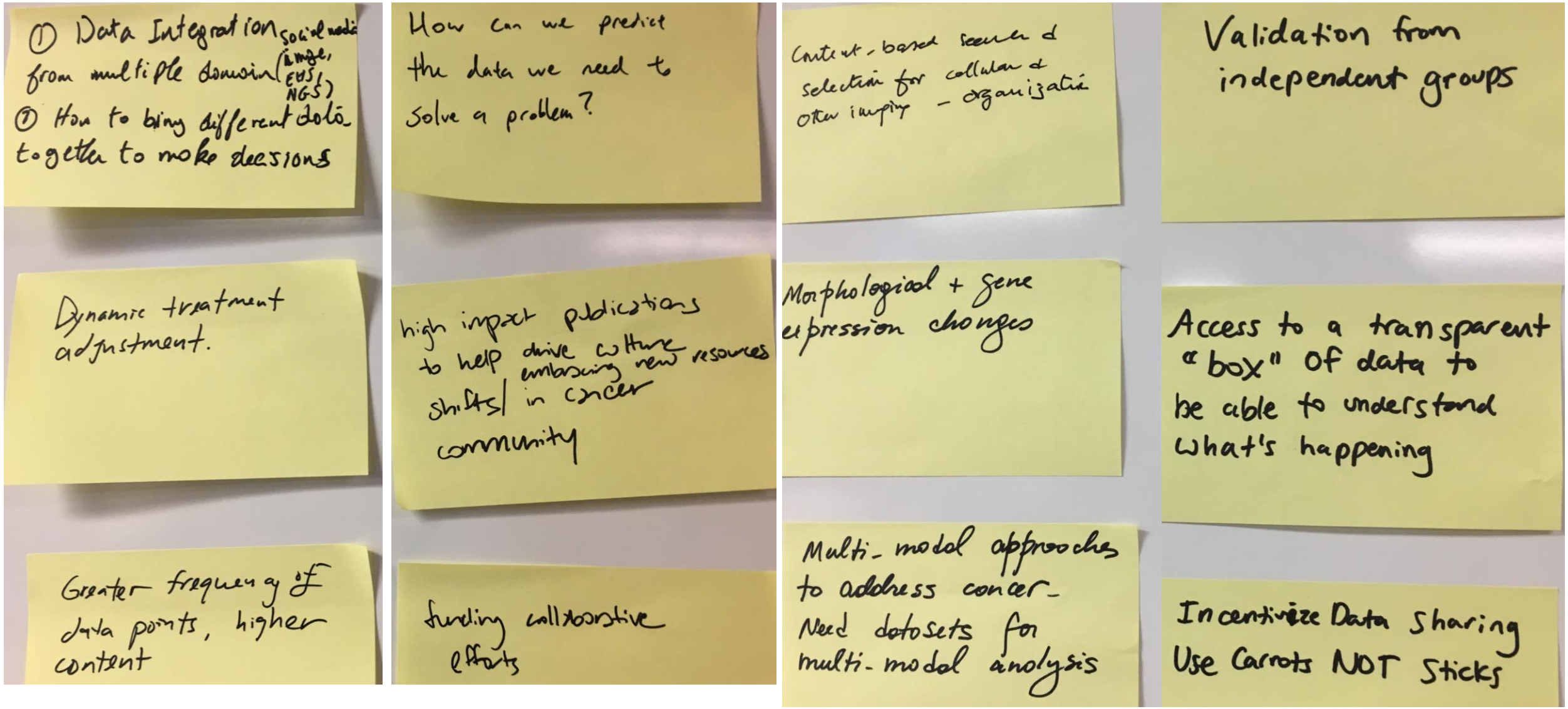
Being able to interrupt mechanisms reliably ↑ disease

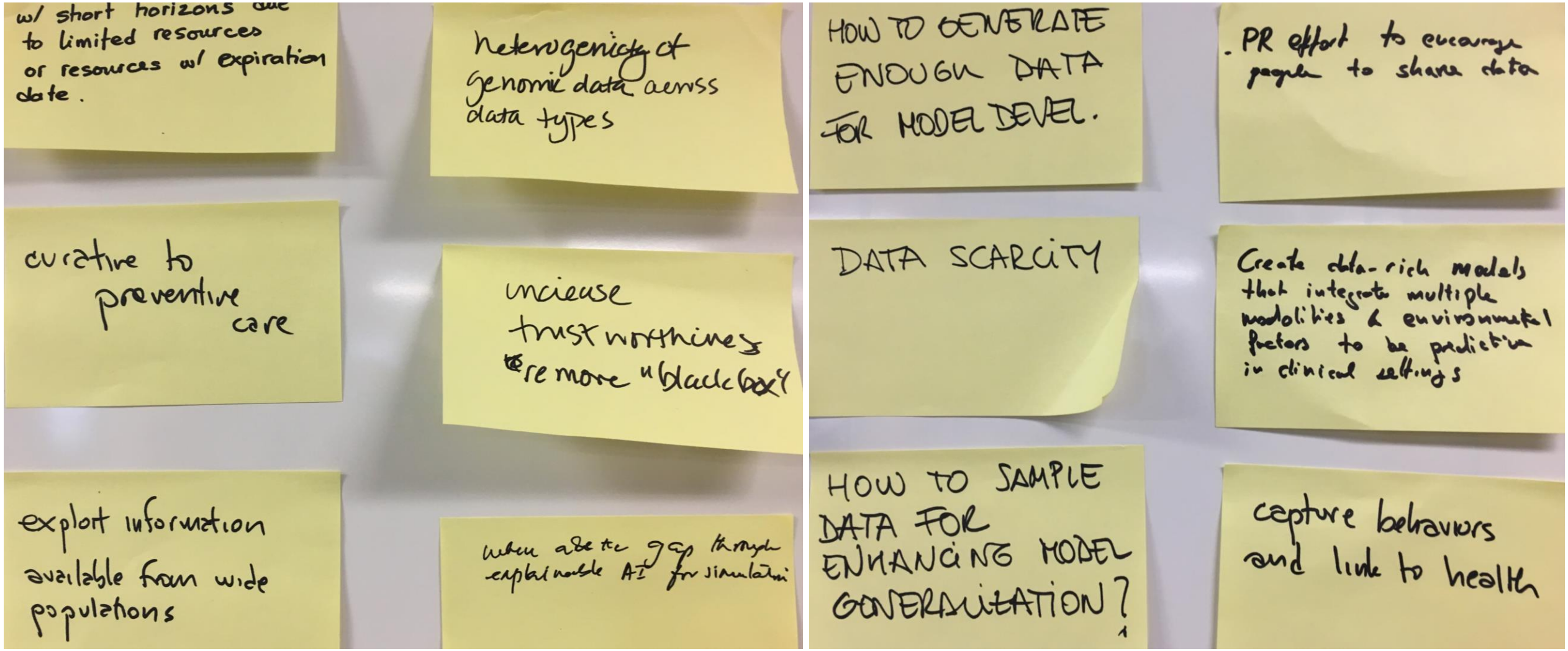
Interpretation of genomic variants of uncertain significance

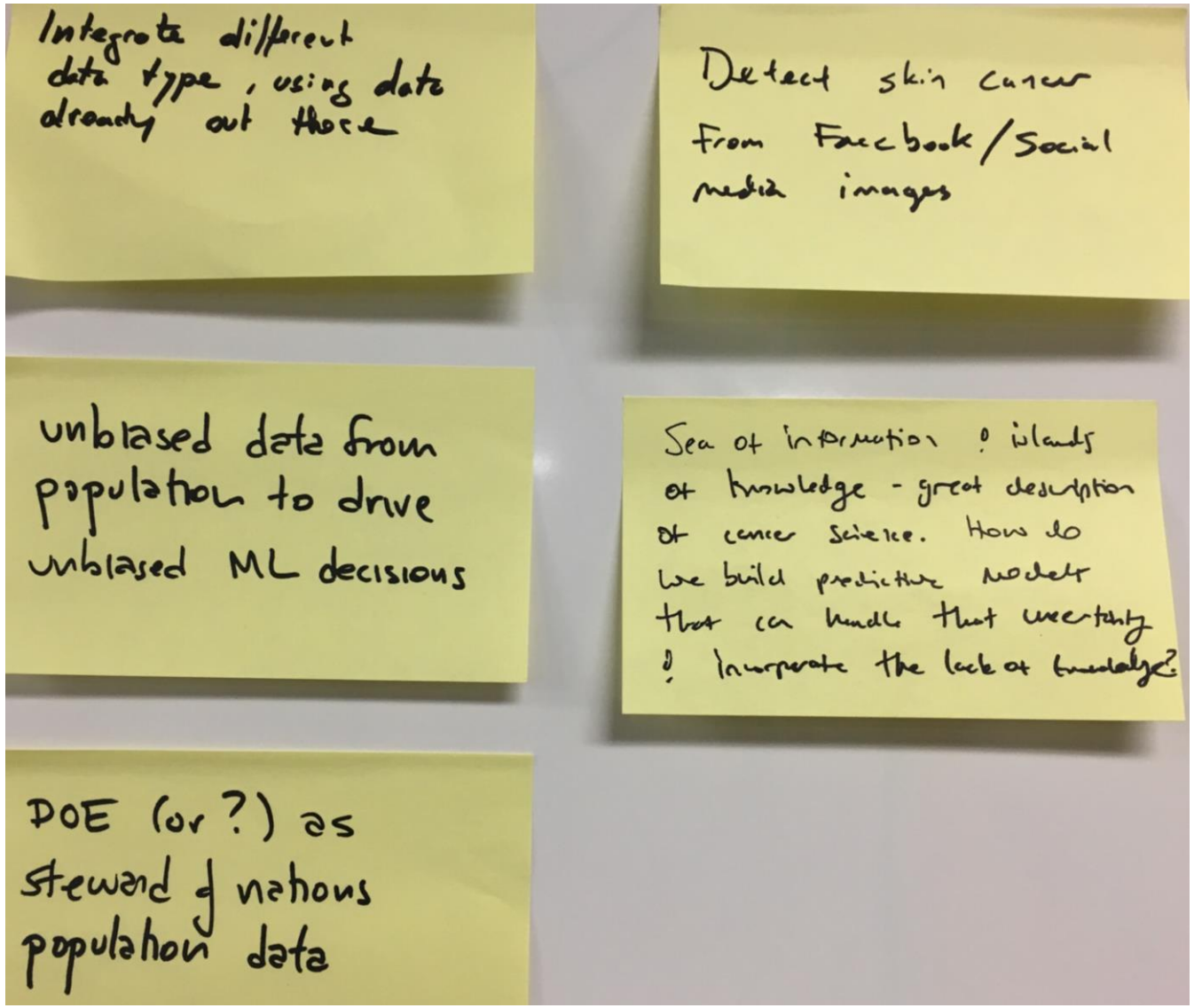
Overcoming steep learning curve for data science to use tools well

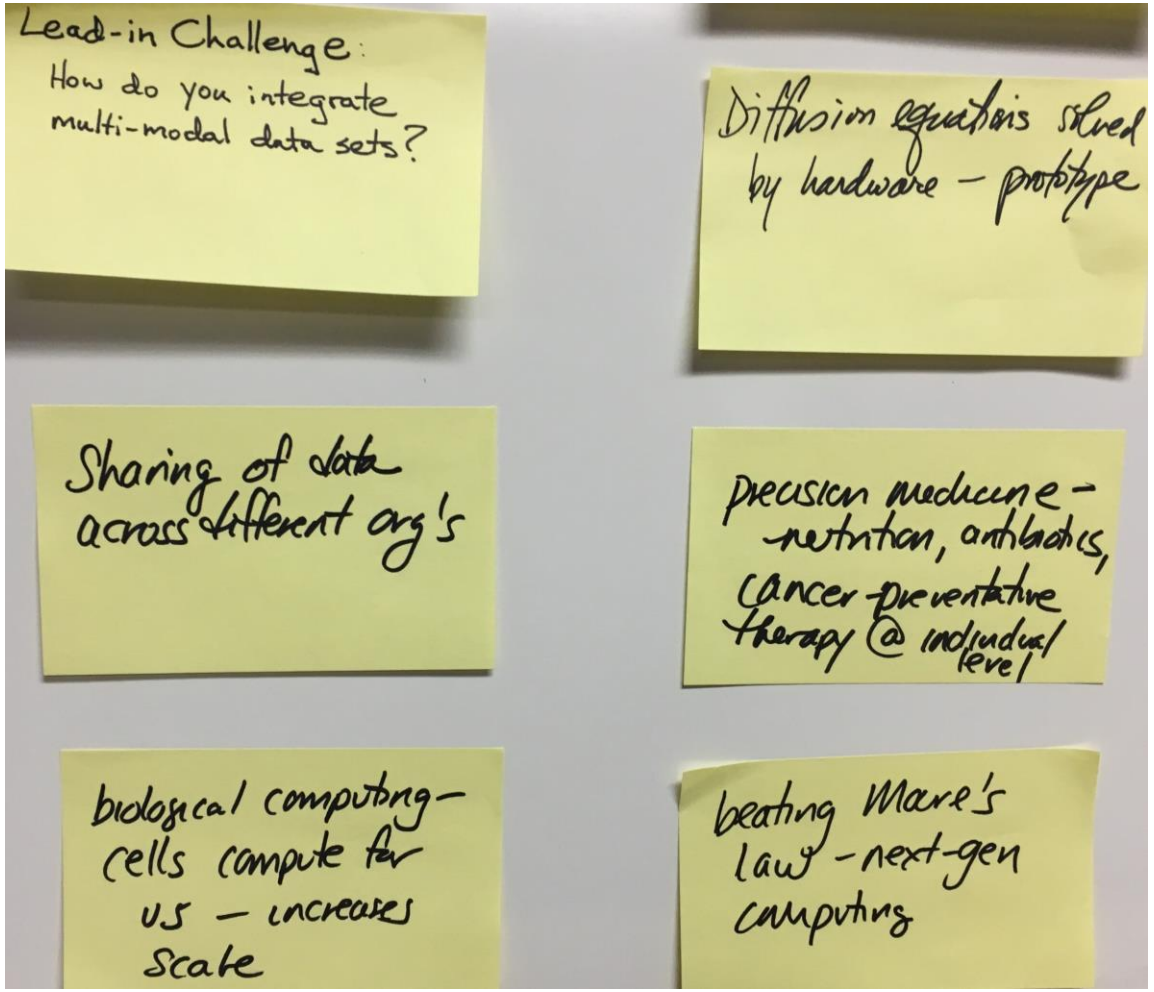
Enabling students to be effective in data science without programming

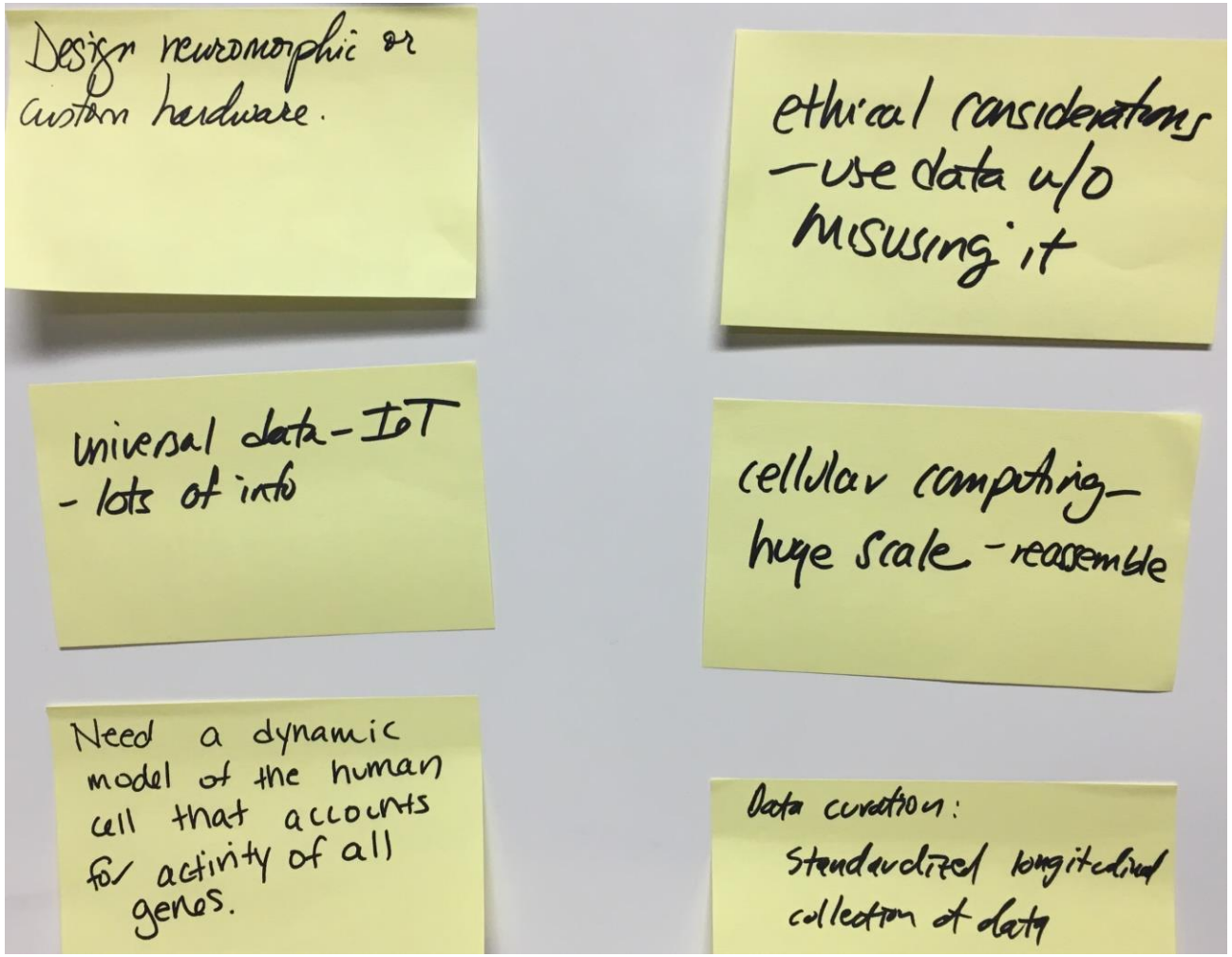
Democratic access to patient data to compute resources

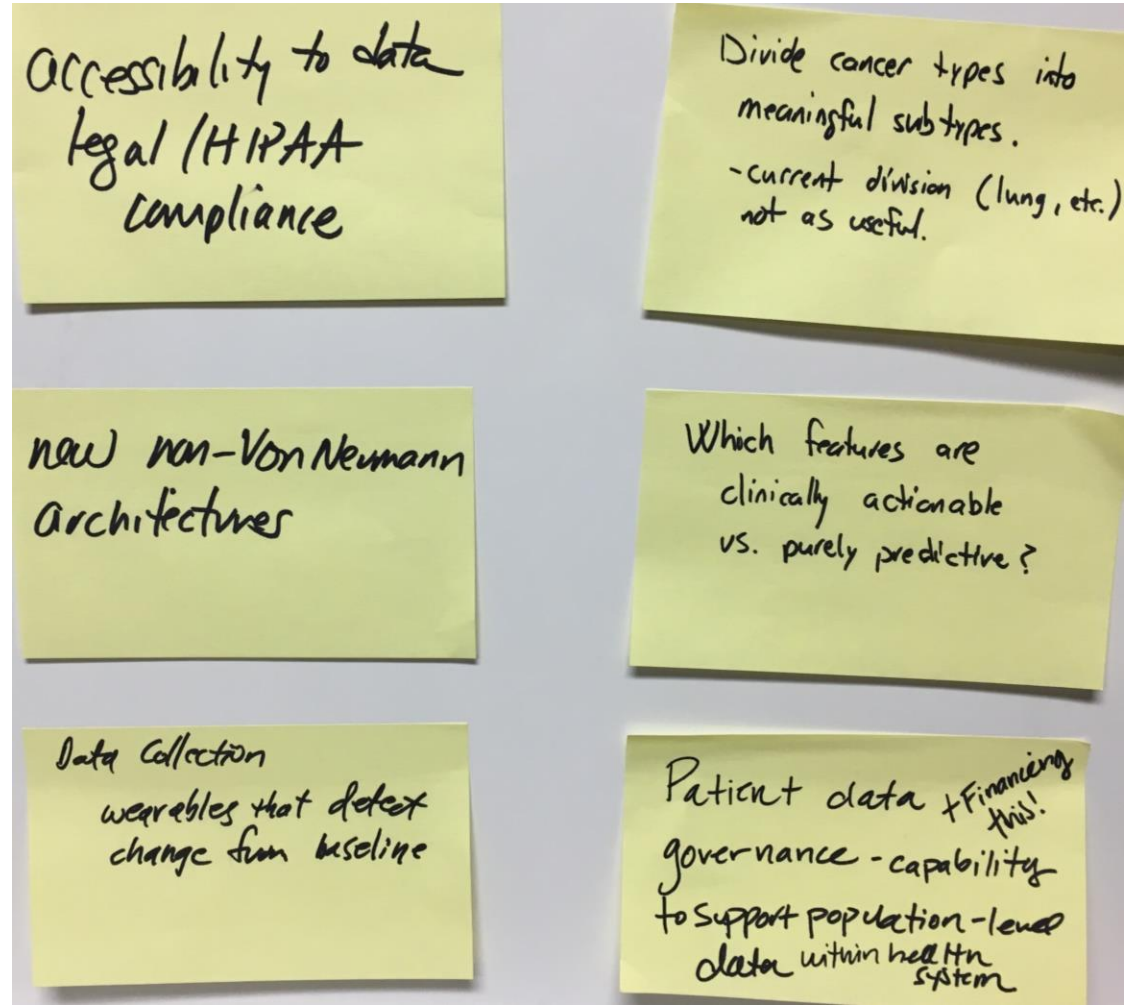

① Data Integration from multiple domains (Social media, Image, EHR, NGS)
② How to bring different data together to make decisions
How can we predict the data we need to solve a problem?
Content-based search & selection for cellular & other imaging - organizati
Validation from independent groups
Dynamic treatment adjustment.
high impact publications to help drive culture shifts/ in cancer community
embracing new resources
Morphological + gene expression changes
Access to a transparent "box" of data to be able to understand what's happening
Greater frequency of data points, higher content
funding collaborative efforts
Multi-modal approaches to address cancer. Need datasets for multi-modal analysis
Incentivize Data Sharing Use Carrots NOT Sticks

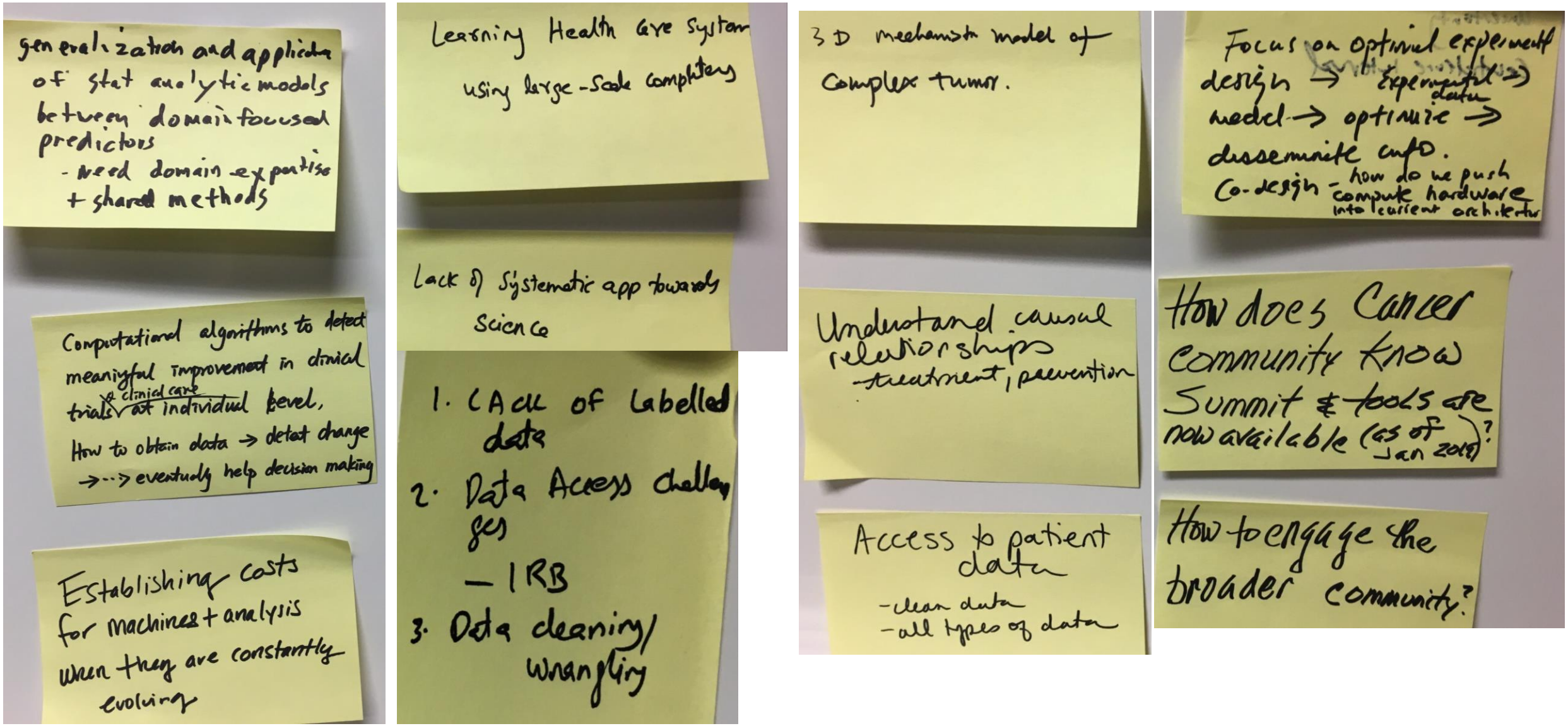

w/ short horizons due to limited resources or resources w/ expiration date.
heterogenicity of genomic data across data types
curative to preventive care
increase trustworthiness & remove "black box"
explort information available from wide populations
HOW TO GENERATE ENOUGH DATA FOR MODEL DEVEL.
PR effort to encourage people to share data
DATA SCARCITY
Create data-rich models that integrate multiple modalities & environmental factors to be predictive in clinical settings
HOW TO SAMPLE DATA FOR ENHANCING MODEL GENERALIZATION?
capture behaviors and link to health


Integrate different data type, using data already out there
Detect skin cancer from Facebook/Social media images
unbiased data from population to drive unbiased ML decisions
Sea of information ! islands of knowledge - great description of cancer science. How do we build predictive models that can handle that uncertainty ! Incorporate the lack of knowledge?
DOE (or ?) as steward of nations population data


Lead-in Challenge:
How do you integrate multi-modal data sets?
Diffusion equations solved by hardware - prototype
Sharing of data across different org's
precision medicine - nutrition, antibiotics, cancer preventative therapy @ individual level
biological computing - cells compute for us - increases scale
beating Moore's law - next-gen computing


Design neuromorphic or custom hardware.
ethical considerations - use data w/o misusing it
universal data - IoT - lots of info
cellular computing - huge scale - reassemble
Need a dynamic model of the human cell that accounts for activity of all genes.
Data curation: Standardized longitudinal collection of data


accessibility to data legal / HIPAA compliance
Divide cancer types into meaningful subtypes.
- current division (lung, etc.) not as useful.
new non-Von Neumann architectures
Which features are clinically actionable vs. purely predictive?
Data collection wearables that detect change from baseline
Patient data + Financing this!
governance - capability to support population-level data within health system



generalization and application of stat analytic models between domain focused predictors
- need domain expertise + shared methods

Learning Health Care System using large-Scale computers

3D mechanistic model of Complex tumor.

Focus on optimal experiment design → experimental data → model → optimize → disseminate info.
Co-design - how do we push compute hardware into current architectur

Computational algorithms to detect meaningful improvement in clinical trials ~~& clinical care~~ at individual level,
How to obtain data → detect change
→ --> eventually help decision making

Lack of Systematic app towards Science

Understand causal relationships
- treatment, prevention

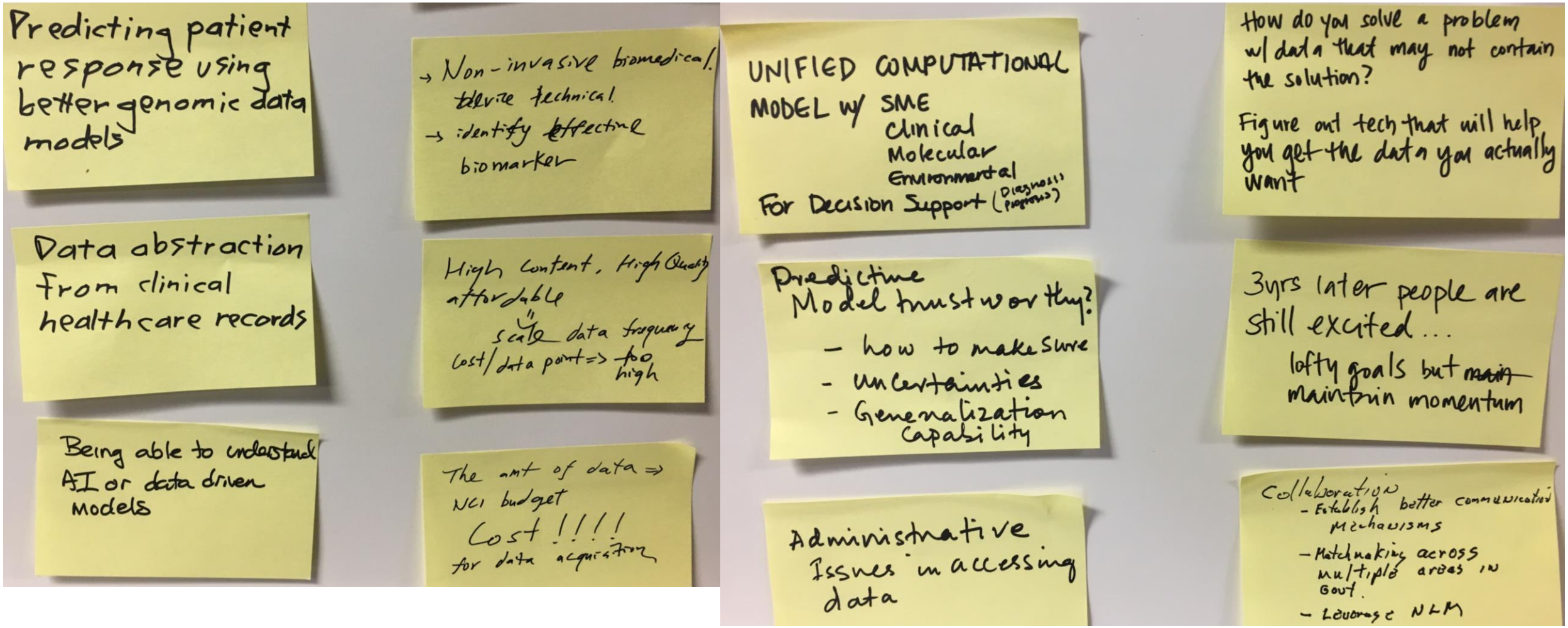
How does Cancer community know Summit & tools are now available (as of Jan 2018)?

Establishing costs for machines + analysis when they are constantly evolving

1. CAcll of labelled data
2. Data Access challenges
   - IRB
3. Data cleaning/ wrangling

Access to patient data
- clean data
- all types of data

How to engage the broader community?



Predicting patient response using better genomic data models

→ Non-invasive biomedical device technical
→ identify effective biomarker

Data abstraction from clinical healthcare records

High content, High Quality affordable
" scale data frequency
cost/data point => too high

Being able to understand AI or data driven models

The amt of data ⇒ NCI budget
Cost !!!!
for data acquisition

UNIFIED COMPUTATIONAL MODEL w/ SME
Clinical
Molecular
Environmental
For Decision Support (Diagnosis Prognosis)

How do you solve a problem w/ data that may not contain the solution?

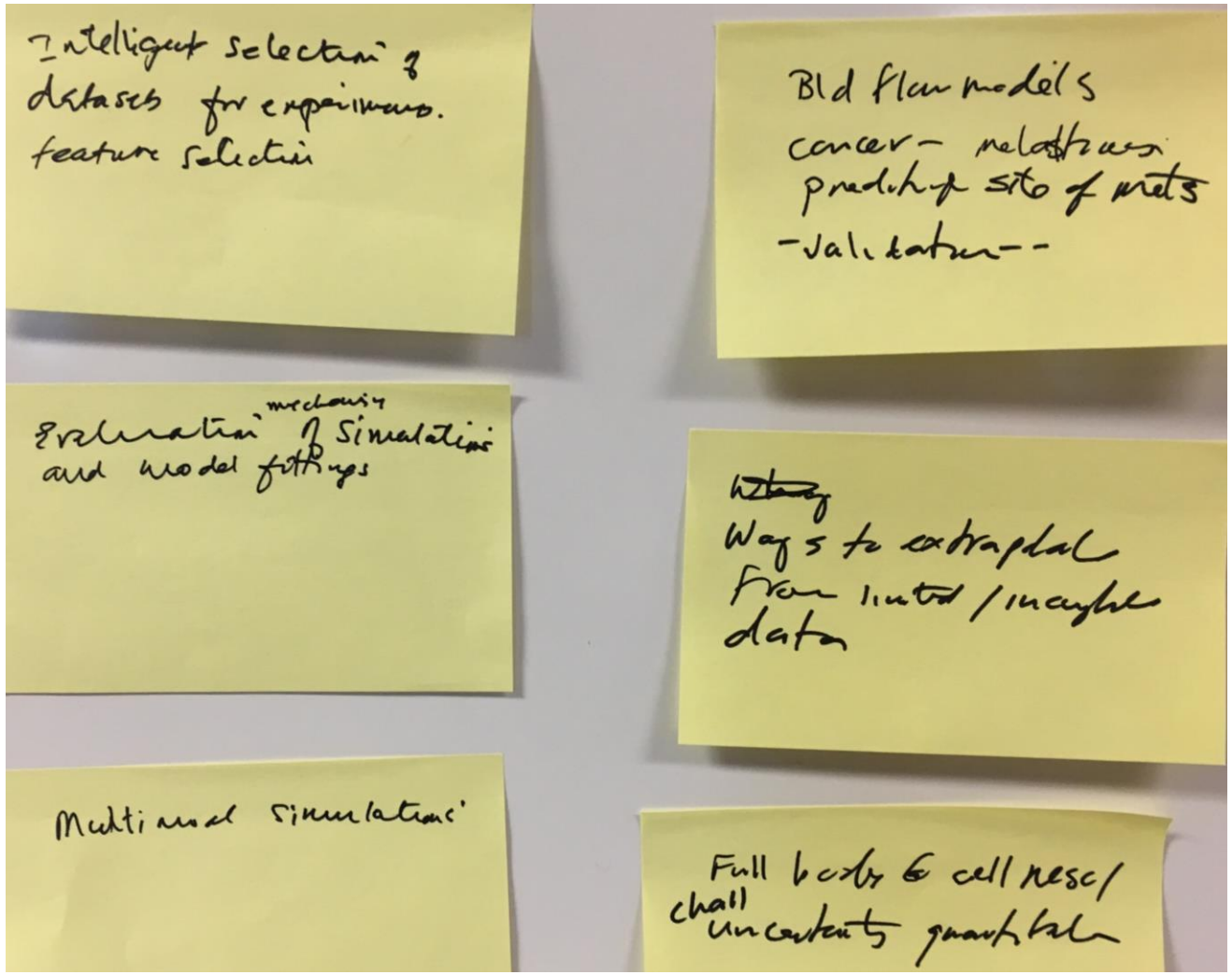
Figure out tech that will help you get the data you actually want

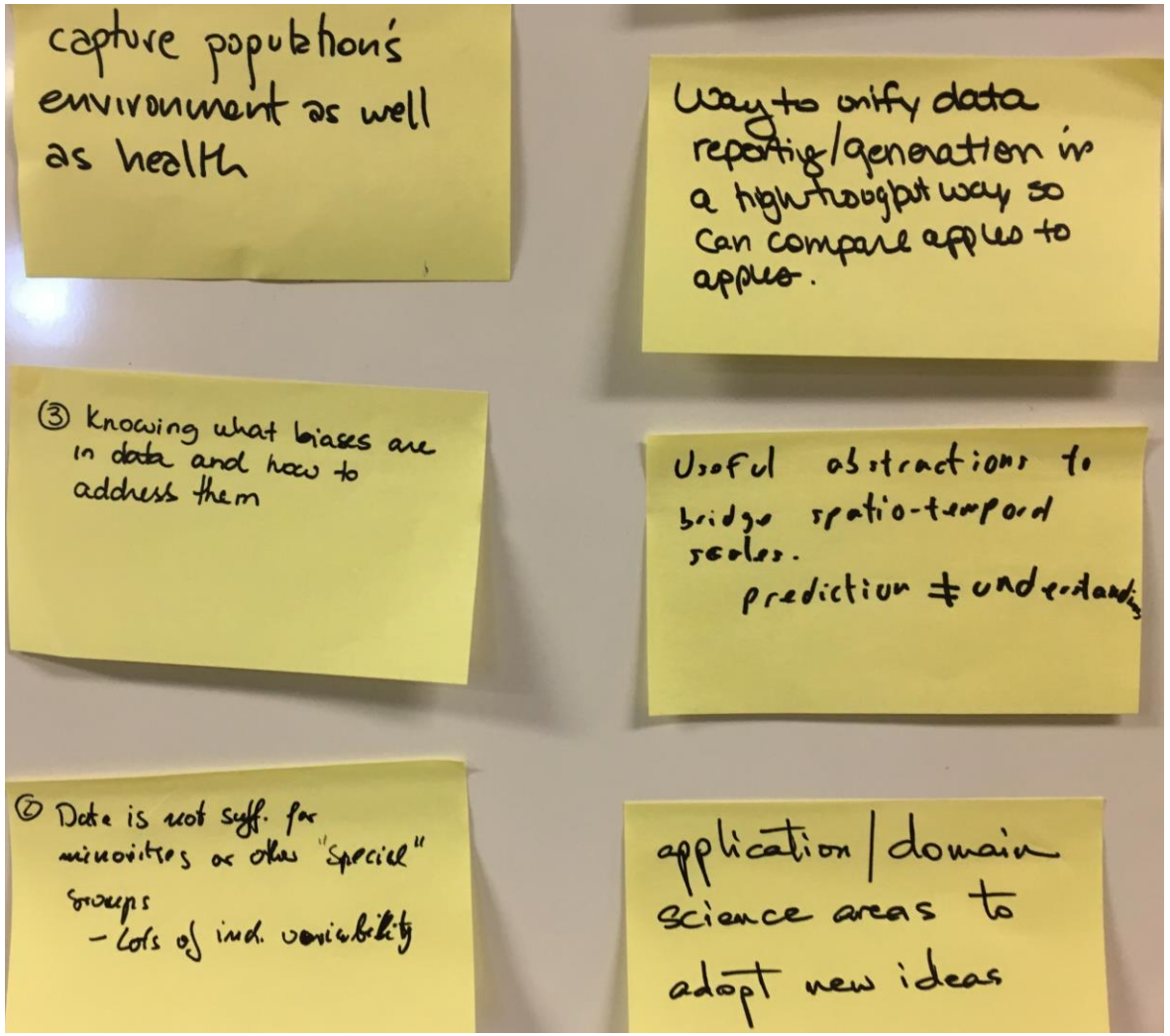
Predictive Model trustworthy?
- how to make sure
- uncertainties
- Generalization capability

3yrs later people are still excited...
lofty goals but maintain momentum

Administrative Issues in accessing data

Collaboration
- Establish better communication mechanisms
- Matchmaking across multiple areas in Gout.
- Leverage NLM


Intelligent selection of datasets for experiments. feature selection
Bld flow models cancer- metastases predict sites of mets -validation--
Evaluation mechanism of Simulations and model fittings
Ways to extrapolate from limited/incomplete data
Multiscale simulations
Full body & cell resc/ chall uncertainty quantification


capture population's environment as well as health
Way to unify data reporting/generation in a high throughput way so can compare apples to apples.
③ Knowing what biases are in data and how to address them
Useful abstractions to bridge spatio-temporal scales. prediction ≠ understanding
⑥ Data is not suff. for minorities or other "special" groups - lots of ind. variability
application/domain science areas to adopt new ideas

HOW TO LEVERAGE EPIDEMIOLOGICAL DATA WITHOUT INTRODUCING BIAS IN THE MODEL?

Common mutation pathways across cancer patients

Validation of algorithm performance across different locations and platforms at a level to support clinical patient care.

Computational infrastructure that allows real-time analysis with simulations

Single cell analysis of gene expression

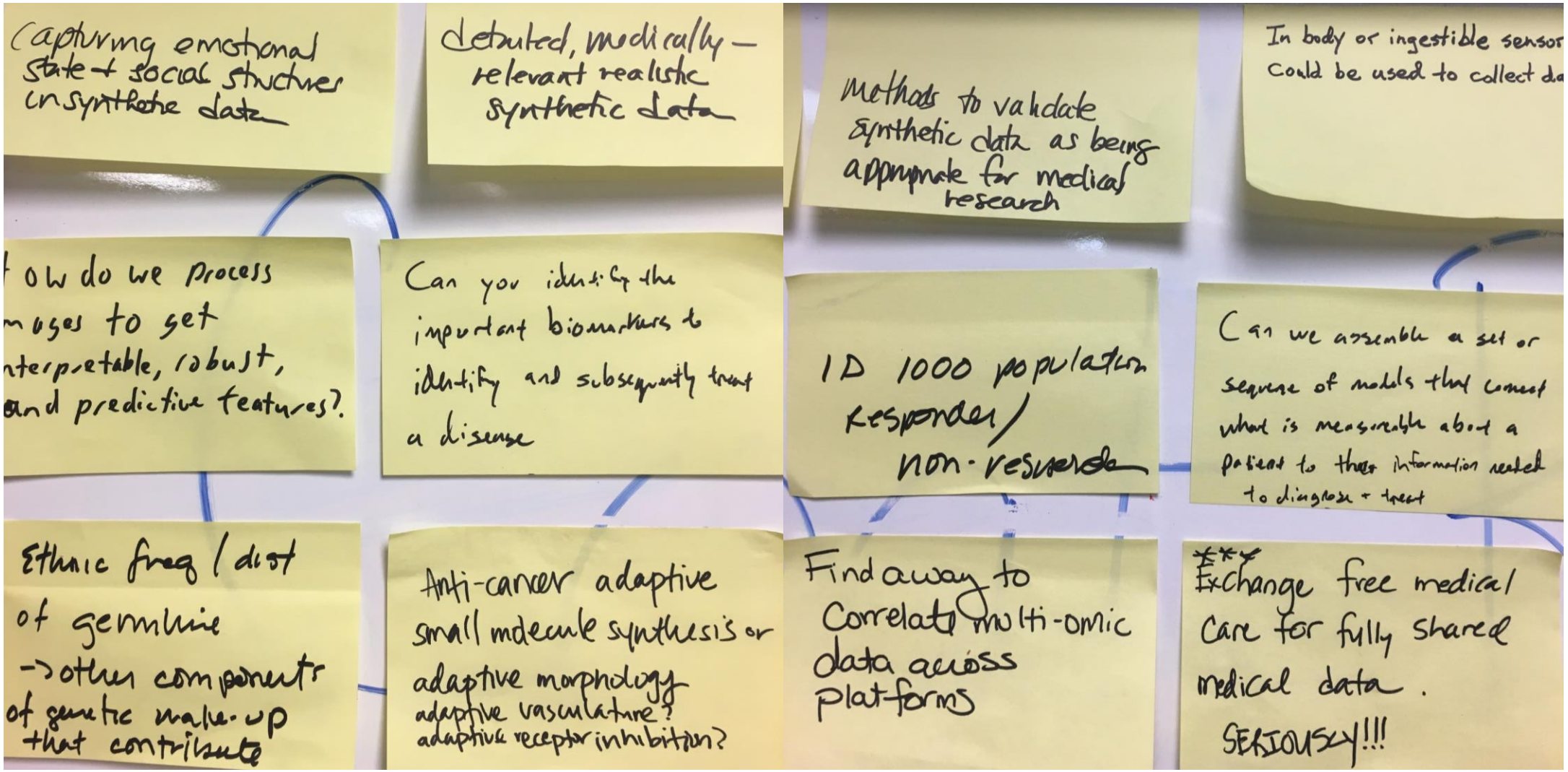
Pipeline for doctors to submit images for immediate diagnosis / suggestion of follow-up tests / treatment

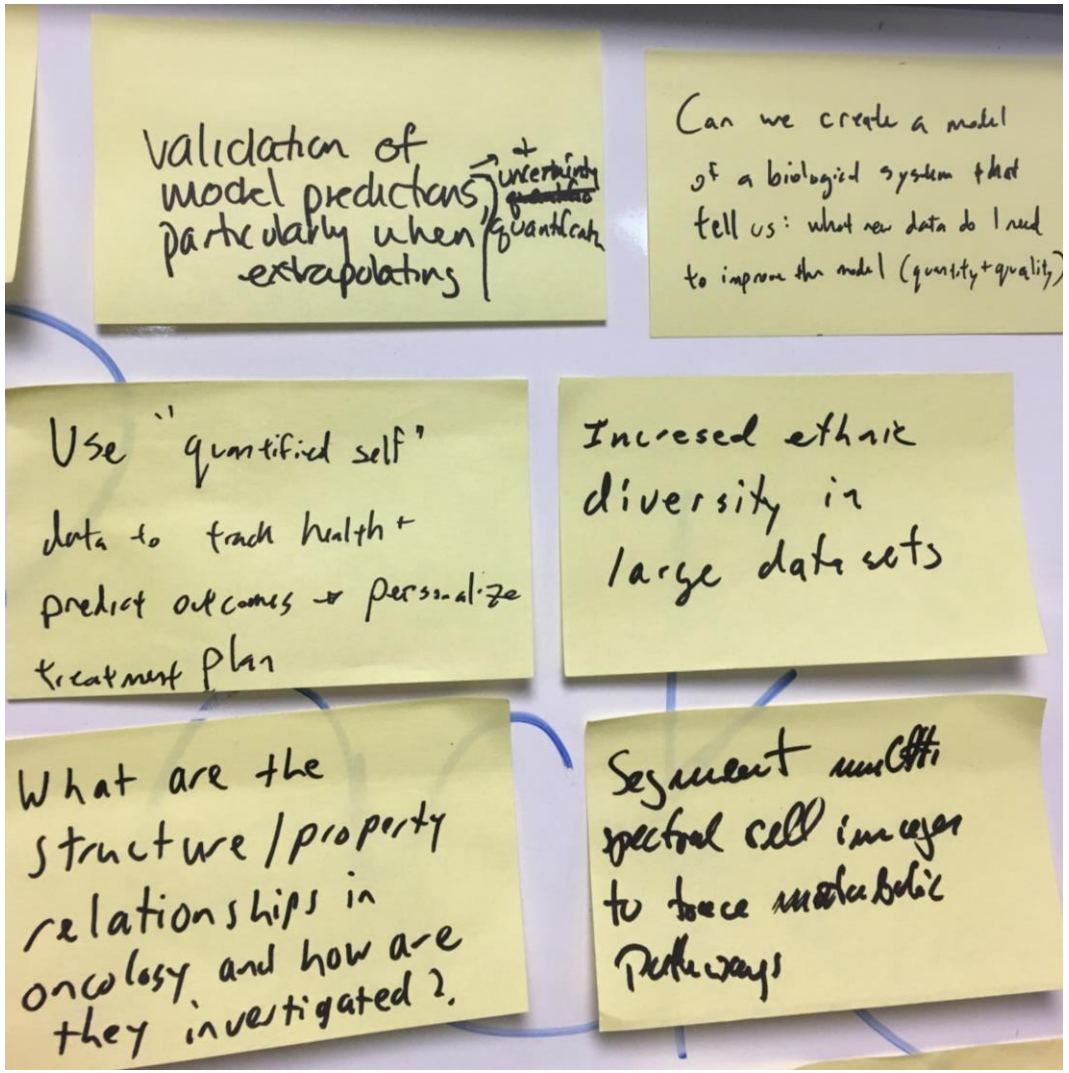
Physic-based modeling

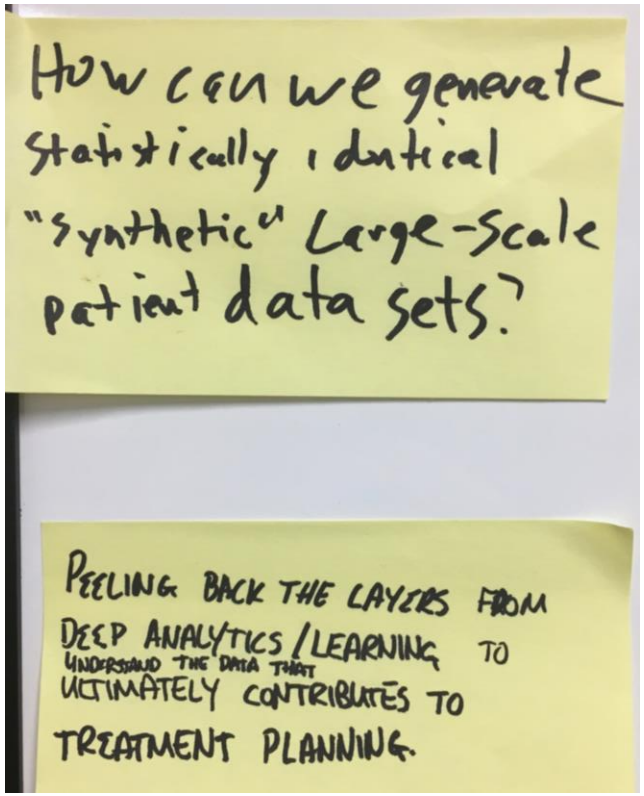
A.I. pipeline for IDing cancer/tumors in images from patients

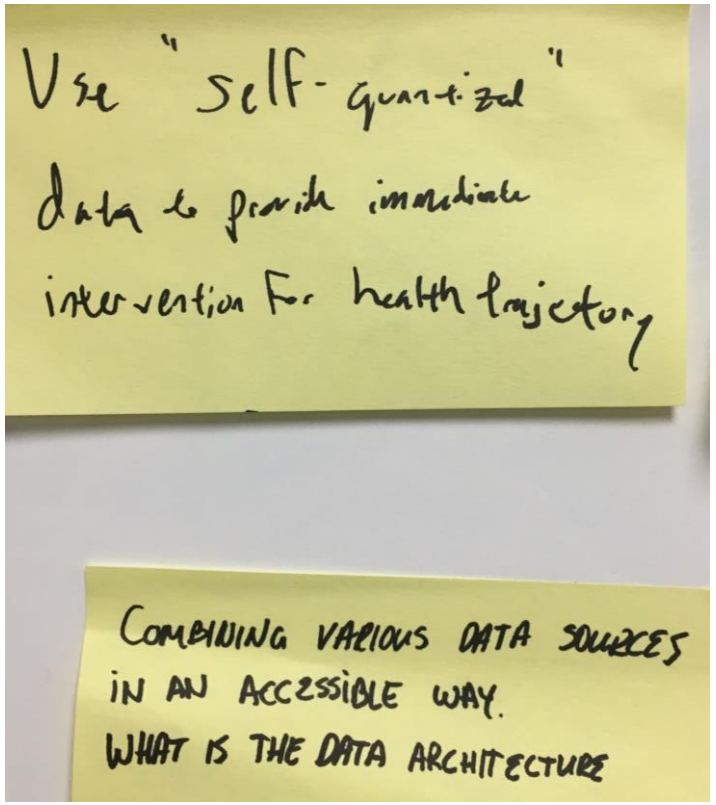
Thursday, March 7, 2018, Morning
Final grouping following Wednesday.

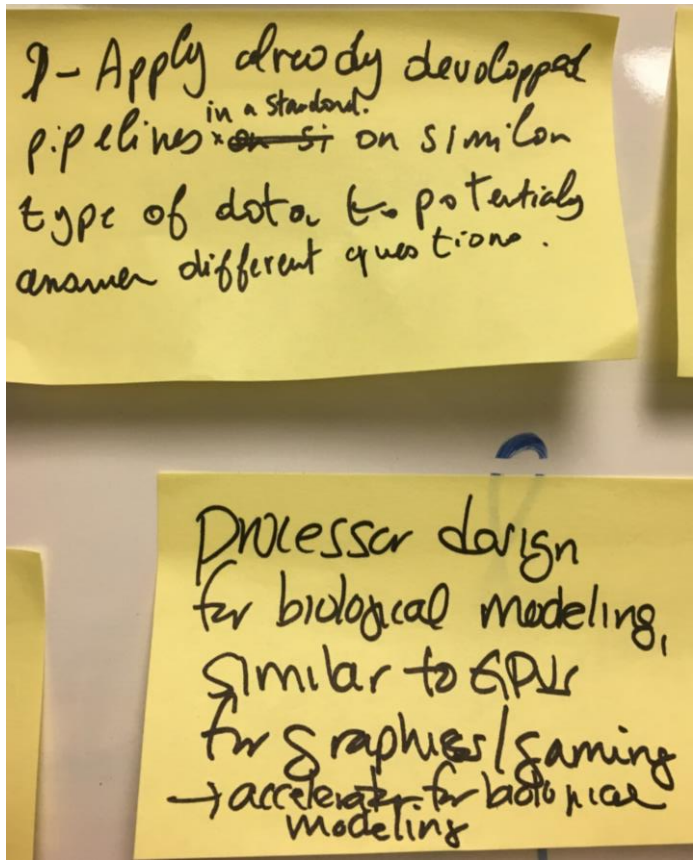
10:45 am – 11:30 am	**A Coffee Break with a Purpose** - Generating ideas for lean-in challenge areas (part 1)
11:30 am – 12:30 pm	**KEYNOTE PRESENTATION** **DOE Success Story: Leading with Science in a Computational Context** Peter Nugent, *LBL*
12:30 pm – 2:00 pm	**Working Lunch** - Generating ideas for lean-in challenge areas (part 2)
2:00 pm – 3:00 pm	**SESSION 2: CHALLENGE AREAS IN CANCER RESEARCH** **Panel:** Gregory Cooper, *University of Pittsburgh* Tina Hernandez-Boussard, *Stanford University* Paul Macklin, *Indiana University* John Quackenbush, *Harvard University* Amanda Randles, *Duke University* William Richards, *Brigham and Women's Hospital* Ilya Shmulevich, *Institute for Systems Biology* Amber Simpson, *MSKCC* **Moderators:** Amy Gryshuk, *LLNL* Roxanne Jensen, *NCI*
3:00 pm – 3:45 pm	**SPEED NETWORKING** - Generating ideas for lean-in challenge areas (part 3)

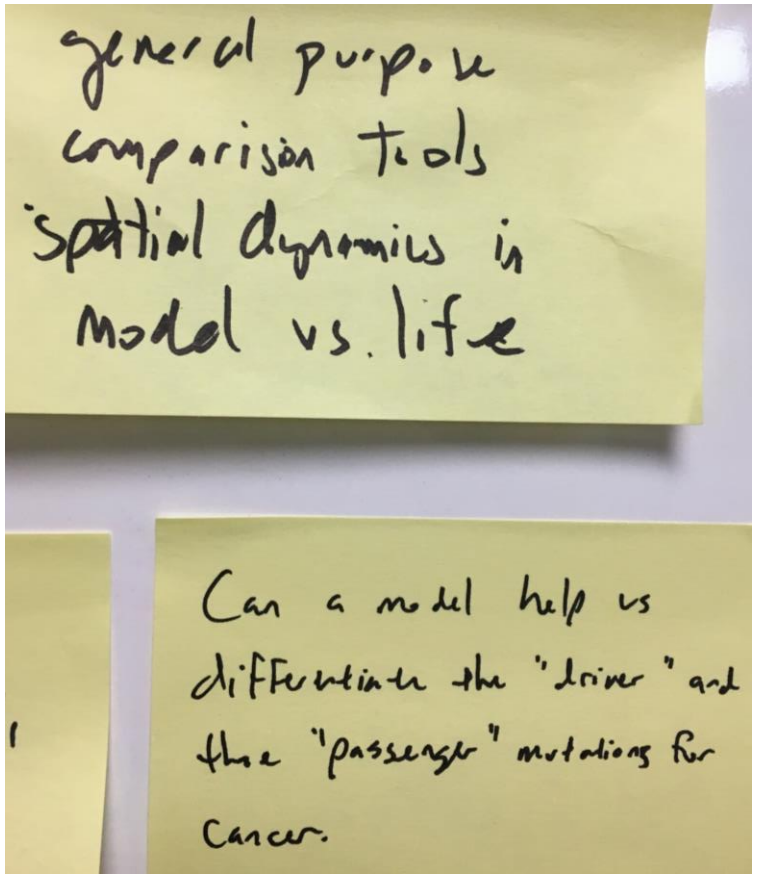

Capturing emotional state + social structures in synthetic data
detailed, medically – relevant realistic synthetic data
Methods to validate synthetic data as being appropriate for medical research
In body or ingestible sensor
Could be used to collect da
How do we process images to get interpretable, robust, and predictive features?
Can you identify the important biomarkers to identify and subsequently treat a disease
ID 1000 population responder / non-responder
Can we assemble a set or sequence of models that connect what is measurable about a patient to their information needed to diagnose + treat
Ethnic freq / dist of germline -> other components of genetic make-up that contribute
Anti-cancer adaptive small molecule synthesis or adaptive morphology adaptive vasculature? adaptive receptor inhibition?
Find away to Correlate multi-omic data across platforms
*** Exchange free medical care for fully shared medical data. SERIOUSLY!!!

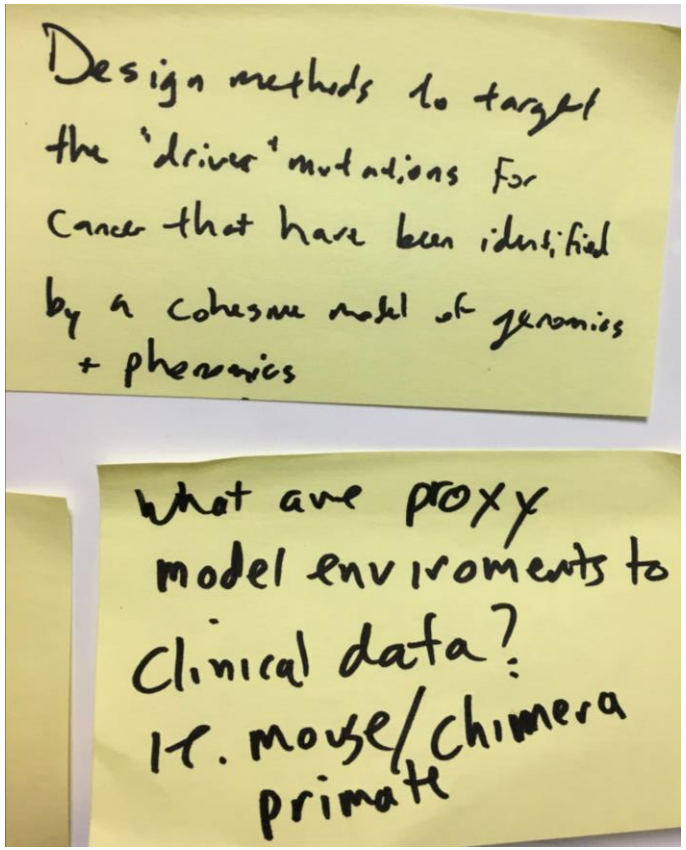

validation of model predictions particularly when extrapolating
+ uncertainty quantification
Can we create a model of a biological system that tell us: what new data do I need to improve the model (quantity + quality)
Use "quantified self" data to track health + predict outcomes + personalize treatment plan
Increased ethnic diversity in large data sets
What are the structure / property relationships in oncology and how are they investigated?
Segment multi spectral cell images to trace metabolic pathways


How can we generate statistically identical "synthetic" Large-scale patient data sets?
PEELING BACK THE LAYERS FROM DEEP ANALYTICS / LEARNING TO UNDERSTAND THE DATA THAT ULTIMATELY CONTRIBUTES TO TREATMENT PLANNING.


Use "self-quantized" data to provide immediate intervention for health trajectory
COMBINING VARIOUS DATA SOURCES IN AN ACCESSIBLE WAY.
WHAT IS THE DATA ARCHITECTURE


1- Apply already developped pipelines in a standard. on silicon type of data to potentially answer different questions.
Processor design for biological modeling, similar to GPU for graphics/gaming → accelerator for biological modeling


general purpose comparison tools spatial dynamics in model vs. life
Can a model help us differentiate the "driver" and the "passenger" mutations for Cancer.


Design methods to target the "driver" mutations for Cancer that have been identified by a cohesive model of genomics + phenomics
What are proxy model environments to Clinical data? ie. mouse/chimera primate

apply "causal inference" algorithms to [redacted] complex clinical data to sub-classify [redacted] tumor type

what range of incentives are needed to get patients to contribute their "Normal"/non-normal clinical data.

- Multimodal data integration, harmonization, and storage & management
- Centralized OR distributed
- Capture metadata
- How about data schemas
- Sustainability ** **

Solutions are developed w/ short horizons due to limited resources or resources w/ expiration date.

increase trustworthiness & remove "black box"

HOW TO LEVERAGE EPIDEMIOLOGICAL DATA WITHOUT INTRODUCING BIAS IN THE MODEL?

unbiased data from population to drive unbiased ML decisions

Way to unify data reporting/generation in a high throughput way so can compare apples to apples.

Heterogeneity of genomic data across data types

③ Knowing what biases are in data and how to address them

Validation of algorithm performance across different locations and platforms at a level to support clinical patient care.

lack of transparency of ML/algorithms outside of clinical trials (ie data)

GET ENOUGH PPL SEQUENCED (AND HAVE DATA IN AN ACCESSIBLE FORMAT) TO BUILD ALL THESE 'FANCY' COMPUTATIONAL MODELS.

curative to preventive care

Single cell analysis of gene expression

evaluation mechanism ? Simulations and model fittings

Common mutation pathways across cancer patients

Multiscale simulations

HOW TO STANDARDIZE DATA GENERATION?

HOW TO DEAL WITH HETEROGENEOUS DATA (USING DIF. STANDARDS)?

Intelligent selection of datasets for experiments. feature selection

Reasoning deep learning neurons for more complex modeling

~~Opti~~ Biology-inspired computing

Computational infrastructure that allows real-time analysis with simulations

New areas of opportunity: Reanalyzing data that has gotten disassociated from experimental Meta data (applies to Nuclear tests AND biology!)

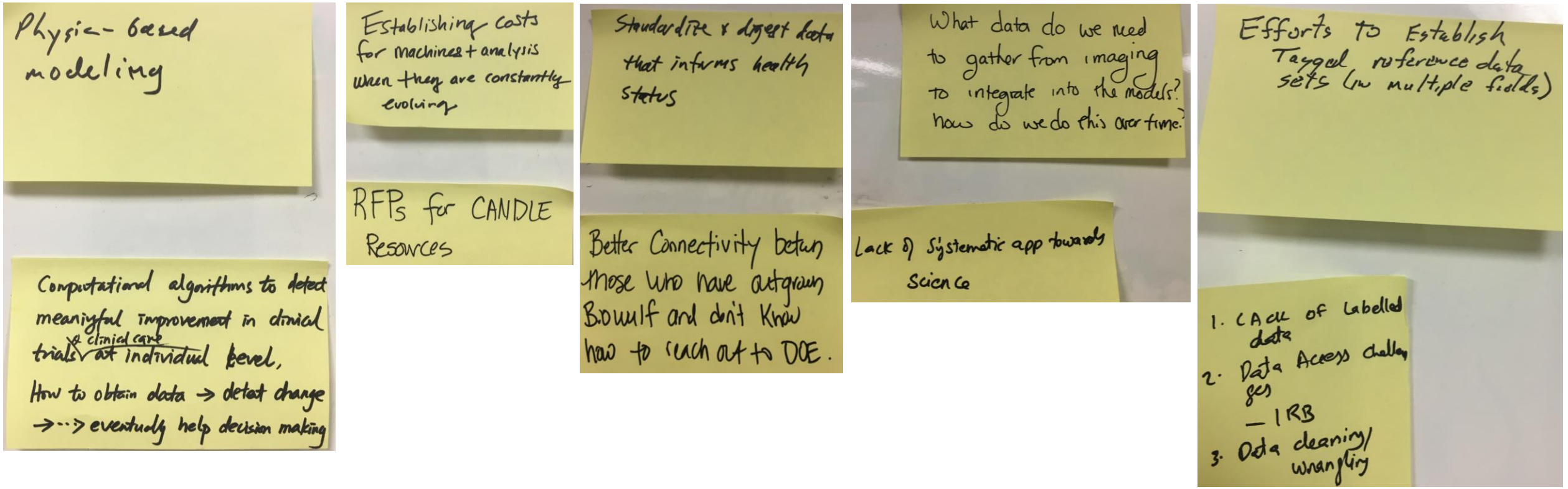
Useful abstractions to bridge spatio-temporal scales.

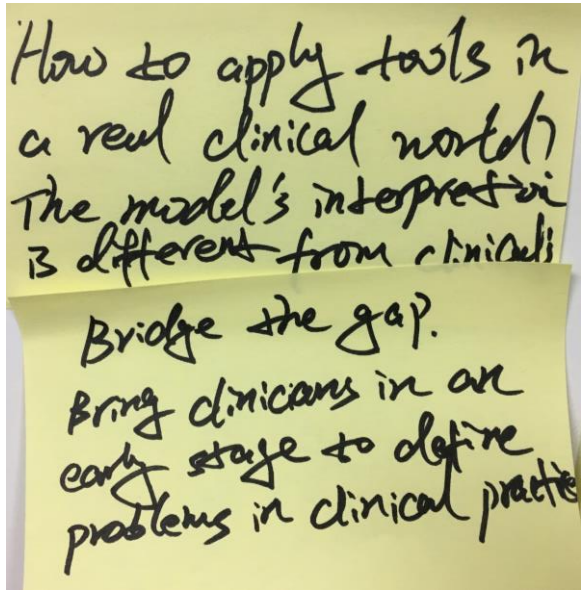
prediction ≠ understanding

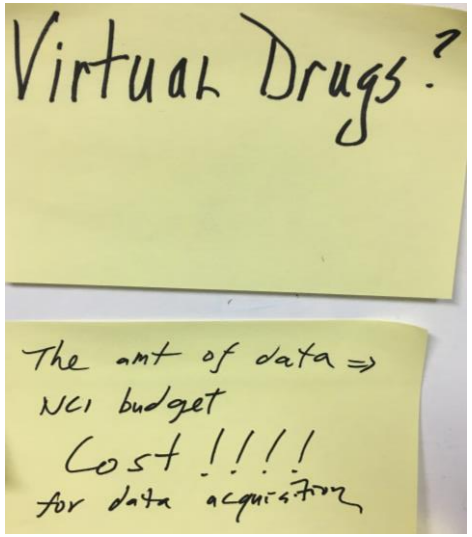
What are the inputs for predictive models? How do we represent biological knowledge in a computational form?

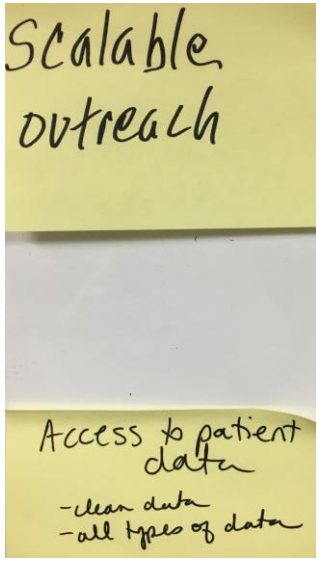
Improve "black box" of UQ & Deep learning For both computational & cancer scientists

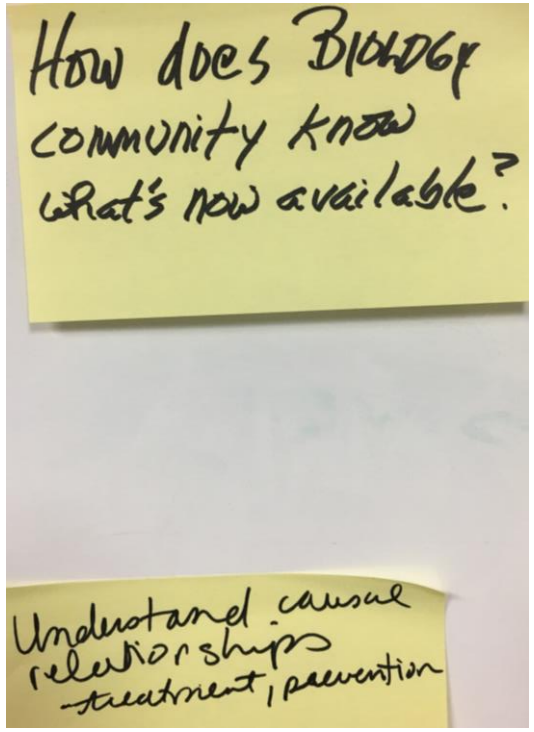
New Board

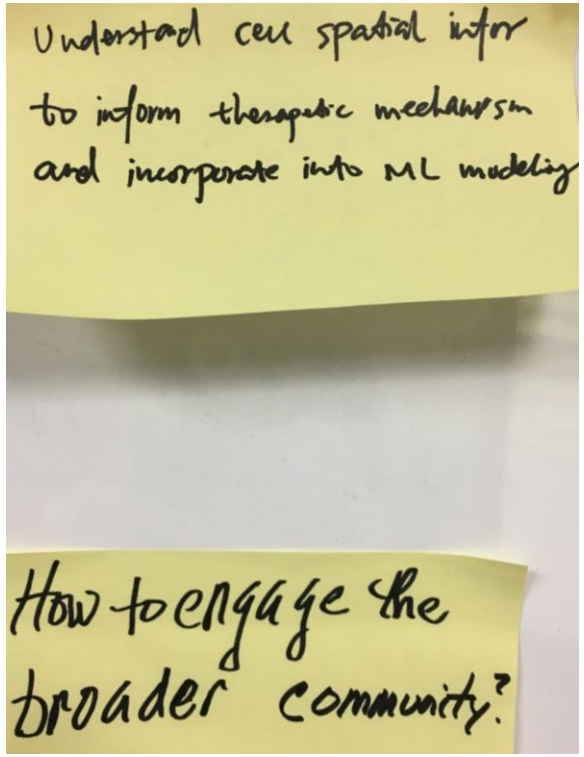

Physics-based modeling
Computational algorithms to detect meaningful improvement in clinical trials ~~& clinical care~~ at individual level,
How to obtain data → detect change →···> eventually help decision making
Establishing costs for machines + analysis when they are constantly evolving
RFPs for CANDLE Resources
Standardize & digest data that informs health status
Better Connectivity betwn those who have outgrown Beowulf and don't know how to reach out to DOE.
What data do we need to gather from imaging to integrate into the models? how do we do this over time?
Lack of systematic app towards science
Efforts to Establish Tagged reference data sets (in multiple fields)
1. CAcu of labelled data
2. Data Access challenges — IRB
3. Data cleaning/ wrangling


How to apply tools in a real clinical world? The model's interpretation is different from clinicali
Bridge the gap. Bring clinicans in an early stage to define problems in clinical practice


Virtual Drugs?
The amt of data ⇒ NCI budget
Cost!!!!
for data acquisition


Scalable outreach
Access to patient data
-clean data
-all types of data


How does Biology community know what's now available?
Understand causal relationships treatment, prevention


Understand cell spatial infor to inform therapeutic mechanism and incorporate into ML modeling
How to engage the broader community?

3D mechanistic model of complex tumor.

How does Cancer community know Summit & tools are now available (as of Jan 2018)?

Focus on optimal experiment design → experimental data → model → optimize → disseminate info.
Co-design - how do we push compute hardware into current architecture

Confidence intervals / uncertainty testing

-Access to data/Standards
-Push more in community, access dataset, data sharing
- In medical/clinical research communities

Internet 2 – VPNs for health
Common IRB
Break regulatory boundaries

Lead-in Challenge:
How do we perform clinical tests and use models to predict outcomes/solutions earlier

How do we integrate modeling and predictive design into clinical workflow?

Collect, right information
How to use that information to feed into models and validate

Validation of comp models
- survival model
- factors on model to understand
- ~~understand causal~~

Data integration
- surveillance
- molecular
- image (path)

Improving clinical structured data collection...
culture shift

How to target all of the data integration to a specific Prediction?

What is the target?

Predicting patient response using better genomic data models

Defining higher level treatment strategies

Being able to understand AI or data driven models

Interpretation of genomic variants of uncertain significance

Data abstraction from clinical healthcare records

Being able to interrupt mechanisms reliably ↑ disease

→ Non-invasive biomedical device technical
→ identify effective biomarker

Reliability of prediction model ⟺ precision medicine
contrast

During the survivorship phase for patients, what info can we combine about patient experience (symptoms, impacts, quality of life) and other clinical data (labs, images, genome) to inform care and patient expectations?

High content, High Quality affordable
scale data frequency
cost/data point => too high

Are there opportunities to leverage DOE monitoring and imaging devices, and computational capabilities to supply clinicians with information about patient outcomes?

① Data Integration from multiple domains (social media, image, EHR, NGS)
② How to bring different data together to make decisions

To identify risk population, to scan more frequency.

Dynamic treatment adjustment.

New data types—
in-body / ingestible sensors:
nutrition
physical activity
medical measurements

Population-wide data collection

Flops are often not a meaningful computing power metric. What are good metrics in this space?

Integration of Mechanistic + machine learning models

Shift to collaborative model from competitive model

Lead-in Challenge:
How do you integrate multi-modal data sets?

Be able to predict patient x response from a to be defined set of patient data

Design neuromorphic or custom hardware.

LOTS of DATA...
...HOW TO DECIDE?

Exposome: nutrition, environmental exposures, behaviors

How do we integrate temporal data to across time frames to identify key leverage points.

Data Collection
wearables that detect change from baseline

ethical considerations
- use data w/o misusing it

multivariate time series data for the whole population to avoid biases

Utilizing data already gathered that may have an effect:
Heart rate, breathing circadian rhythms

Need integration of multiscale ~~integration of~~ model, data, and expertise from cellular to tissue and population levels allowing more realistic prediction.

Sharing of data across different org's

beating Moore's law - next-gen computing

Reproducibly of Comp. Model

Consolidate drug response data or generate new data in an annotated, standardized way.

Looking at relationships within trends to look more holistic approach over time

accessibility to data
legal / HIPAA compliance

Can we prioritize which features are predictive?
- which groups of mutations are the important ones?

answers to questions are in the data
- need for laws similar to Newton's laws for biology

cellular computing - huge scale - reassemble

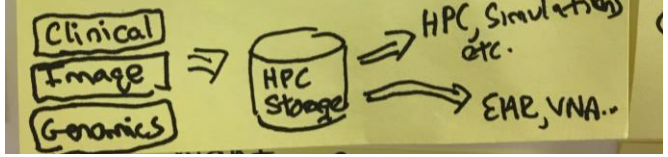
Data curation:
Standardized longitudinal collection of data

High Energy Phys
Promote HEP collaborative model to cancer community

Turning "real data" into computationally actionable data

How to maximize utility of clinical data?

We need to re-architect how we capture, store and access patient data so it can be easily available for research efforts.



Diffusion equations solved by hardware – prototype

How might cancer cells/ tumors inspire computing capabilities?

new non-Von Neumann architectures

generalization and application of stat analytic models between domain focused predictors
- need domain expertise + shared methods

a.) How do we interpret models/predictive models.
⇒ Why does it work
⇒ How do we intervene

1) Addressing gaps in Registries
- Missing longitudinal data
- Missing quantitative test results.

Predictive modeling w/patient trajectories

Democratize access to protect data to compute resources

Make tools from JDACS4C used more broadly

Enabling students to be effective in data science without programming

Want a
Software package that helps you design drugs or small molecules. Preferably starting with a broad, non-target specific base, but that can later be target-specific.

Overcoming steep learning curve for data science to use tools well

Engage extramural community to resources like CANDLE

Joining millions of individual projects data across modalities

Institutional issues
HIPAA, data use, etc.
How to ingest data?

Moving beyond genomics causal studies to other modalities of data

Cancer prevention:
baseline description of health and definition of cancer-related deviation

Using imaging data (path + CT), genomics and multiscale data effectively to predict well

Multi-modal approaches to address cancer-
Need datasets for multi-modal analysis

Greater frequency of data points, higher content

Morphological + gene expression changes

To be able cancer in situ in a dynamic way, ~~in~~ on time ~~the~~ for ~~the~~ treatment

Content-based search & selection for cellular & other imaging - organizati

UNIFIED COMPUTATIONAL MODEL W/ SME
Clinical
Molecular
Environmental
For Decision Support (Diagnosis Prognosis)

More resources in
1) computation
2) software tools
3) People

① Knowing what data elements to use; AND (data, features) EHR, social media, genomics

Semantic integration of different datasets (even the same variable in 2 ≠ datasets may have slightly ≠ meaning!)

Can we use computing to guide developing a new experimental technique that could give us information we need?

logistical issue: NCI have high computing, enable it

Predict cancer risk based on genetic + molecular + environment at birth (young age)

Predictive Model trustworthy?
- how to make sure
- uncertainties
- Generalization capability

New Board

In 10-15 yrs. —
- Virtual epidemiology model
- Series of models that inform one another

In 10-15yrs. — Want personalized patient specific modeling — make "patient relevant" specific predictions

In 10-15 yrs. -
Move beyond snapshot of patient trajectory
↑ towards
now
today
tomorrow

In 10-15 yrs - Desire for Systems of Systems Optimization of prevention of health (model & optimize)

In 10-15 yrs. - Virtual models of the patient, spanning the spatial & temporal scales

How to integrate different types of data

In 10-15 yrs. - Can we scale to timescale of patient w/ reaurrence

Team Science needs not just experts in different domains...
... but also people who are not experts, rather they have a good understanding in several domains.
These people are "Liaison" people.

DERIVING BIOMARKERS FOR RESPONSE TO RADIATION
- LONGITUDINAL IMAGING
- LINK TO GENOMIC
- NEED LARGE SAMPLE SIZES

universal data - IoT
- lots of info

precision medicine - nutrition, antibiotics, cancer preventative therapy @ individual level

How does one provide basis of model for regulatory work.

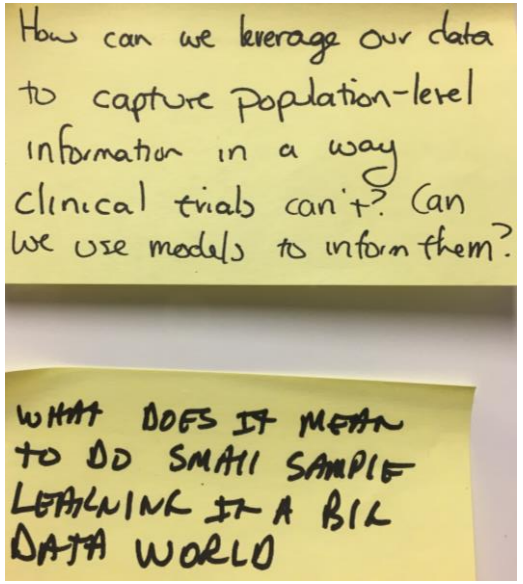
biological computing - cells compute for us - increases scale

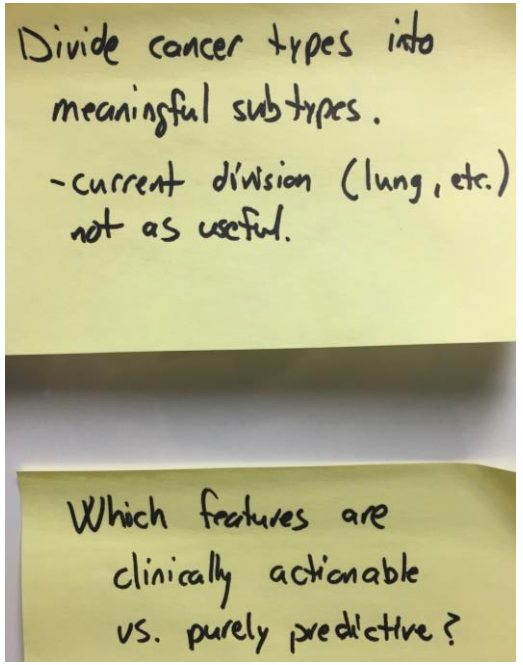
Model validation to actually believe ~~Models~~ to be "true"

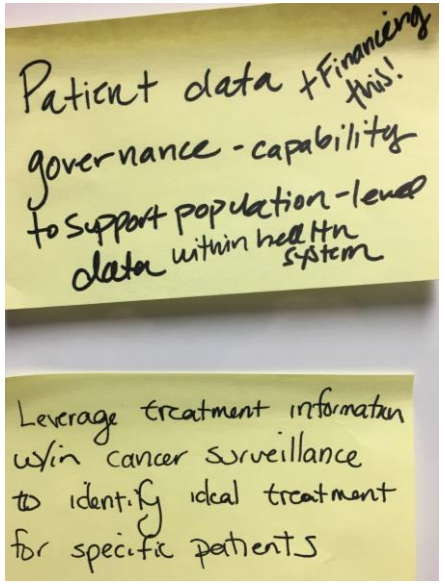
Science needs to be in streamline fashion.

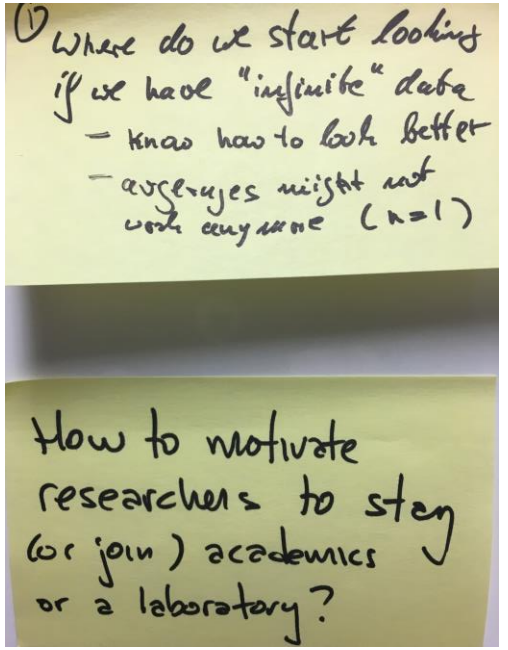
Co-design of hardware, software & algorithms to solve bio problems. Could you design custom hardware for the bio you are doing?

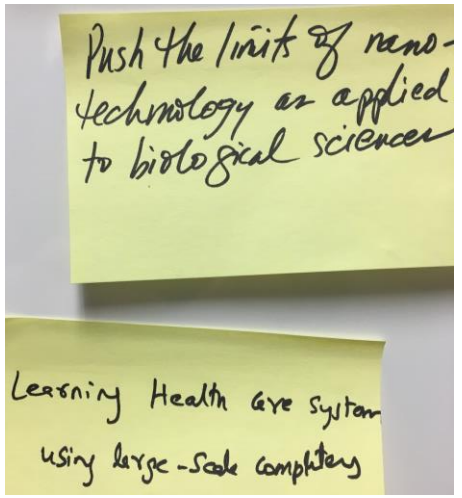
Need a dynamic model of the human cell that accounts for activity of all genes.


How can we leverage our data to capture population-level information in a way clinical trials can't? Can we use models to inform them?
WHAT DOES IT MEAN TO DO SMALL SAMPLE LEARNING IN A BIG DATA WORLD


Divide cancer types into meaningful subtypes.
- current division (lung, etc.) not as useful.
Which features are clinically actionable vs. purely predictive?


Patient data + governance - capability to support population-level data within health system
Financing this!
Leverage treatment information w/in cancer surveillance to identify ideal treatment for specific patients


Where do we start looking if we have "infinite" data
- know how to look better
- averages might not work anymore (n=1)
How to motivate researchers to stay (or join) academics or a laboratory?


Push the limits of nano-technology as applied to biological sciences
Learning Health Care system using large-scale computers

Cancer prevention:
Accurate detection

Bridging the Gaps between experimentist and CS?
1) reducing ~~effo~~ duplication of efforts
2) enabling knowledge sharing.

funding collaborative efforts

Personalization of treatment in Rare Disease

Response models
Useful predictions using existing data (prostate cancer use case)

How can we predict the data we need to solve a problem?

How do you solve a problem w/ data that may not contain the solution?

Figure out tech that will help you get the data you actually want

high impact publications to help drive culture shifts/ embracing new resources in cancer community

3yrs later people are still excited...
lofty goals but ~~main~~ maintain momentum

Administrative Issues in accessing data

Collaboration
- Establish better communication mechanisms
- Matchmaking across multiple areas in Gout.
- Leverage NLM

Benchmarking

trust / model interpretability - how to build for clinical community, so they will use AI?

IoT - ideas identifying the signal

quantitative imaging - info in images (that we may or may not be aware of)

imaging informing other domains (eg ~~non~~ mechanistic)

multi scale physics needed
→ human cell atlas
human tumor atlas

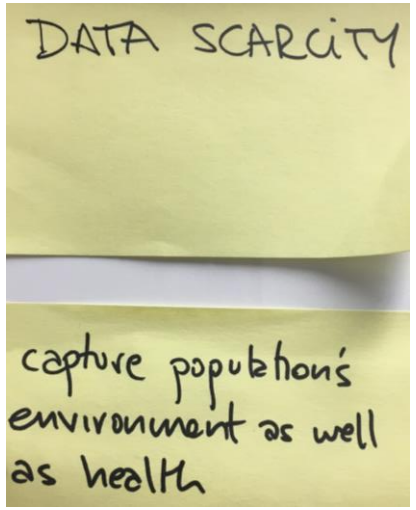
Data integration is a challenge

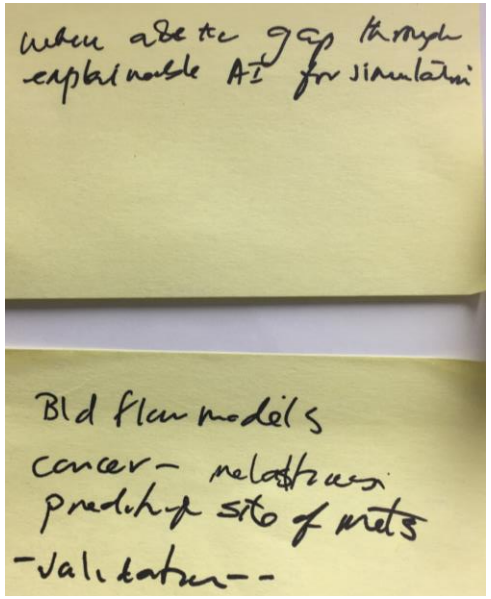
- We need new types of data.
- Need to record quantitative observations of behaviors.

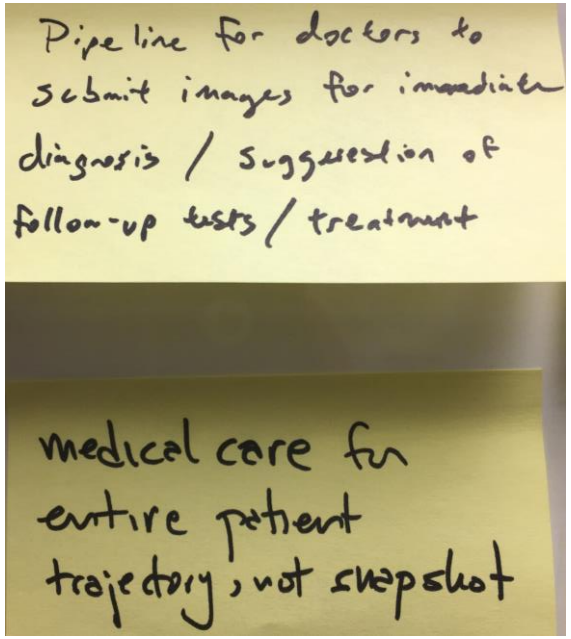
more forward from predictive to prescriptive
- why can't we ~~have~~ a model that is!

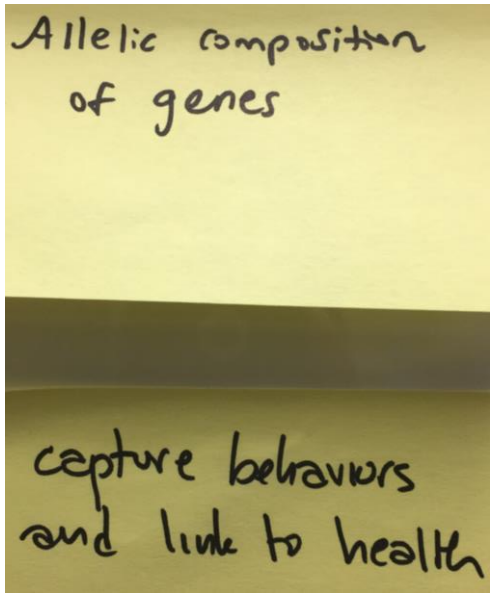
Can we frame a bio. problem similarly to AlphaGo, with similar success? (and be robust)

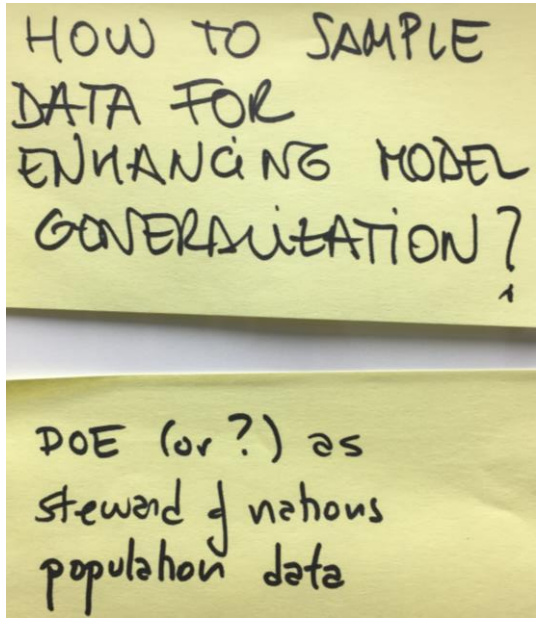
New Board

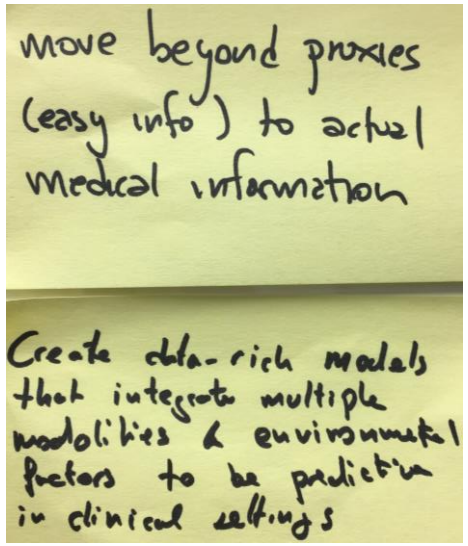

DATA SCARCITY
capture population's environment as well as health


Bld flow models
cancer- metastases
predicting site of mets
-Validation--


Pipeline for doctors to submit images for immediate diagnosis / suggestion of follow-up tests / treatment
medical care for entire patient trajectory, not snapshot


Allelic composition of genes
capture behaviors and link to health


HOW TO SAMPLE DATA FOR ENHANCING MODEL GENERALIZATION?
DOE (or?) as steward of nations population data


move beyond proxies (easy info) to actual medical information
Create data-rich models that integrate multiple modalities & environmental factors to be predictive in clinical settings

- PR effort to encourage people to share data

A.I. pipeline for IDing cancer/tumors in images from patients

Single Cell Gene Expression profiles for all cell types in human body (37 BIL cells) Capture healthy & disease states

What are possibilities w/ cancer image data, real-time processing, time-domain imaging ~~&~~ temporal imaging?

Experiment Workflow languages - DAGs

training in use of new tools and resources to promote dissemination

Incentivize Data Sharing
Use Carrots NOT Sticks

AI for clinical problems that matter (eg. usefulness of predictive models)

explort information available from wide populations

application/domain science areas to adopt new ideas

⑥ Data is not suff. for minorities or other "special" groups
- Lots of ind. variability

Ways to extrapolate from limited/incomplete data

temporal + spatial
imaging - new analytic techniques for data viz
⇒ building databases + atlases

why can't we have real-time ML?

CLASSIFY THE CHAOS OF ANEUPLOIDY AND ALLOW IT TO CONTRIBUTE TO MODELLING.

how can we develop tools that can actually be used in real-world settings? not just theoretical models

Q: What does it take to trust a machine learning output to make a prescriptive decision?

Can "aggressiveness" of a tumor be modeled, ~~instead~~ in addition to just "staging"?

Sea of information ! islands of knowledge - great description of cancer science. How do we build predictive models that can handle that uncertainty ! Incorporate the lack of knowledge?

Data sharing

Integrate different data type, using data already out there

UNDERSTANDING & MEASURING TUMOUR HETERO..?

Detect skin cancer from Facebook/Social media images

KEDAR: powerful imaging technique. molecular
~~show~~ support tumor disease progression.
prediction supplementary to genomic/proteomic data.

Access to a transparent "box" of data to be able to understand what's happening

Manual annotation

Amount of time it takes to label data for DL (e.g. for recurrence)

Full body & cell resc/ chall

uncertainty quantification

dynamism - systems epidemiology

better entangling ~~the~~ not a lot of data genomic features, tumor & drug features is expensive + difficult

Science of ~~data~~ science toolkit

HOW TO GENERATE ENOUGH DATA FOR MODEL DEVEL.

measure what you can & model the rest.

Building trust in models is key.

Data Integration

Application of analysis methods from other fields (physics) to biological and clinical data

BIOLOGISTS NEED A LOT MORE HAND HOLDING THAN IS BEING ASSUMED!

* NEED MORE DATA SCIENTISTS "CLOSER TO THE GROUND" & "101 HANDBOOKS".

Rapid approximations of solutions versus precision which may take longer

- Commonly labelled data sets available to the community

Heisenberg compensator Adaptive Multiphysics Model for tumor metastasis

- Demonstration of computational model and validation thereof are required for new

How to extrapolate from limited/incomplete data.

MOVING BEYOND H&E STAINING!
Difficulty in extracting useful imaging features. Real-time imaging and disease diagnosis's and prognosis prediction.

Validation from independent groups

New Board

Large Insurance Payors (United, Anthem, Humana, etc.) will provide all data under a BAA (Biz. Assoc. Agreement) for specific purposes (research)

Our national labs can have access to a lot of data!

HHS has an initiative to make patients own their own data. In addition, HHS is imposing fines on EHS vendors blocking patient access to their data.

We can look forward to patients contributing their data, like All of Us.

generic analysis of spatial dynamic system

Dynamical models of tumors @ cellular + subcellular resolution

patient scale predictive models spanning molecular to full body scales

accelerated DNN accelerated systems of ODEs

What are Virtual drugs?

Patients trust that their privacy and best-interest are protected & embraced across the ecosystem

IP concerns often create silos

SEMI-SUPERVISED MODELS TO HELP PREDICTION RARE (R) DISEASE

1- Code acceleration for agent based model that include cell specific gene expression evolution.

Develop biological models for mechanistic understanding of cell. structure-function relationship

How can we get feedback on how codes/products/APIs are used by the community and identify where they see gaps? Can we identify new sources of data to incorporate?

Accelerate Path from research findings to pt impact (ie case providers)

How can pts Monetize their own data?

What strategies can we use to engage scientists, clinicians, patients and others, either to contribute to research or to use end products?
- Engagement, ux as a sociological research question

Co-develop a Pipeline to better understand cancer & progression of the disease

tumor - metastasis
- model transition
- ID therapeutic interventions

EXPLOITING ML TO UNDERSTAND BIOMECHANICS OF CANCER CELL INTERACTIONS & TUMOR FORMATION

Can computing help determine who is missing from a clinical cohort?

Use HPC to select a new (or the next) cohort for clinical trials

MEDICAL
GATHERING DATA FROM BIRTH TO DEVELOP DIGITAL TWINS FOR HUMANS

MACHINE LEARNING THAT CAN IDENTIFY INTERACTIONS

Impact Patient centered outcomes
Integrated Data across scales Opt Tech Design

Disease evolution
Understanding tipping point from "non-cancer" to cancer

Cancer is ~~the~~ a disease of the tissue

Clinician
holistic Advisor at point of care

Non invasive dynamic Imaging - RFID cells

Radiogenomics at the pathway level (metabolism)

Use immediate patient-provided data to fill in full + predictive patient trajectory

How do we target a model for a treatment once all the pieces are pulled together?

Create virtual models of patients to test treatment plans on

A system of all systems to model + optimize for patient outcomes

Patient-specific cancer modeling that makes relevant predictions for better outcomes

Scale predictive modeling from atomistic to patient outcome level

What can we do at the intersection of real-world, noisy, empirical data and mechanistic models?

Can you provide an immediate risk value to a clinician for a patient for a given treatment?

Are there sources of information we don't currently think of as "data" that could be used in research? How do we turn them into "data"? What makes something data?

Create models that could explore new questions, eg. "what happens when I block this protein?"

Determine drug targets (competing) from advanced models

Assume all powerful models will make exploration (interactively) and interpretability paramount

How can treatments for + knowledge of common cancers inform treatments for + knowledge of rare cancers?

Can you recognize the perfect treatment when it has been created + roll it out immediately?

Explore USING ML To develop Effective
1) DATA MODELS + classification systems for Cancer Data at Multi-Scale.
2) Develop data sharing Paradigms

What biases could we unintentionally be introducing into our machine learning algorithms? How do we recognize them? what impact could they have on patient health outcomes?

How can HPC be used to connect biological + social determinants of health and reduce health disparities?

Couple ML & mechanistic models to improve interpretability

Build individual patient models first, rather than apply a population scale model to a patient

Understanding and Modeling the cellular (and other) components and interactions in the tumor microenvironment.

Data privacy- Synthetic data development

How to understand and model the whole causal/ Mechanistic system that is generating and sustaining Cancer.

Develop methods to learn patient-specific causal models (e.g. graph models)

Perform Active learning per patient to personalize treatments and drive further experiments

Problem #1.

Personalized Drug Design ..

Combo therapy

Problem #2 !

Active Learning on Patient → Drug

Problem #3

↙ AL

Deep omics characterization of each Patient.

Problem #4

Long tail of Cancers.. exploring MODELS ↓

Problem #5 AI

Adaptive Search

Self-aware SEARCH meta Search.

How to integrate patient Reported Outcomes into predictive models?

Continuous Improving Patient Specific TUMOR MODELS.

Why do we need medical privacy if none of us retain it?

What are the potential applications from Peter's pipelines to biological imaging?

Most challenging computational problems are in combinatorial optimization. Simulations are "easy".

How do you retrospectively address the limited nature of patient consents to broaden the ability to share tissue left-over biospecimens?

Privacy issues are being addressed by the iDash community. Scalable solutions are being developed.

Can we predict how the mutational landscape of a tumor tissue react to treatment.

How can we mine data from health records of people not on trials? Can consents be developed?

DATA INFRASTRUCTURE

INTEGRATION (MULTI MODAL)

CURATE. COHORT SPECIFIC

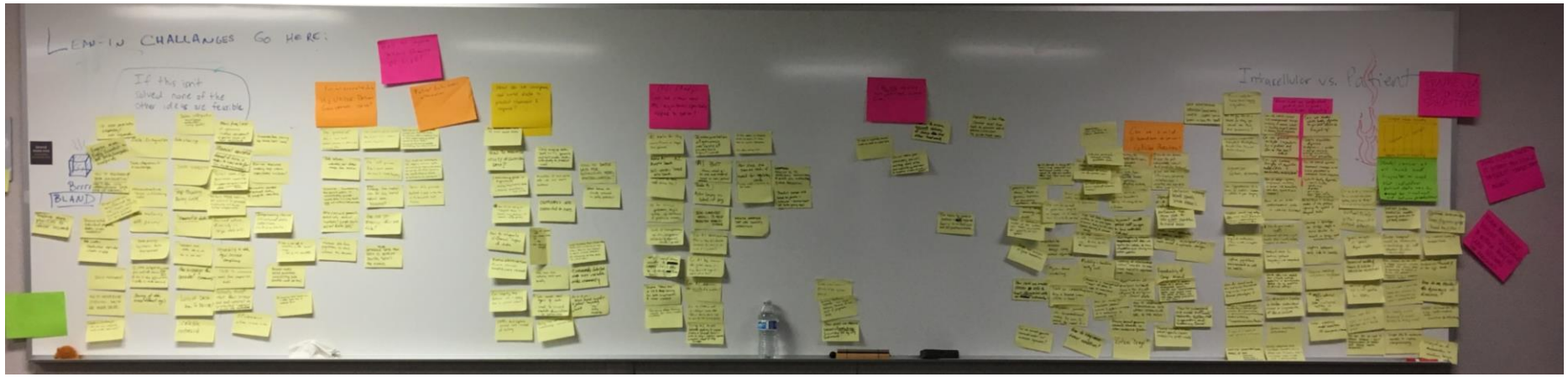
Queries → FEED ML Beast.

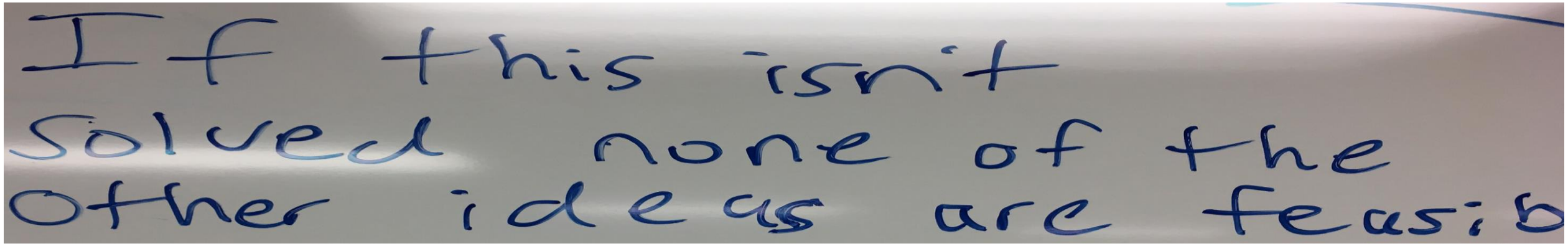
Omics, cellular, tissue level imaging integration

IN-SILICO MODELS TO TRANSLATE IN-VIVO EXP. RESULTS TO HUMAN PRIOR TO CLINICAL APPLICATIONS

Thursday, March 7, 2018, Lunch

March 7	
8:00 am – 8:45 am	**ARRIVAL AND CHECK-IN**
8:45 am – 9:00 am	**RECAP DAY 1 | OVERNIGHT IDEAS**
9:00 am – 10:00 am	**KEYNOTE PRESENTATION** **Blue Sky Possibilities at the Intersection of Oncology and Computing** Warren Kibbe, *Duke University*
10:00 am – 11:00 am	**BREAK OUT GROUPS** - Generating ideas for lean-in challenge areas (part 4)
11:00 am – 12:00 pm	**SYNTHESIS OF LEAN-IN CHALLENGE AREAS**
12:00 pm – 1:00 pm	**LUNCH**
1:00 pm – 1:30 pm	**PRIORITIZATION OF LEAN-IN CHALLENGE AREAS & WRITING GROUP FORMATION**





Data Integration

Data sharing

Data integration is a challenge

ID 1000 population responder/ non-responder

DATA SCARCITY

Incentivize Data Sharing
Use Carrots NOT Sticks

Data integration
- surveillance
- molecular
- image (path)

How can predictive modeling help inform experimental validation?

Patient data + finance governance - capability to support population-level data within health system

③ Knowing what biases are in data and how to address them

Administrative issues in accessing data

IP concerns often create silos

Improving clinical structured data collection... culture shift

Manual annotation
Amount of time it takes to label data for DL (e.g. for recurrence)

Engage more Young Scientist to get fresh/energetic viewpoints

Shift to collaborative model from competitive model

Democratize access to patient data to compute resources

Accessibility to data legal/HIPAA compliance

Involving more physicists to cancer research

⑦ Data is not suff. for minorities or other "special" groups
- Lots of ind. variability

Response models
Useful predictions using existing data (prostate cancer use case)

Joining million individual projects data across modalities

Data curation:
Standardized longitudinal collection of data

If this isn't solved none of the other ideas are feasible

Stop throwing away data

Perform Active learning per patient to personalize treatments and drive further experiments

Patients trust that their privacy and best-interest are protected & embraced across the ecosystem.

① Data Integration from multiple domains (image, EHR, NGS, social media)
② How to bring different data together to make decisions

Institutional issues
HIPAA, data use, etc.
How to ingest data?

Data privacy-
Synthetic data development

Ethnic freq / dist of germline
→ other components of genetic make-up that contribute

HOW TO STANDARDIZE DATA GENERATION?
HOW TO DEAL WITH HETEROGENEOUS DATA (USING DIF. STANDARDS)?

How to engage the broader community?

Scalable outreach

HOW TO GENERATE ENOUGH DATA FOR MODEL DEVEL.

Data enclaves
diff. privacy

Privacy issues are being addressed by the Dash community. Scalable solutions are being developed.

HOW TO GENERATE ENOUGH DATA FOR MODEL DEVEL.

Data privacy-
Synthetic data development

Accessibility to data
legal / HIPAA compliance

Institutional issues
HIPAA, data use, etc.
How to ingest data?

Sharing of data across different org's

Lead-in Challenge:
How do you integrate multi-modal data sets?

HOW TO GENERATE ENOUGH DATA FOR MODEL DEVEL.

Democratize data

LOTS OF DATA.
...HOW TO DECIDE

DATA MANAGEMENT

Patient generated data

H2 Utilze Patient Generated Data?

How to improve Patient Quality of Life?

How can pts Monetize their own data?

How can we generate statistically identical "synthetic" Large-scale patient data sets?

Use "quantified self" data to track health + predict outcomes + personalize treatment plan

HHS has an initiative to make patients own their own data. In addition, HHS is imposing fines on EHS vendors blocking patient access to their data. We can look forward to patients contributing their data, like All of Us.

Use immediate patient-provided data to fill in full + predictive patient trajectory

Use "self-quantized" data to provide immediate intervention for health trajectory

How might we interrogate the adaptive immune system of individuals (ongoing, min. cursus) to monitor its data regarding the health/disease state to initiate early intervention?

Challenge: Incentivizing the general public to consistently provide health data in a way that's easy and uniform/interoperable

MEDICAL
GATHERING DATA FROM BIRTH TO DEVELOP DIGITAL TWINS FOR HUMANS

Data Collection
wearables that detect change from baseline

unbiased data from population to drive unbiased ML decisions

Exchange free medical care for fully shared medical data.
SERIOUSLY!!!

Deliver daily, personalized Feedback to people to keep them on the best possible health trajectory.

How do we incorporate real world data to predict treatment & response?

Commonly labelled data sets available to the community

EMRs that capture clinical care instead of billing

Large Insurance Payors (United, Anthem, Humana, etc.) will provide all data under a BAA (Biz. Assoc. Agreement) for specific purposes (research)
Our national labs can have access to a lot of data!

1) Addressing gaps in Registries
- Missing longitudinal data
- Missing quantitative test results.

How to MAXIMIZE UTILITY of CLINICAL DATA?

1. LACK of labelled data
2. Data Access challenges
   - IRB
3. Data cleaning/ wrangling

OUTCOMES are connected to care

In 10-15 yrs. -
Move beyond snapshot of patient trajectory
towards ↑ now today tomorrow

- We need new types of data.
- Need to record annotate observations of behaviors.

HOW TO SAMPLE DATA FOR ENHANCING MODEL GENERALIZATION?

translation of real world data into real world evidence

Using imaging data (path + CT), genomics and multiscale data effectively to predict well

How do we integrate temporal data to across time frames to identify key leverage points.

Which features are clinically actionable vs. purely predictive?

How to integrate different types of data

Can computing help determine who is missing from a clinical cohort?

Make tools from JDACS4C used more broadly

Turning "real data" into computationally actionable data

How to make best use of real world data

Data abstraction from clinical healthcare records

ML / Analysis

Can we invent new ML algorithms specifically applied to cancer?

why can't we have real-time ML?

SEMI-SUPERVISED MODELS TO HELP PREDICTION RARE(R) DISEASE

Predict cancer risk based on genetic + molecular + environment at birth (young age)

Can we create a model of a biological system that tell us: what new data do I need to improve the model (quantity + quality)

lack of transparency of ML/algorithms outside of clinical trials (for data)

EXPLOITING ML TO UNDERSTAND BIOMECHANICS OF CANCER CELL INTERACTIONS & TUMOR FORMATION

MACHINE LEARNING THAT CAN IDENTIFY INTERACTIONS

How does one provide basis of model for regulatory work.

↑ in silico models & standards for those models?

A.I. pipeline for IDing cancer/tumors in images from patients

Interpretation of genomic variants of uncertain significance

apply "causal inference" algorithm to [redacted] complex clinical data to sub-classify [redacted] tumor type

Using AI to ask patients questions (in clinical practice or clinical trials) in order to obtain patient-reported outcomes instead of long surveys.

Improve "black box" of UQ & Deep learning for both computational & cancer scientists

Problem #2 !

Active Learning on Patient → proj

Can A.I. direct experiments that can test A.I. identified correlations for which are the cause of a condition?

Problem #5 AI

ADAPTIVE SEARCH

SELF-AWARE SEARCH

meta search.

Can A.I. determine the causality of a condition?

IN-SILICO MODELS TO TRANSLATE IN-VIVO EXP. RESULTS TO HUMAN PRIOR TO CLINICAL APPLICATIONS

* FYI, ATOM is working on this

Reasoning deep learning networks for more complex modeling

AI BOT

Know what q's to ask and respond based on indiv patient

AI for clinical problems that matter (eg. usefulness of predictive models)

Can we develop cancer new sensors across Scale?

Patients Like Me
· Develop real-time tools to provide info to patient / clinician to guide treat choice

learning healthcare system - real time evolving standard of care, built in pragmatic trials

In body or ingestible sensors could be used to collect data

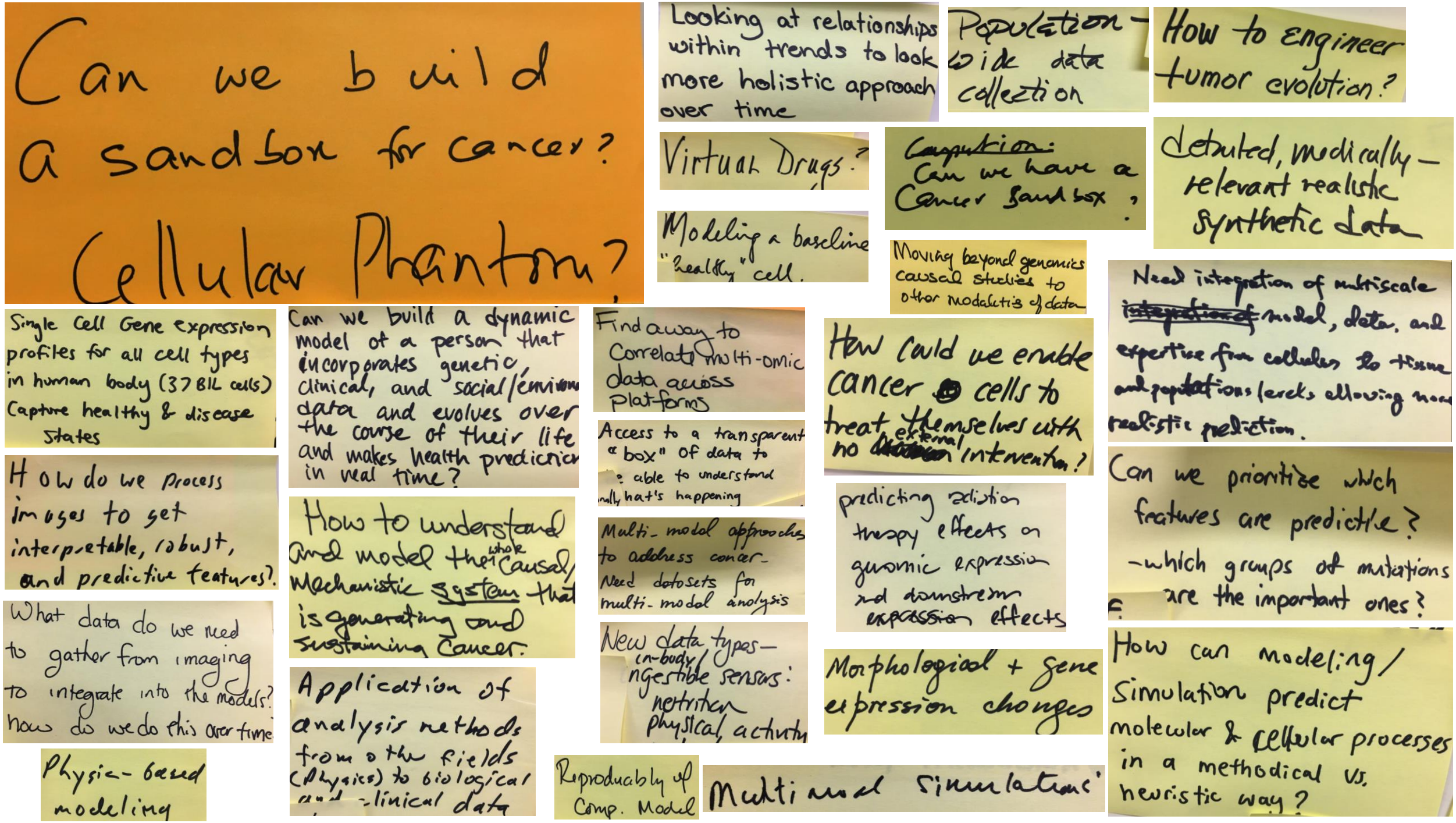
Create virtual models of Patients to test treatment plans on

How might we measure Patient outcomes (pain, fatigue, sleep) during survivorship NOT by using Questionnaires?

Can we measure pain (without asking patient) and stay ahead of the pain administering meds based on trajectory of pain for indiv. Patients?

Robotics to ensure accurate delivery of cancer [redacted] and other treatments.

Be able to predict patient x [redacted] txt response from a to be defined set of patient data


Can we build a sandbox for cancer?
Cellular Phantom?
Looking at relationships within trends to look more holistic approach over time
Population-wide data collection
How to engineer tumor evolution?
Virtual Drugs?
Computation: Can we have a Cancer Sandbox?
detailed, medically-relevant realistic synthetic data
Modeling a baseline "healthy" cell.
Moving beyond genomics causal studies to other modalities of data
Need integration of multiscale model, data, and expertise from cellular to tissue and population levels allowing more realistic prediction.
Single Cell Gene expression profiles for all cell types in human body (37 BIL cells) Capture healthy & disease states
Can we build a dynamic model of a person that incorporates genetic, clinical, and social/environ data and evolves over the course of their life and makes health predictio in real time?
Find away to Correlate multi-omic data across platforms
How could we enable cancer cells to treat themselves with no external intervention?
Access to a transparent "box" of data to able to understand what's happening
How do we process images to get interpretable, robust, and predictive features?
Multi-modal approaches to address cancer- Need datasets for multi-modal analysis
predicting radiation therapy effects on genomic expression and downstream expression effects
Can we prioritize which features are predictive?
-which groups of mutations are the important ones?
How to understand and model the whole Causal, Mechanistic system that is generating and sustaining Cancer.
What data do we need to gather from imaging to integrate into the models? how do we do this over time
New data types- in-body/ ingestible sensors: nutrition physical activity
Morphological + Gene expression changes
How can modeling/ Simulation predict molecular & cellular processes in a methodical vs. heuristic way?
Application of analysis methods from other fields (Physics) to biological and clinical data
Physic-based modeling
Reproducibly of Comp. Model
Multiscale Simulation

Can we build a sandbox for cancer?

Cellular Phantom?

Patient-specific cancer modeling that makes relevant predictions for better outcomes

Develop methods to learn patient-specific causal models (e.g. graph models)

To be able cancer in situ in a dynamic way, in time for treatment

DATA INFRASTRUCTURE INTEGRATION (MULTI MODAL). CURATE. COHORT SPECIFIC QUERIES → FEED ML BEAST.

HOW TO LEVERAGE EPIDEMIOLOGICAL DATA WITHOUT INTRODUCING BIAS IN THE MODEL?

Multivariate time series data for the whole population to avoid biases

Can you identify the important biomarkers to identify and subsequently treat a disease

How To:
Accelerate vaccine design using structural information

How To Develop a library of "Virtual" cells implemented with Agent-based models — and linked TO NCI-60 and/or other NCI Tissue banks (well annotated)

Can you identify the important biomarkers to identify and subsequently treat a disease

What are possibilities w/ cancer image data, real-time processing, time-domain imaging & temporal imaging?

Segment multi-spectral cell images to trace metabolic pathways

Hypothesis generation via computational models. Eg can I predict the impact of a new drug?

Could we computationally design a targeted immune cell for a tumor?

What are the structure/property relationships in oncology and how are they investigated?

Continuous improvement Patient Specific TUMOR MODELS

Can we connect genomic characterization w/ molecular dynamics?

How to include biological variability into mechanistic model.

Single cell analysis of gene expression

Need a dynamic model of the human cell that accounts for activity of all genes.

How could we understand, predict, and engineer cancer trajectory?

generic analysis of spatial dynamic system

Can we build a "virtual twin" for an individual that enables exploring options for therapy? (or project patient trajectories?)

Develop a paradigm to bridge length + time scales to model a single cell mechanistically

medical care for entire patient trajectory, not snapshot

Omics, cellular, tissue level imaging integration

Can we design dynamic, longitudinal monitoring of cancer risk biomarkers, both from liquid & solid biopsies?

Co-develop a pipeline to better understand cancer & progression of the disease

How can HPC be used to connect biological + social determinants of health and reduce health disparities?

- while population dynamics
  - ↳ dynamics of epidem.
  - ↳ merge socio
  - ↳ merge mechanistic
  - ↳ virtual prevention

In 10-15 yr. – Virtual models of the patient spanning the spatial & temporal scales

How do you build a model for thing you cannot see that one predictive?

How do we bridge spatiotemporal scales in cancer biology?

Can we predict optimal treatment strategies across diverse populations to understand health trajectories

DNN accelerated systems of ODEs

dynamism-systems epidemiology

How could we understand, predict, and engineer Cancer trajectory?

Virtual trials
Virtual prevention
Virtual education

Dynamic treatment adjustment.

Design a Computational model for a "Normal" Cell in terms of organization in space/time/compartmental reaction biochem pathways.

"Computational Cellular Phantom"

Whole-body dynamics
- immune function
- cardiovasc.
- digestive
- paracrine...
- plus tumor sites, dissemination

Can we model whole body dynamics to predict disease trajectory?

UNIFIED COMPUTATIONAL MODEL w/ SME
Clinical
Molecular
Environmental
For Decision Support (Diagnosis, prognosis)

Create data-rich models that integrate multiple modalities & environmental factors to be predictive in clinical settings

Can we predict multiple trajectories for a patient and choose the best one?

capture behaviors and link to health

How could we understand, predict, and engineer Cancer trajectory?

Predictive modeling w/ patient trajectories

Dynamical models of tumors @ cellular + subcellular resolution

Ways to extrapolate from limited / incomplete data

Can we inform cancer risk management through modeling of adverse events in healthy people & patients to develop health trajectory projections?

Heisenberg Compensator Adaptive Multiphysics Model for tumor metastasis

How do small scale mo molecular trajectories map onto large scale disease phenotypes?

A system of all systems to model + optimize for patient outcomes

~~KUDAR~~: powerful imaging technique ~~show~~ support tumor disease progression prediction supplementary to molecular genomic/proteomic data.

relating molecular scale (nutrients, dynamics, etc) → cellular scale → tumor scale → individual scale → pop scale → environmental scale

Next 5-10 years 20x to 100x speedup for ABMs

How could we understand, predict, and engineer cancer trajectory?

Can "aggressiveness" of a tumor be modeled, ~~instead~~ in addition to just "staging"?

Create models that could explore new questions, e.g. what happens when I block this protein?

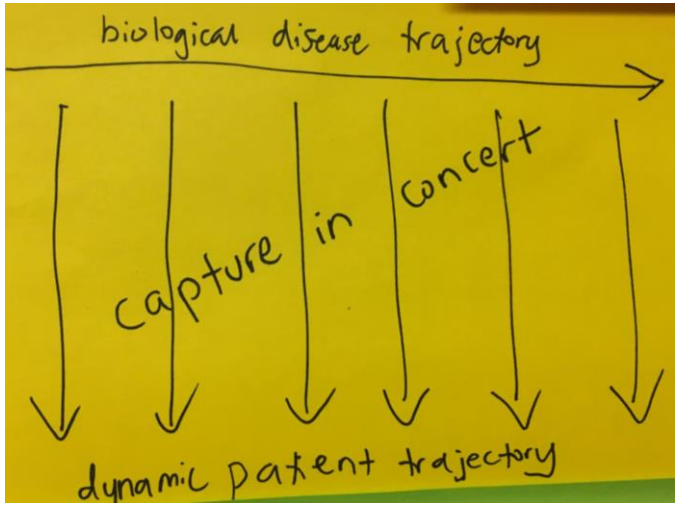
tumor - metastasis
- model transition
- ID therapeutic intervention

capture population's environment as well as health

Can we combine strengths of population models/data & mechanistic models/data?

How do you model health?

A diseased cell/body is just deviation from a healthy cell/body
Perspective of disease.



Couple ML & mechanistic models to improve interpretability

Can we create a personalized "monte carlo" simulation to explore how interventions can impact a patient's health?

Can computing automate the process of scientific discovery by identifying causality?

3D mechanistic model of complex tumor.

Model cancer at all levels and trajectories such that individual patient data can be used for treatment and outcome predictions

What strategies can we use to engage scientists, clinicians, patients and others, either to contribute to research or to use end products?
- Engagement, ux as a sociological research question

How do we model the dynamics of disease?

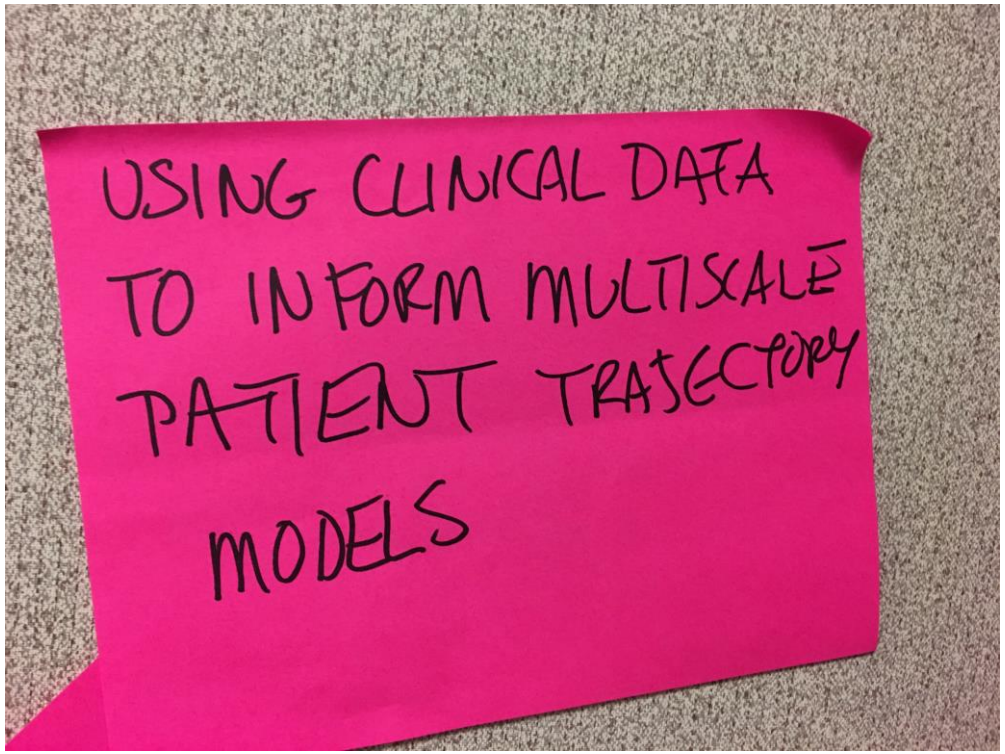
How to create mechanistic 3D models at the diff. levels?

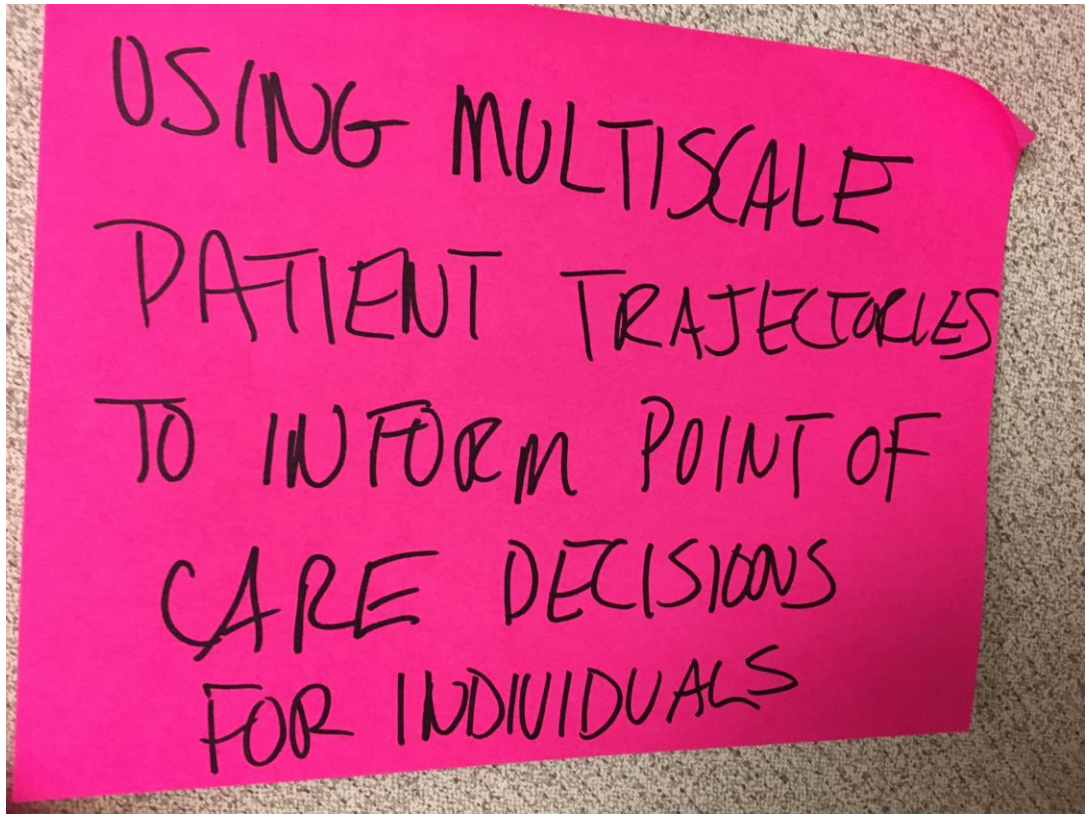
patient scale predictive models spanning molecular to full body scales

Develop biological models for mechanistic understanding of cell. structure-function relationship

Useful abstractions to bridge spatio-temporal scales.
prediction ± understanding

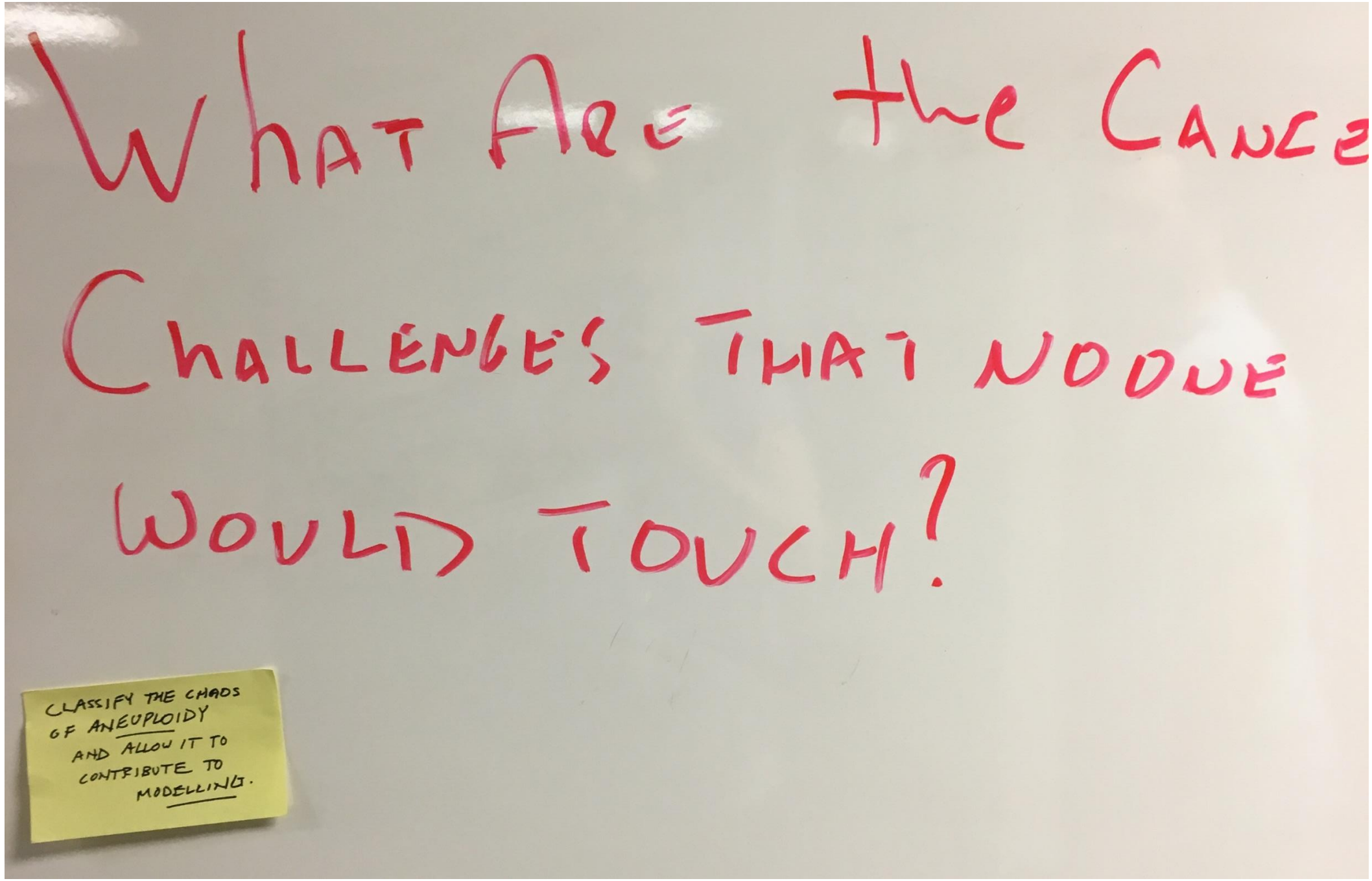
Integration of Mechanistic + Machine learning models


USING CLINICAL DATA
TO INFORM MULTISCALE
PATIENT TRAJECTORY
MODELS

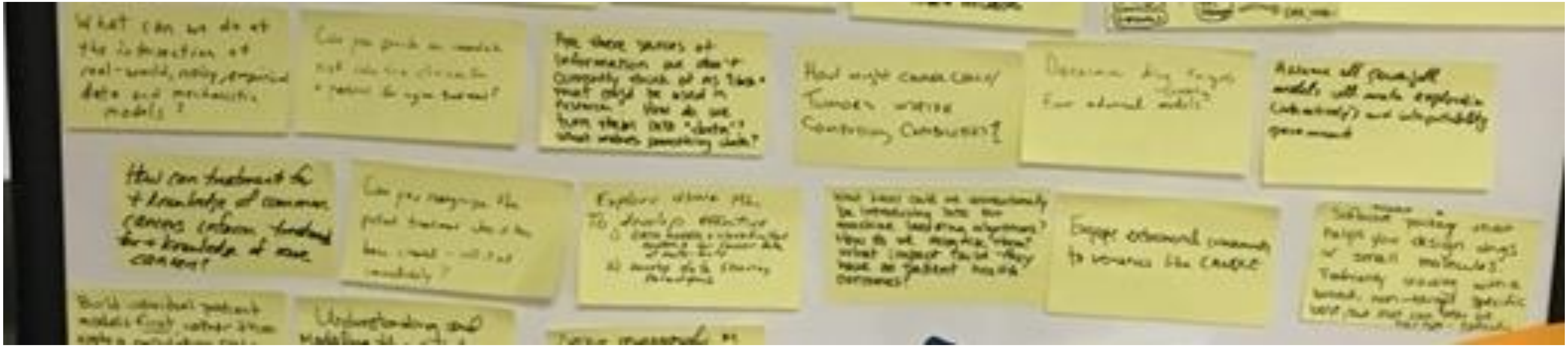

USING MULTISCALE
PATIENT TRAJECTORIES
TO INFORM POINT OF
CARE DECISIONS
FOR INDIVIDUALS

Thursday, March 7, 2018

Final Grouping of Post-Its


What Are the Cance
Challenges That NODDE
Would Touch?
CLASSIFY THE CHADS
OF ANEUPLOIDY
AND ALLOW IT TO
CONTRIBUTE TO
MODELLING.


What are virtual drugs?
Cancer is a disease of the tissue
Sabbaticals for computational people at biology institutions


What can we do at the intersection of real-world, noisy, empirical data and mechanistic models?
Engage external community to enhance the CANCER

Leverage treatment information w/in cancer surveillance to identify ideal treatment for specific patients

IoT - ideas identifying the signal

1- Code acceleration for agent based model that include cell specific gene expression evolution.

Anti-cancer adaptive small molecule synthesis or adaptive morphology adaptive vasculature? adaptive receptor inhibition?

What are Virtual drugs?

Exposome: nutrition, environmental exposures, behaviors

Where do we start looking if we have "infinite" data
- know how to look better
- averages might not work anymore (n=1)

more forward from predictive to prescriptive
- why can't we ~~have~~ a model that is!

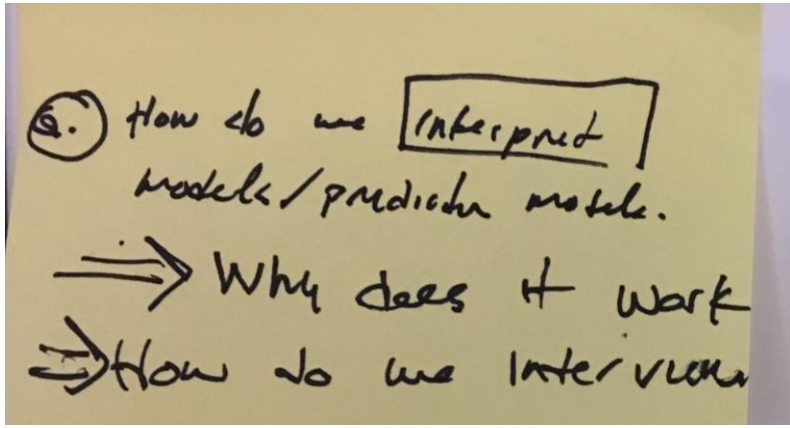
Cancer Challenge Areas (Cont.)

Understanding Physics of Disease

a) Develop integrated multi-scale predictive models of disease (molecular scale to systems biology scale). Can patient specific information then be used to create personalized predictive models?
   a) Deep tissue 3D optical imaging of live animals: https://www.biorxiv.org/content/early/2018/10/18/447433 - opportunity to map down to molecular level starting from whole animal
   b) Scaling up agent-based modeling (models individual cells as agents – position, velocity, phenotype): http://mathcancer.org/Projects.php
   c) Advancing cryo-EM algorithms

b) Develop close feedback loop between patient-care outcomes, continued model development, and directions for experimental basic science research. ("Learning Healthcare System")

Related NCI efforts: HTAN, CSBC, PSON



How can we get feedback on how codes/products/APIs are used by the community and identify where they see gaps? Can we identify new sources of data to incorporate?

Accelerate path from research findings to pt impact (ie case providers)

Diffusion equations solved by hardware – prototype

What are the next set of theories we need to investigate to give us rigorous bounds on the performance of machine learning & deep learning models?

Science needs to be in streamline fashion.

Combination of social & clinical psychology (behavior) & ML to understand & predict patient decisions re: cancer prevention practices, diagnosis, treatment.

Answers to questions are in the data – need for laws similar to Newton's laws for biology

Use HPC to select a new (or the next) cohort for clinical trials

3yrs later people are still excited...
lofty goals but maintain momentum

beating Moore's law – next-gen computing

Impact Patient centered outcomes

Integrated Data across scales

Opt Tech Design

Disease evolution

Understanding tipping point from "non-cancer" to cancer

Cancer is ~~the~~ a disease of the tissue

Holistic Clinical Advisor at point of care

Now from dynamic Imaging - RFID cells

new non-Von Neumann architectures

How do we target a model? For a treatment one all the pieces are pulled together?

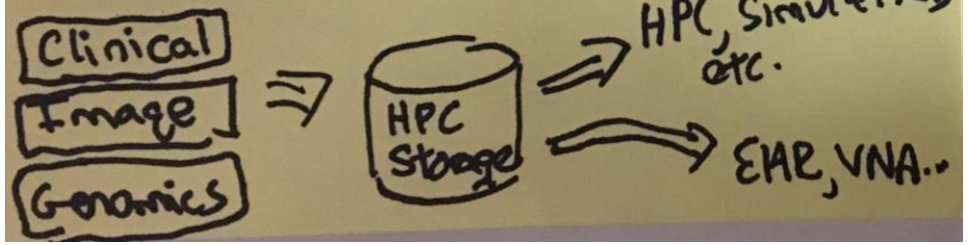
Consolidate drug response data or generate new data in an annotated, standardized way.

Cancer prevention:

baseline description of health and definition of cancer-related deviation

Scale predictive modeling from atomistic to patient outcome level

We need to re-architect how we capture, store and access patient data so it can be easily available for research efforts.



Radiogenomics at the pathway level (metabolism)

Build individual patient models first, rather than apply a population scale model to a patient

Understanding and Modeling the cellular (and other) components and interactions in the tumor microenvironment.

Design neuromorphic or custom hardware.

Team Science needs not just experts in different domains...

... but also people who are not experts, rather they have a good understanding in several domains.

These people are "Liaison" people.

Employing Quantum mechanics to get to the bottom of cancer cell mechanics

Promote cancer related problems in engineering including

Sabbaticals for Computational people at Biology Institutions

How could sensors enable real-time dosing w/ pain meds so patients have no pain w/ the minimal amount of pain meds?

Source of data from healthy patients for early detection

What are the gaps
- knowledge
- technology
- mechanisms

↳ computational/ solutions

How do you solve a problem w/ data that may not contain the solution?

Figure out tech that will help you get the data you actually want

high impact publications to help drive culture shifts/ embracing new resources in cancer community

Can we build models of the specific correlations of patient health information in order to create personalized health strategies for a patient?

ethical considerations — use data w/o misusing it

Think / teach abstract models for biology highlighting intervention strategies?
eg:- Dominant negatives
- Synthetic lethals

Utilizing data already gathered that may have an effect:
Heart rate, breathing
circadian rhythms

Study the Normals
Establish baselines for everyone

focus on enabling technologies for generating early detection data & imaging & biomarkers.

In 10-15 yrs. — Can we scale to timescale of patient w/ recurrence

Develop AUTONOMOUS + ADAPTIVE
↳ DRUGS
ADAPT TO PATIENT/TUMORS

Go back & reanalyse drug hits that failed later — identify the subset they worked on

Everyone wants prevention & early detection. Leverage that desire to trade continuous surveillance/ data collection — throw in free health care. Basic health care should be tracked like a [illegible]

More Stratification
we have the data for this
Allow the raw data to flow

Biology seems to be moving away from acknowledging what is viewed as purely technical contributions — that is a mistake, it is core to team science, money is insufficient compensation for smart people who want to contribute

Recognition

Can we build a central tech platform for data sharing and analysis that adheres to the FAIR principles, respects patient privacy rights, and minimizes workload and culture barriers to participation for researchers?

Flops are often not a meaningful computing power metric. What are good metrics in this space?

Bridge the gap between experimentalist & computational scientist by developping tutorials, lessons learned, and use cases from common pipelines.

in silica drug development
w/ prediction of off
target effects that
won't revealed until
drugs are in general pop

Universal data - IoT
- lots of info

Cancer prevention:
Accurate detection

What could embedded
and connected sensors
detect at an early
state?

Collaboration
- Establish better communication
Mechanisms
- Matchmaking across
multiple areas in
Gout.
- Leverage NLM

Can computing accelerate the
exploration of combination treatments
by designing/suggesting new clinical
trials?

elimination of simple
bias in risk scores

OBTAINING BIOMARKERS
FOR RESPONSE TO RADIOTHERAPY
- LONGITUDINAL IMAGING
- LINK TO GENOMICS
- NEED LARGE SAMPLE SIZES

validation of
model predictions,
particularly when
extrapolating
+ uncertainty quantification

WHAT DOES IT MEAN
TO DO SMALL SAMPLE
LEARNING IN A BIG
DATA WORLD

hyper parameter
optimization of DL
models of clinical
data, imaging data, txt
medical

capturing emotional
state + social structures
in synthetic data

Predictive
Model trustworthy?
- how to make sure
- uncertainties
- Generalization capability

Bridging the Gaps between experimentalist and CS:
1) reducing ~~effo~~ duplication of efforts
2) enabling knowledge sharing.

how to allow groups to share data (tech solution that addresses IP, privacy, legal concerns to overcome cultural barrier

VIRTUAL TWIN TO ADAPT PERSONALIZE AND ADVANCE CLINICAL TRIAL DRUG COMBINATION (i.e. CROSS OVER, DROP OUT, NEW TREATMENT)
- RARE CANCER VARIANTS
- SEVERE AEs / IMMUNOTHER
TO SPEED UP CANCER TRIALS

Methods to validate synthetic data as being appropriate for medical research

eliminating need for registries b/c we can capture/query all the info the registries record + store structure in a real time consistent fashion across the entire healthcare system

updating predictive models based on pre txt data to for post txt longitudinal info into w little post txt info

1- Apply already developped pipelines ~~or si~~ in a standard. on similar type of data to potentially answer different questions.

true integrative analytics => full spectrum omics -> imaging -> clinical -> behavioral -> environment

Design methods to target the 'driver' mutations for cancer that have been identified by a cohesive model of genomics + phenomics

Incorporation of non-traditional data to understand whole life trajectory (eg outside care setting)

Proof of concept with focused effort towards pancreatic cancer

PEELING BACK THE LAYERS FROM DEEP ANALYTICS / LEARNING TO UNDERSTAND THE DATA THAT ULTIMATELY CONTRIBUTES TO TREATMENT PLANNING.

increase trustworthiness & remove "black box"

- Multimodal data integration, harmonization, and storage & management
- Centralized OR distributed
- Capture metadata
- How about data schemas
- Sustainability ** **

Solutions are developed w/ short horizons due to limited resources or resources w/ expiration date.

COMBINING VARIOUS DATA SOURCES IN AN ACCESSIBLE WAY.
WHAT IS THE DATA ARCHITECTURE

Can a model help us differentiate the "driver" and the "passenger" mutations for cancer.

heterogeneity of genomic data across data types

What are proxy model environments to Clinical data?
i.e. mouse / chimera primate

Validation of algorithm performance across different locations and platforms at a level to support clinical patient care.

Processor design for biological modeling, similar to GPUs for graphics / gaming → accelerate for biological modeling

Push the limit of nano-technology as applied to biological sciences

what range of incentives are needed to get patients to contribute their "Normal" / non-normal Clinical data.

GET ENOUGH PPL SEQUENCED (AND HAVE DATA IN AN ACCESSIBLE FORMAT) TO BUILD ALL THESE 'FANCY' COMPUTATIONAL MODELS.

How can we leverage our data to capture population-level information in a way clinical trials can't? Can we use models to inform them?

Way to unify data reporting/generation in a high throughput way so can compare apples to apples.

Co-design of hardware, software & algorithms to solve bio problems. Could you design custom hardware for the bio you are doing?

logistical issue: NCI value high computing, enable it

Common mutation pathways across cancer patients

How can we do long-time, predictive assessments?

Computational infrastructure that allows real-time analysis with simulations

curative to preventive care

Learning Health Care System using large-scale computers

① Knowing what data elements to use; AND (data, features) EHR, social media, genomics

② Semantic integration of different datasets (even the same variable in 2 ≠ datasets may have slightly ≠ meanings!)

Retention & Development of Comp. Sc. workforce in academia

- Can we frame a bio. problem similarly to AlphaGo, with similar success? (and be robust)

What are the inputs for predictive models? How do we represent biological knowledge in a computational form?

Evaluation mechanism & simulations and model fittings

Intelligent selection of datasets for experiments. feature selection

Life cycle of data - from start to finish

Biology-inspired computing

How might we solve the social component of connecting the different comm. Cancer, ML, Computing, Math, etc.

New areas of opportunity: Reanalyzing data that has gotten disassociated from experimental Meta data (applies to Nuclear tests AND biology!)

How do we fund excellent teams (long term) rather than projects (think career devel., tenure, etc.)

capturing + ~~analysis~~ interpret info as it is generated from the clinical care team, patient

Advancements in ENG - sensors/devices more data - more ML DL

Can we design a Computational Cellular Tracker?

how can we inform/use/understand both n of 1 and generalizable to the population

HOW TO STRUCTURE ~~Def~~ SCIENTIFIC CALCULATIONS ACROSS SCALES.

IS THERE A UNIVERSAL POLICY SCHEME OR AUTHZ FRAMEWORK FOR DATA ACCESS?

~~automat~~ real time, ~~auto~~ automated pipelines for cleaning and incorporating new data streams

understanding of the quality and validity of data sources

Total simulation of multi-cell tissue (or even single cell, which hasn't been done...)

think beyond development of computational algorithms to IMPLEMENTATION + INTEGRATION ~~o~~ EMR/portals

how we know when we need on-the-fly, real-time training of algorithms for clinical care? When we don't know what features are input for specific pts

We need a way to detect a pre-cancerous state in a patient

Example:
Computational antibody design for a specific cancer protein to be injected into a patient for detection purposes.

Develop a standard language for the representation of models of cancer at all levels

Use HPC to explore and find effective heuristic strategies for searching the space of cancer models that are supported by available data & knowledge.

How to increase:
Computational power?
This is my limit now

How can sensors enable intervention when it is needed for:
- people contemplating suicide
- people struggling w/ addiction

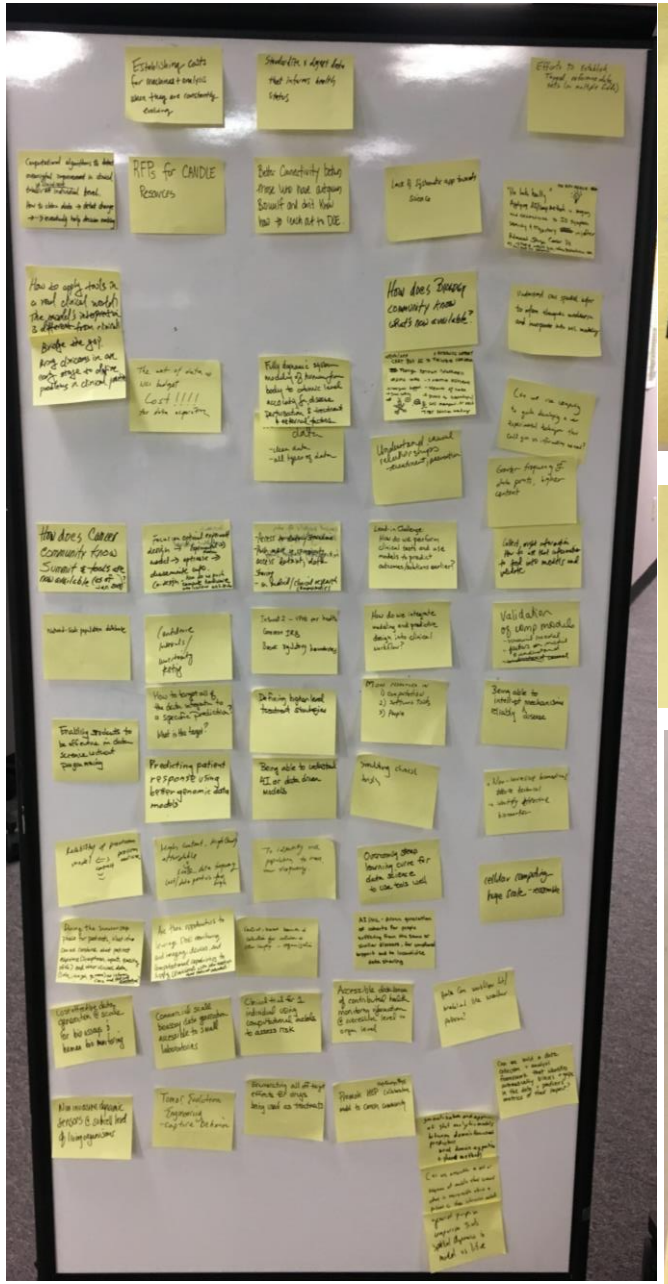
Use HPC to help develop and implement new cancer measurement technologies.

How to integrate population-level digital data & model risk?

Can we build a virtual tissue compilation tissue environment to study across specimen?

How do we model continuous measurement?

Can the concept of a digital twin be useful for predictive oncology?



Computational algorithms to detect meaningful improvement in clinical trials ~~& clinical care~~ at individual level,
How to obtain data → detect change
→ ··· → eventually help decision making

Better Connectivity betwn those who have outgrown Beowulf and don't know how to reach out to DOE.

How does Biology community know what's now available?

Standardize & digest data that informs health status

RFPs for CANDLE Resources

Efforts to Establish Tagged reference data sets (in multiple fields)

Establishing costs for machines + analysis when they are constantly evolving

"He looks healthy"

THE ALEX TABBEK IDEA

Applying AI/Comp Methods in Imaging and telemedicine to ID symptom severity & trajectory ~~onset~~ in/after Advanced Stage Cancer DX
eg → rapid weight loss, inform palliative care
pre/post DX compare.

Lack of Systematic app towards Science

How to apply tools in a real clinical world? The model's interpretation is different from clinician's

Bridge the gap.
Bring clinicians in an early stage to define problems in clinical practice

mHealth/APP & ESTABLISH RAPPORT
CHAT BOX AI TO FACILITATE FREE-FORM
PRECISE PATIENT GENERATED
HEALTH INFO -> ADAPTIVE ASSESSMENT
-> caregiver support. - experience of care
-> goal setting
-> Direct to knowledge/
self-management, or visit
-> TRT Decision making

Understand causal relationships
-treatment/prevention

Fully dynamic system modeling of human from body to atomic level accounting for disease perturbation & treatment & external factors

data
- clean data
- all types of data

The amt of data => NCI budget
Cost!!!!
for data acquisition

Understand cell spatial infor to inform therapeutic mechanism and incorporate into ML modeling

Can we use computing to guide developing a new experimental technique that could give us information we need?

Greater frequency of data points, higher content

How does Cancer Community Know

Summit & tools are now available (as of Jan 2018)

Focus on optimal experiment design → experimental data → model → optimize → disseminate info.

Co-design - how do we push compute hardware into current architecture

-Access to data/Structure

-Push more in community, access dataset, data sharing

- in medical/clinical research communities

Lead-in Challenge:

How do we perform clinical tests and use models to predict outcomes/solutions earlier?

How do we integrate modeling and predictive design into clinical workflow?

Collect, right information How to use that information to feed into models and validate

Validation of comp models

- survival model
- factors on model to understand

Internet 2 – VPNs for health

Common IRB

Break regulatory boundaries

Confidence intervals/ uncertainty testing

National-Scale population database

Enabling students to be effective in data science without programming

How to target all of the data integration to a specific prediction?

What is the target?

Defining higher level treatment strategies

Being able to interrupt mechanisms reliably ↑ disease

cellular computing — huge scale - reassemble

simulating clinical trials

Predicting patient response using better genomic data models

Being able to understand AI or data driven models

→ Non-invasive biomedical device technical

→ identify effective biomarker

More resources in
1) computation
2) software tools
3) People

Reliability of prediction model ⟺ precision medicine
contrast w

During the survivorship phase for patients, What info can we combine about patient experience (symptoms, impacts, quality of life) and other clinical data (labs, images, genome) to inform care and patient expectations?

Overcoming steep learning curve for data science to use tools well

To identify risk population, to scan more frequency.

High Content, High Quality affordable
" scale data frequency
cost/data point => too high

Content-based search & selection for cellular & other imaging - organizati

AI/ML - driven generation of cohorts for people suffering from the same or similar diseases, for emotional support and to incentivize data sharing

Are there opportunities to leverage DOE monitoring and imaging devices, and Computational capabilities to supply Clinicians with information about patient outcomes?

Cost effective data generation @ scale for bio assays & human bio monitoring

Clinical trial for 1 individual using computational models to assess risk

Can we build a data collection + analysis framework that identifies automatically biases + gaps in the data + produces metrics of their impact?

Non invasive dynamic sensors @ subcell level of living organisms

Enumerating all off target effects for drugs being used as treatments

generalization and application of stat analytic models between domain focused predictors
- need domain expertise + shared methods

Commercial scale bioassay data generation accessible to small laboratories

Accessible database of contributed health monitoring information @ subcellular level to organ level

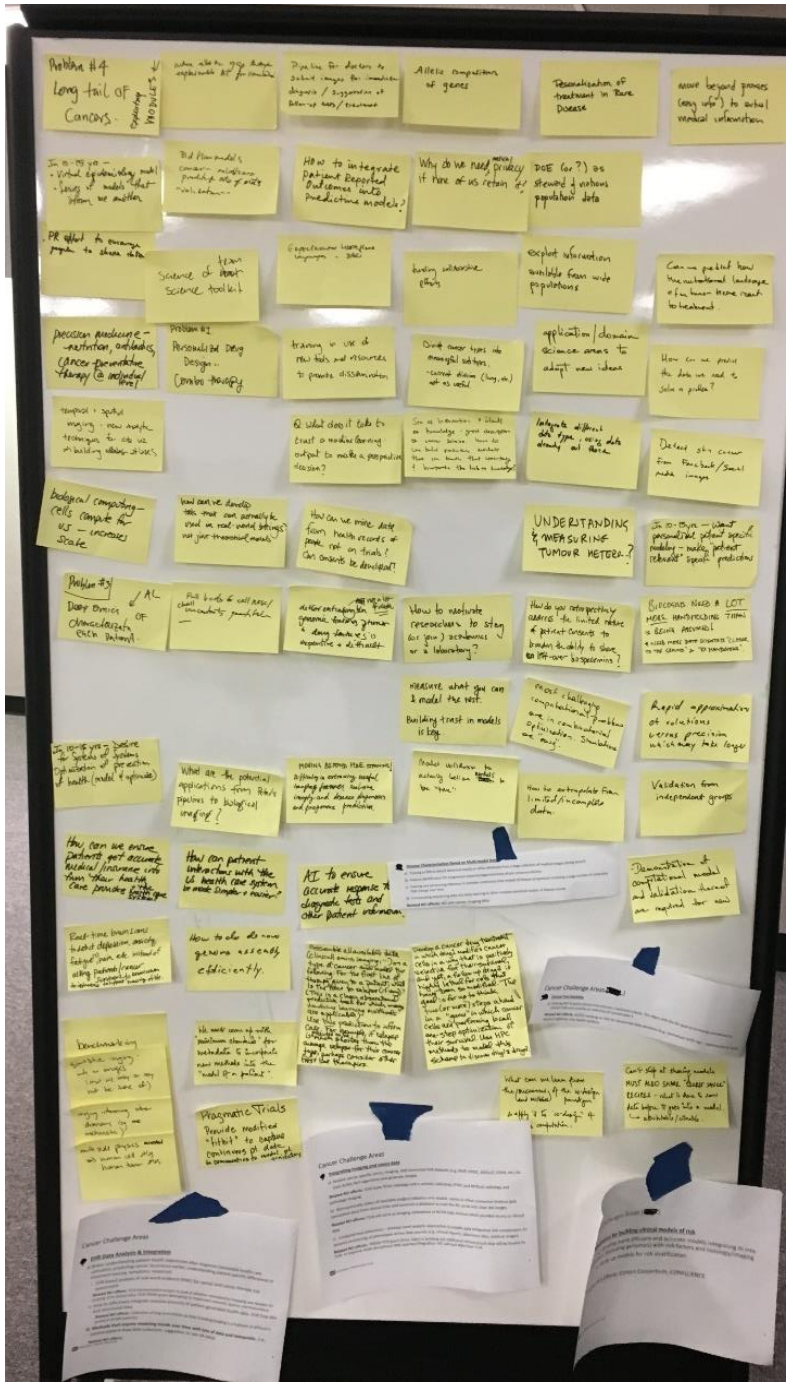
Health Care workflow sit
modeled like weather
patterns?

Can we assemble a set or sequence of models that connect what is measurable about a patient to their information needed

Tumor Evolution
- Engineering
- capture behavior

Promote HEP (High Energy Phys) collaborative model to cancer community

general purpose comparision tools
spatial dynamics in model vs. life



Pipeline for doctors to submit images for immediate diagnosis / suggestion of follow-up tests / treatment

How to integrate Patient Reported Outcomes into Predictive models?

Personalization of treatment in Rare Disease

DOE (or ?) as steward of nations population data

Problem #4

Long tail of Cancers..

exploring MODULES

In 10-15 yrs. —
- Virtual epidemiology model
- Series of models that inform one another

When able to close the gap through explainable AI for simulation

Bld flow models
cancer- metastases
predicting site of mets
-Validation--

Allelic composition of genes

move beyond proxies (easy info) to actual medical information

Why do we need medical privacy if none of us retain it?

precision medicine - nutrition, antibiotics, cancer preventative therapy @ individual level

. PR effort to encourage people to share data

Can we predict how the mutational landscape of a tumor tissue react to treatment.

Experimental workflows languages - DAGs

How can we predict the data we need to solve a problem?

training in use of new tools and resources to promote dissemination

science of team science toolkit

problem #1.

Personalized Drug Design ..

Combo therapy

funding collaborative efforts

explort information available from wide populations

Divide cancer types into meaningful subtypes.

- current division (lung, etc.) not as useful.

application / domain science areas to adopt new ideas

temporal + spatial
imaging - new analytic
techniques for data viz
⇒ building databases atlases

biological computing—
cells compute for
us — increases
scale

Q: What does it take to
trust a machine learning
output to make a prescriptive
decision?

How can we mine data
from health records of
people not on trials?
Can consents be developed?

Integrate different
data type, using data
already out there

UNDERSTANDING
& MEASURING
TUMOUR HETERO..?

Detect skin cancer
from Facebook/Social
media images

In 10-15yrs. — Want
personalized patient specific
modeling — make "patient
relevant" specific predictions

how can we develop
tools that can actually be
used in real-world settings
not just theoretical models

Sea of information ! islands
of knowledge - great description
of cancer science. How do
we build predictive models
that can handle that uncertainty
! Incorporate the lack of knowledge?

Problem #3

Deep Omics AL of characterizatn each PatienA.

How to motivate researchers to stay (or join) academics or a laboratory?

BIOLOGISTS NEED A LOT MORE HANDHOLDING THAN IS BEING ASSUMED!

* NEED MORE DATA SCIENTISTS "CLOSER TO THE GROUND" & "101 HANDBOOKS".

How do you retrospectively address the limited nature of patient consents to broaden the ability to share left-over biospecemins?

Full body & cell resc/ chall uncertainty quantification

measure what you can & model the rest.

Building trust in models is key.

In 10-15 yrs - Desire for Systems of Systems Optimization of prevention of health (model & optimize)

Validation from independent groups

Rapid approximations of solutions versus precision which may take longer

Most challenging computational problems are in combinatorial optimization. Simulations are "easy".

How to extrapolate from limited/incomplete data.

MOVING BEYOND H&E STAINING

Difficulty in extracting useful imaging features. Real-time imagery and disease diagnosis and prognosis prediction.

What are the potential applications from Peter's pipelines to biological imaging?

Model validation to actually believe ~~Models~~ to be "true."

~~Easy~~ not a lot of data to differentiating btn genomic features, tumor & drug features is expensive & difficult

Disease Characterization Based on Multi-modal Data

a) Training a CNN to detect abnormal results or other attributes from a large collection of medical images (being do

b) Feature identification for progression potential/aggressiveness of pre-cancerous lesions

c) Training and performing inference in complex continuous time models of disease progression involving a large nu that change over time

d) Incorporating sequencing data into deep learning or other complex statistical models of disease course

Related NCI efforts: NCI pre-cancer imaging atlas

How can we ensure patients get accurate medical/insurance info from their health care providers + the health care system?

How can patient interactions with the US health care system be made simpler + easier?

-Demonstration of computational model and validation thereof are required for new

AI to ensure accurate response to diagnostic tests and other patient information.

How to do de novo genome assembly efficiently.

Real time brain scans to detect depression, anxiety, fatigue(?), pain etc. instead of asking patients/cancer survivors to know when to intervene to support quality of life.

Assemble all available data (clinical, omics, imaging, ...) on a type of cancer and model the following: For the first line of therapy given to a patient, what is the time to relapse (if any)? (This is a classic observational predictive task for which many machine learning methods are applicable.) Use this prediction to inform care. For example, if relapse is predicted to be much shorter than the average relapse for this cancer type, perhaps consider other first line therapies.

Develop a cancer drug treatment in which drug1 modifies cancer cells in a way that is positively selective for their survival, and yet, a follow up drug2 is highly lethal for cells that have been so modified. The goal is for us to think two (or more) steps ahead in a "game" in which cancer cells are performing local, one-step optimization of their survival. Use HPC methods to model this scheme to discover drug1 & drug2.

benchmarking

quantitative imaging - info in images (that we may or may not be aware of)

imaging informing other domains (eg mechanistic)

multiscale physics connected => human cell atlas human tumor atlas

We must come up with "minimum standards" for metadata to incorporate new methods into the "model of a patient".

What can we learn from the (successes and mistakes) of the co-design paradigm & apply it to "co-design" of biology & computation.

Can't stop at sharing models MUST ALSO SHARE "SECRET SAUCE RECIPES - what is done to raw data before it goes into a model. ↳ attributable / citeable.

Pragmatic Trials

Provide modified "fitbit" to capture continuous pt. data in communities to model pt. trajectory

Cancer Challenges Areas ([redacted])

Simulations for building clinical models of risk

a.) Develop more efficient and accurate models integrating AI into biomarkers (including genomics) with risk factors and histology/imaging

b.) Scale up models for risk stratification

Related NCI Efforts: Cohort Consortium, CONFLUENCE

ancer Challenge Areas

Integrating imaging and omics data

a) Analyze cancer specific omics, imaging, and connected EHR datasets (e.g. MVP, CPTAC, APOLLO, HTAN, etc.) to train AI/ML/NLP algorithms and generate models

Related NCI efforts: TCIA hosts TCGA radiology and is actively collecting CPTAC and APOLLO radiology and pathology imaging

b) Retrospectively collect all available images/radiomics and related -omics or other connected medical data (annotated data) from clinical trials and construct a database to train the AI; scrub and clean the images

Related NCI efforts: TCIA will serve as imaging component of NCTN Data Archive which provides access to clinical data

c) Computational phenomics – develop novel analytic approaches to enable data integration and computation for dynamic monitoring of phenotypes across data sources (e.g. clinical reports, laboratory data, medical images)

Related NCI efforts: PRISM ITCR grant (Prior, Saltz) is building out additional infrastructure that will be hosted by TCIA to improve multi-disciplinary data queries/integration; IDC will pull data from TCIA

5

Cancer Challenge Areas

EHR Data Analysis & Integration

a) Better understanding patient health trajectories after diagnosis (avoidable health care utilization, predicting cancer recurrence earlier, understanding patient-specific differences in treatment toxicity, symptoms, response)

i. EHR-based analysis of real-word evidence (RWE) for cancer and cancer therapy risk assessment

Related NCI efforts: TCIA harmonization project to look at whether retrospective mapping was feasible for existing TCIA clinical data; ITCR PRISM grant attempting to implement semantic queries and featurebase

ii. How to effectively integrate massive amounts of patient-generated health data, EHR free text and structured data

Related NCI efforts: Collection of long-term follow-up data including imaging is a hallmark of APOLLO-5 protocol (8,000 patients)

iii. Methods that require modeling trends over time with lots of data and timepoints. (i.e., passive patient-level data collection; suggestion to use VA data)

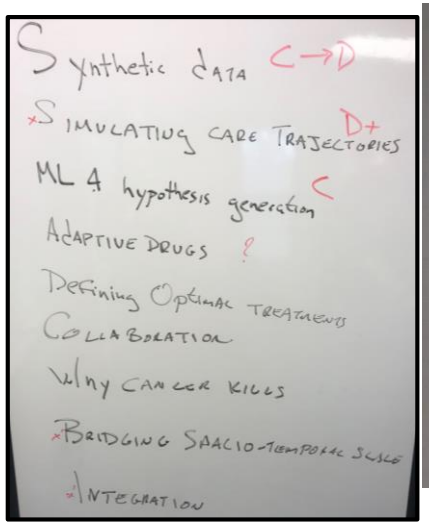
Related NCI efforts:

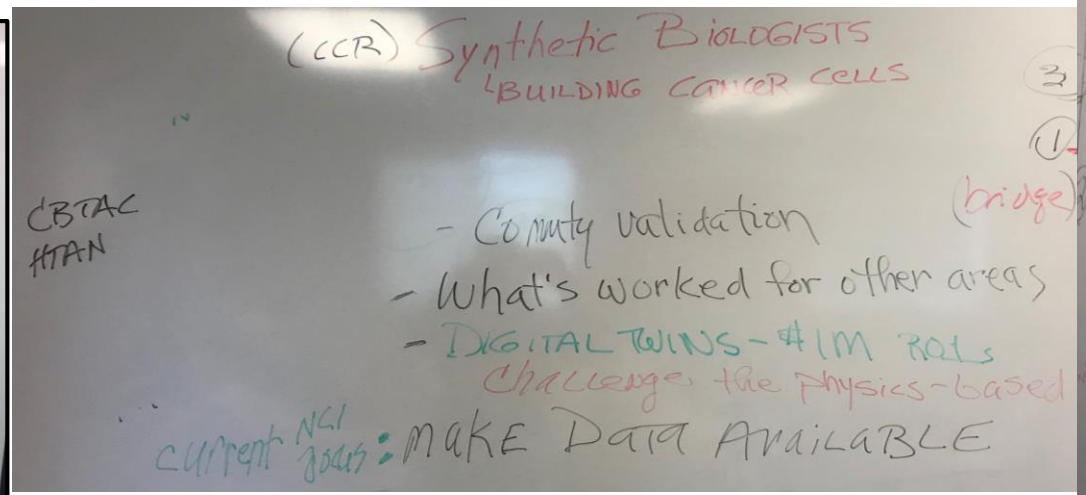
6

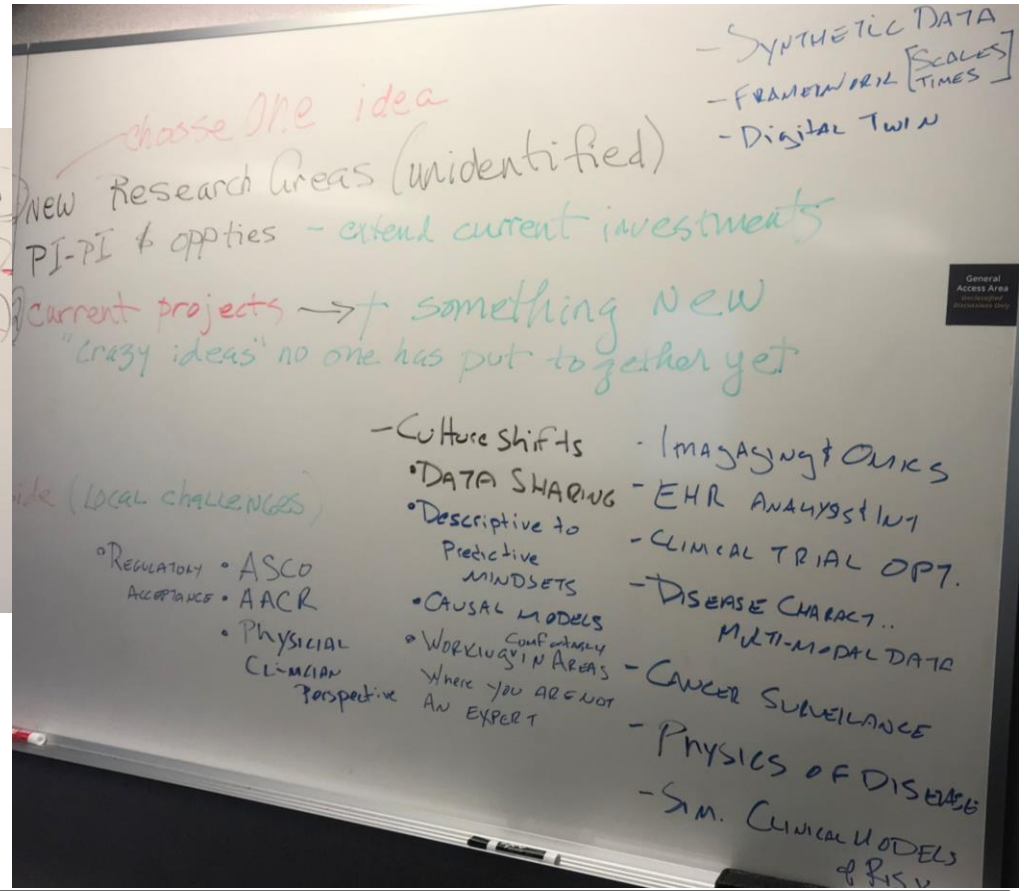
NATIONAL CANCER INSTITUTE

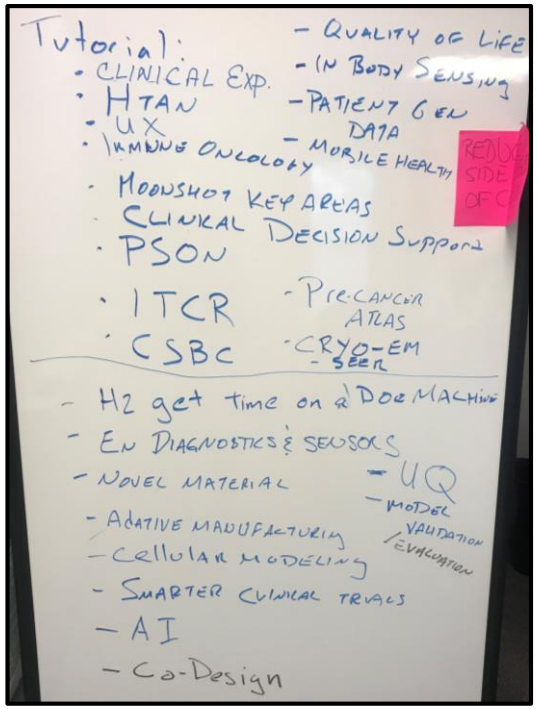
Friday, March 8, 2019

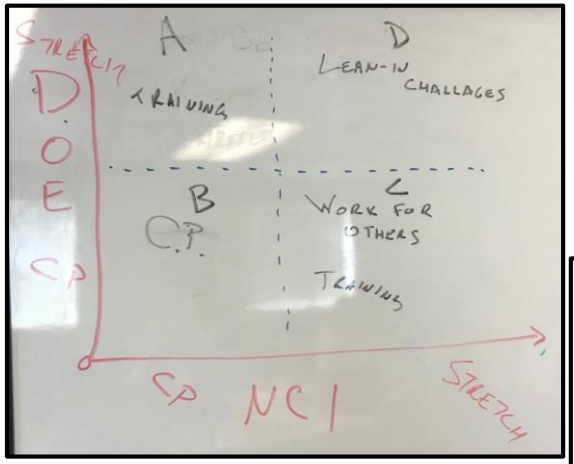
Photos from Planning Team Summary Meeting

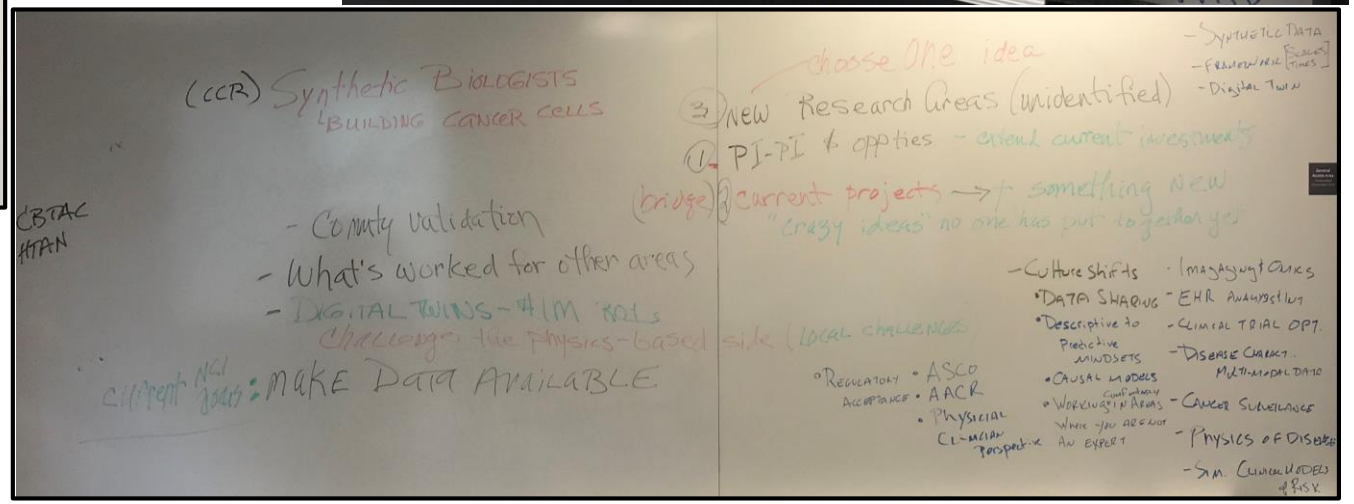

Synthetic data C→D
Simulating care trajectories D+
ML 4 hypothesis generation C
Adaptive drugs ?
Defining Optimal Treatments
Collaboration
Why cancer kills
Bridging spacio-temporal scale
Integration


(CCR) Synthetic Biologists
Building Cancer Cells
CBTAC
HTAN
- County validation
- What's worked for other areas
- Digital Twins - 4IM RAIs
Challenge the physics-based
current NCI goals: make Data Available


choose One idea
- Synthetic Data
- Framework [Scales Times]
- Digital Twin
3 New Research Areas (unidentified)
1 PI-PI & oppties - extend current investments
(bridge) current projects → something New
"crazy ideas" no one has put together yet
- Culture shifts
• Data Sharing
• Descriptive to Predictive Mindsets
• Causal Models
• Working in Areas Complementary Where you are not An Expert
- Imaging & Omics
- EHR Analysis 1M
- Clinical Trial Opt.
- Disease Charact.. Multi-modal data
- Cancer Surveillance
- Physics of Disease
- Sm. Clinical Models of Risk
side (local challenges)
• Regulatory Acceptance
• ASCO
• AACR
• Physicial Clinician Perspective


Tutorial:
• Clinical Exp.
• HTAN
• UX
• Immune Oncology
• Moonshot Key Areas
Clinical Decision Support
• PSON
• ITCR
• CSBC
- Quality of Life
- In Body Sensing
- Patient Gen Data
- Mobile Health
- Pre-Cancer Atlas
- Cryo-EM
- SEER
- H2 get time on a DOE Machine
- En Diagnostics & Sensors
- Novel Material
- Adative Manufacturing
- Cellular Modeling
- Smarter Clinical Trials
- AI
- Co-Design
- UQ
- Model Validation / Evaluation


Stretch
DOE
CP
A
Training
D
Lean-in Challages
B
C.P.
C
Work for others
Training
CP
NCI
Stretch


(CCR) Synthetic Biologists
Building Cancer Cells
CBTAC
HTAN
- County validation
- What's worked for other areas
- Digital Twins - 4IM RAIs
Challenge the physics-based
current NCI goals: make Data Available
choose One idea
- Synthetic Data
- Framework [Scales Times]
- Digital Twin
3 New Research Areas (unidentified)
1 PI-PI & oppties - extend current investments
(bridge) current projects → something New
"crazy ideas" no one has put together yet
- Culture shifts
• Data Sharing
• Descriptive to Predictive Mindsets
• Causal Models
• Working in Areas Complementary Where you are not An Expert
- Imaging & Omics
- EHR Analysis 1M
- Clinical Trial Opt.
- Disease Charact.. Multi-modal data
- Cancer Surveillance
- Physics of Disease
- Sm. Clinical Models of Risk
side (local challenges)
• Regulatory Acceptance
• ASCO
• AACR
• Physicial Clinician Perspective