Modeling the Manufacturing Process in Regenerative Medicine

Theresa Kotanchek

Evolved Analytics LLC

Applying Systems Thinking to Regenerative Medicine- A Workshop

October 22-23, 2020

Session V:

Addressing Regenerative Medicine Manufacturing and Supply Chain Challenges with Systems-Level Approaches

Grand Challenges in Cell Manufacturing

- Lack of **Reproducibility, Standards, and Quality-by-Design (QbD)**
- What **Quality Attributes** Make a Cell Safe and Effective → **What to Measure?**
  - Which cells are "good" in the midst of heterogeneity and which are "bad"?
- **How to Measure Critical Quality Attributes (CQAs), In-line, During Manufacturing?**
- **How to Grow** Billions of Safe and Potent Cells from a Patient/donor?
- **How to "Predict" Safety and Potency** for Specific Indications and Patients?
- How to **Purify, Store, Freeze, Package,** and **Transport** Cells WITHOUT Compromising Quality?
- End-to-end Manufacturing at **Low Cost and High Quality?**
- Lack of **Trained Cell Manufacturing Workforce**

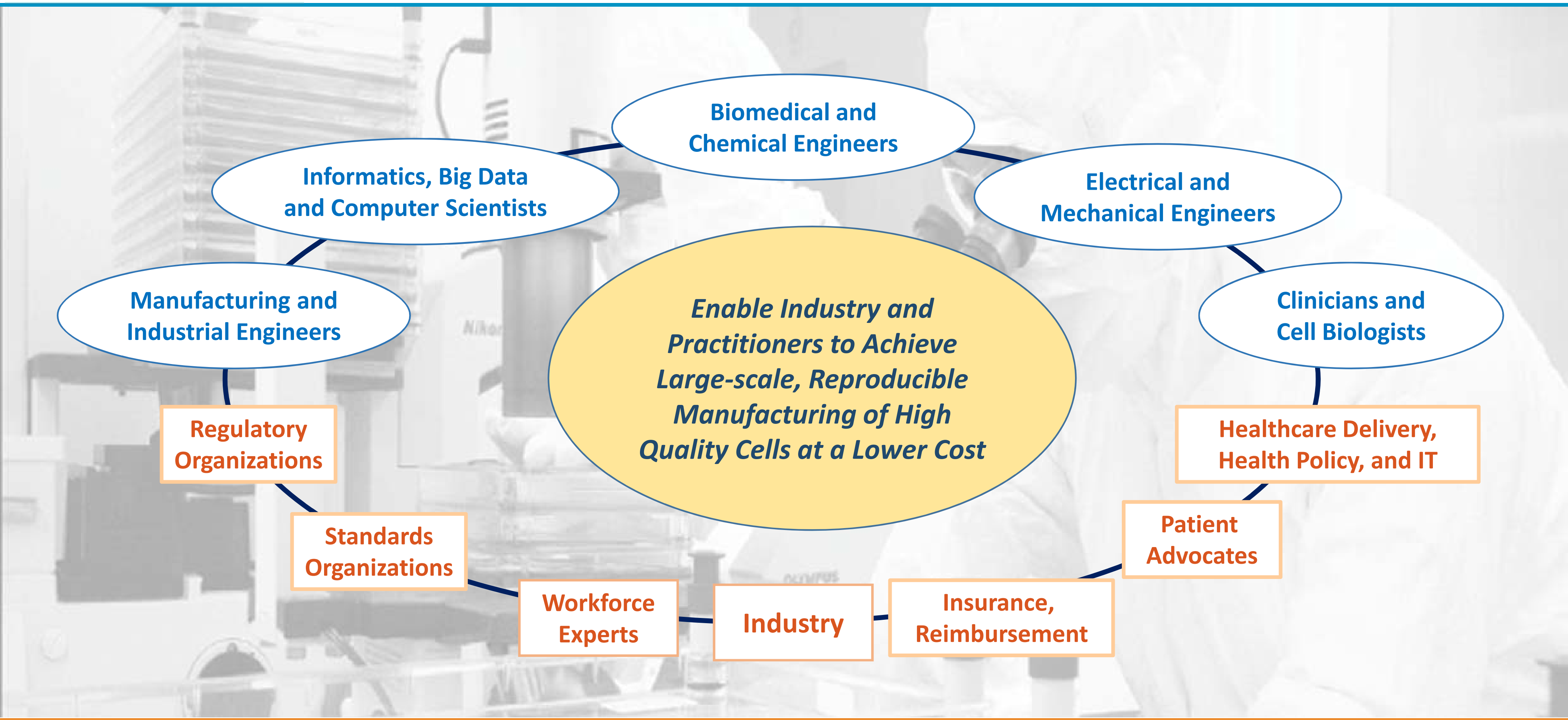
Adopted from *"Achieving Large-scale, Cost-effective, Reproducible Manufacturing of Therapeutic Cells: A Technology Roadmap to 2025" – The NIST/AMTech National Cell Manufacturing Roadmap*

CMaT's Vision

*To **transform the manufacture of cell-based therapeutics** into a **large-scale, lower-cost, reproducible**, and **high-quality engineered process** for broad industry and clinical use.*

*To **become a visionary and strategic international resource and an exemplar** for developing **new knowledge, innovative technologies, diverse workforce** and **enabling standards** for cell-production and characterization processes.*

Bringing All Stakeholders On-deck



CMaT is Strategically Positioned to Address the Challenges Faced by the Emerging Cell Manufacturing Industry

Quality

- **Multivariate CQAs** maximize efficacy and safety
- **Multivariate CPPs** improve process control and reproducibility
- **Real-time monitoring** and predictive analytics
- **Physiologically relevant**, **personalized** assays (human)
- **Standards**, best practices
- Quality-by-Design (**QbD**) and **Flexible Automation**

Cost

- **Lower** failure **risk** and supply-chain risk
- **Reduced** need for **scale-up** through maximizing quality
- Lower cost through **maximizing safety**
- Readily available highly trained, and diverse **workforce**
- **Automation**

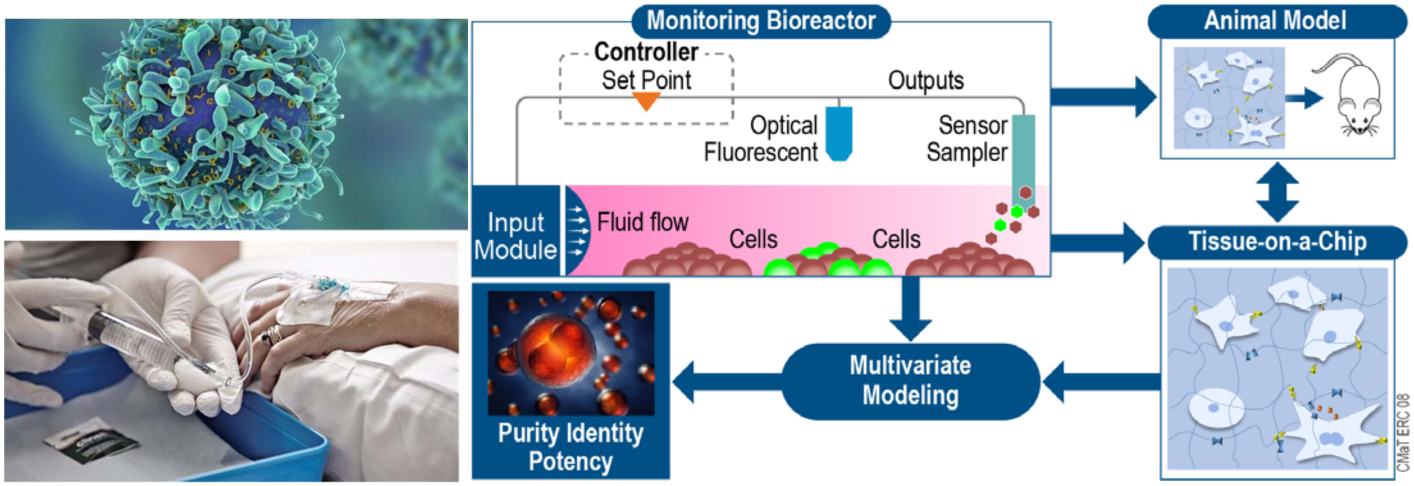
Speed

- Develop **faster and more efficient** bioprocesses
- **Lower failure** rate
- **Faster batch release**
- Direct input from **industry** and **faster translation** to industry and clinic
- Readily available, highly trained, and diverse **workforce**

Agility

- Ideal **supply-chain and logistics** model
- **"What if" models** allow faster response to issues
- **Nimble operation and adaptability** to changing science and industry needs
- **Inclusive Innovation ecosystem** of all stakeholders

Dynamic Sampling Platform

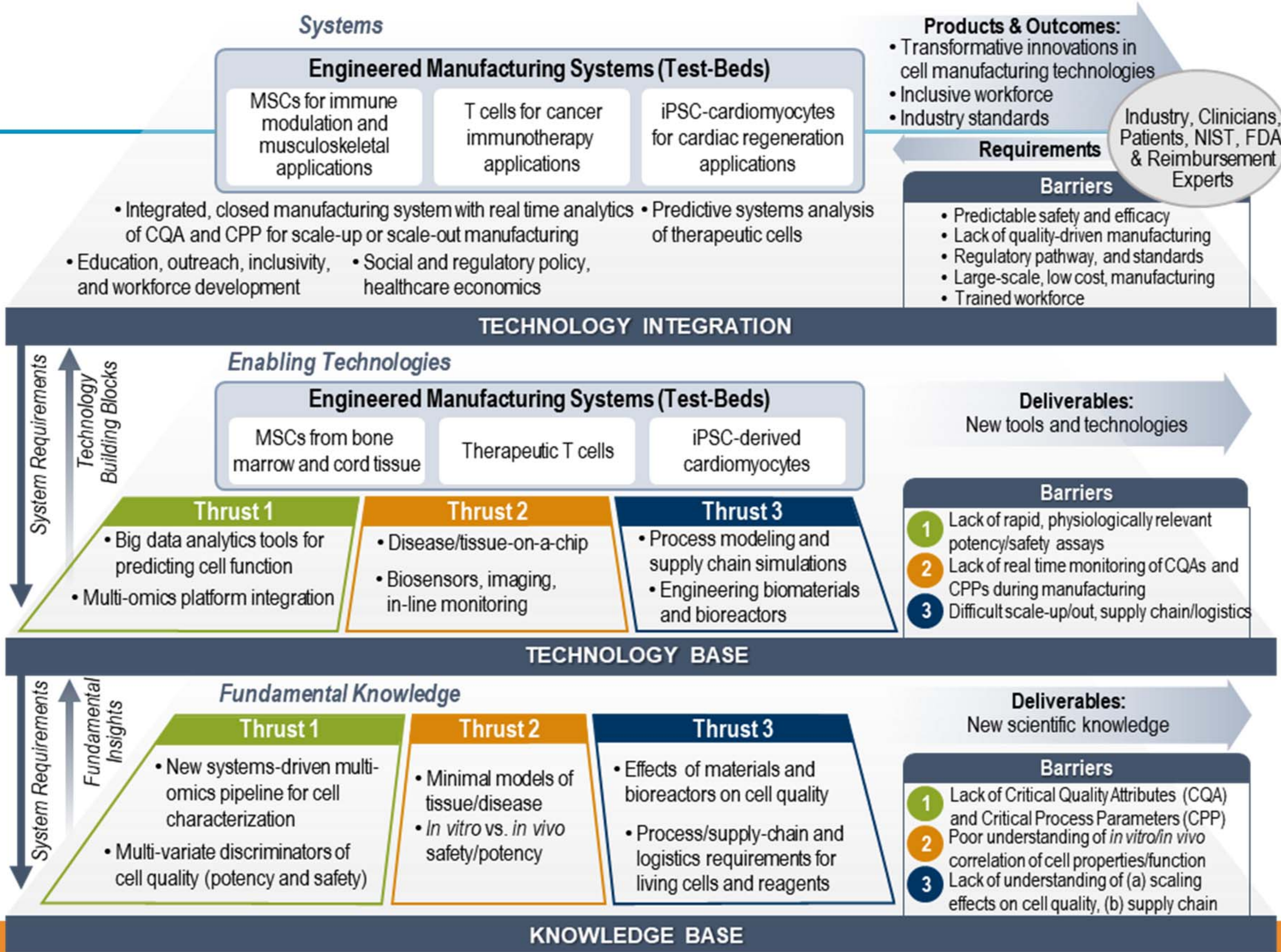


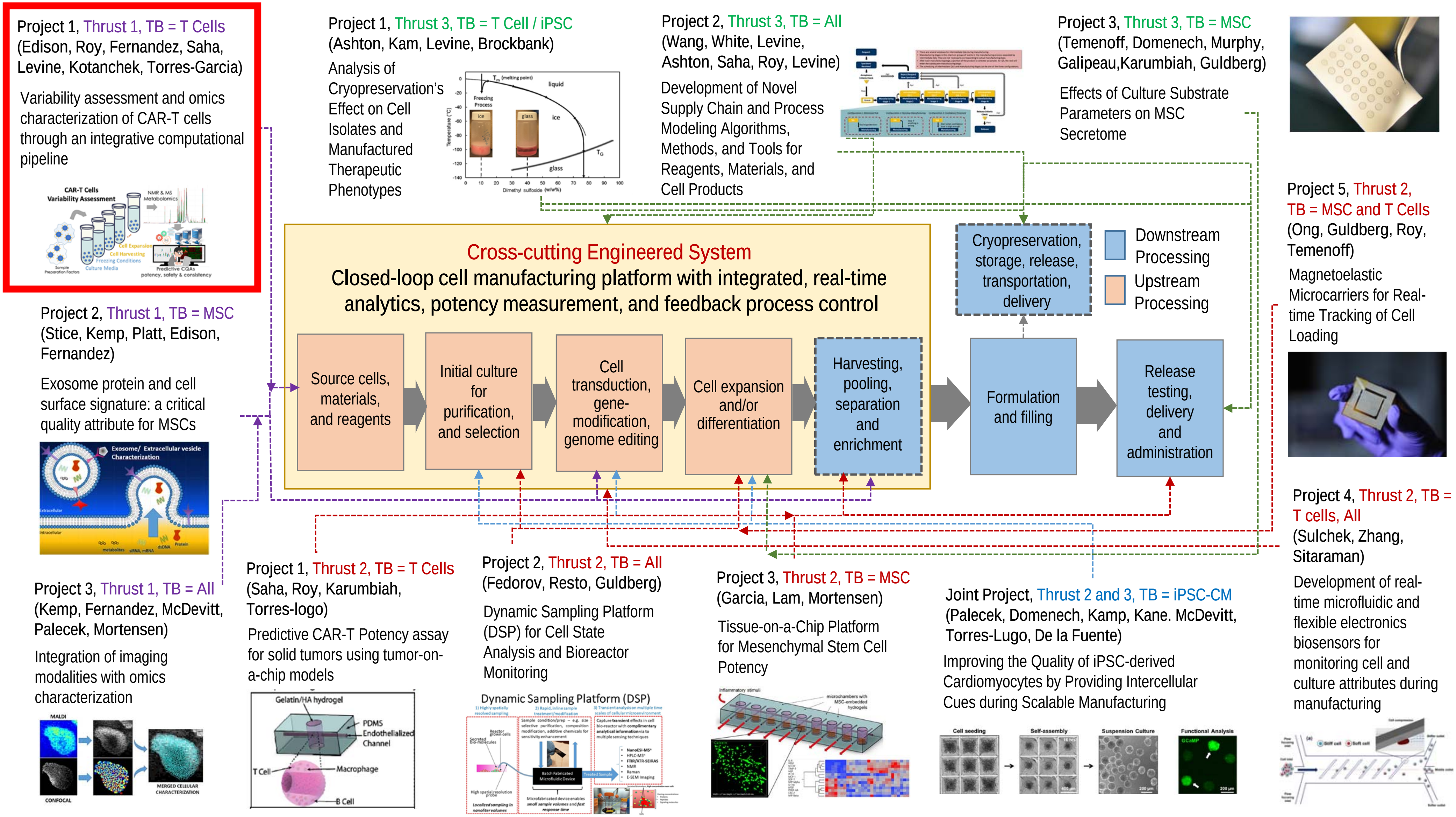
How do we measure & predict quality?

Engineer reproducible, predictive measurement & assay technologies that enable batch & continuous monitoring of cell state & product

- non-destructive, in-line, closed system analysis using real-time sampling, reporters/ sensors & imaging tools as process analytical technologies
- "potency-on-a-chip" 3D disease & organoid models

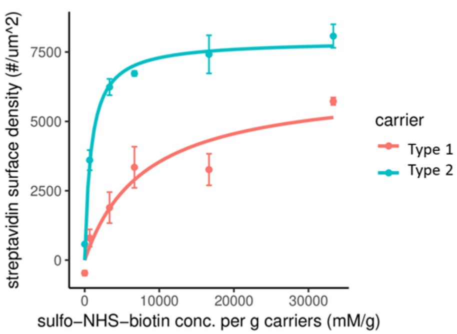
The Three Plane Chart



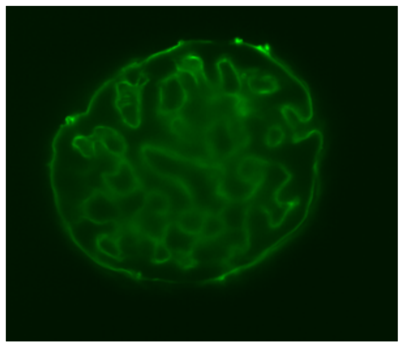

Project 1, Thrust 1, TB = T Cells
(Edison, Roy, Fernandez, Saha, Levine, Kotanchek, Torres-Garcia)
Variability assessment and omics characterization of CAR-T cells through an integrative computational pipeline
CAR-T Cells Variability Assessment
NMR & MS Metabolomics
Predictive CQAs potency, safety & consistency
Project 1, Thrust 3, TB = T Cell / iPSC
(Ashton, Kam, Levine, Brockbank)
Analysis of Cryopreservation's Effect on Cell Isolates and Manufactured Therapeutic Phenotypes
liquid
ice
glass
Freezing Process
Dimethyl sulfoxide (w/w%)
Project 2, Thrust 3, TB = All
(Wang, White, Levine, Ashton, Saha, Roy, Levine)
Development of Novel Supply Chain and Process Modeling Algorithms, Methods, and Tools for Reagents, Materials, and Cell Products
Project 3, Thrust 3, TB = MSC
(Temenoff, Domenech, Murphy, Galipeau, Karumbiah, Guldberg)
Effects of Culture Substrate Parameters on MSC Secretome
Project 5, Thrust 2, TB = MSC and T Cells
(Ong, Guldberg, Roy, Temenoff)
Magnetoelastic Microcarriers for Real-time Tracking of Cell Loading
Cross-cutting Engineered System
Closed-loop cell manufacturing platform with integrated, real-time analytics, potency measurement, and feedback process control
Source cells, materials, and reagents
Initial culture for purification, and selection
Cell transduction, gene-modification, genome editing
Cell expansion and/or differentiation
Harvesting, pooling, separation and enrichment
Formulation and filling
Release testing, delivery and administration
Cryopreservation, storage, release, transportation, delivery
Downstream Processing
Upstream Processing
Project 2, Thrust 1, TB = MSC
(Stice, Kemp, Platt, Edison, Fernandez)
Exosome protein and cell surface signature: a critical quality attribute for MSCs
Exosome/ Extracellular vesicle Characterization
Project 3, Thrust 1, TB = All
(Kemp, Fernandez, McDevitt, Palecek, Mortensen)
Integration of imaging modalities with omics characterization
MALDI
CONFOCAL
MERGED CELLULAR CHARACTERIZATION
Project 1, Thrust 2, TB = T Cells
(Saha, Roy, Karumbiah, Torres-logo)
Predictive CAR-T Potency assay for solid tumors using tumor-on-a-chip models
Gelatin/HA hydrogel
PDMS
Endothelialized Channel
T Cell
Macrophage
B Cell
Project 2, Thrust 2, TB = All
(Fedorov, Resto, Guldberg)
Dynamic Sampling Platform (DSP) for Cell State Analysis and Bioreactor Monitoring
Dynamic Sampling Platform (DSP)
Project 3, Thrust 2, TB = MSC
(Garcia, Lam, Mortensen)
Tissue-on-a-Chip Platform for Mesenchymal Stem Cell Potency
Joint Project, Thrust 2 and 3, TB = iPSC-CM
(Palecek, Domenech, Kamp, Kane. McDevitt, Torres-Lugo, De la Fuente)
Improving the Quality of iPSC-derived Cardiomyocytes by Providing Intercellular Cues during Scalable Manufacturing
Cell seeding
Self-assembly
Suspension Culture
Functional Analysis
Project 4, Thrust 2, TB = T cells, All
(Sulchek, Zhang, Sitaraman)
Development of real-time microfluidic and flexible electronics biosensors for monitoring cell and culture attributes during manufacturing

Degradable Microscaffold (DMS) Microcarrier Cultures Can Expand More Central Memory & LN Homing T cells

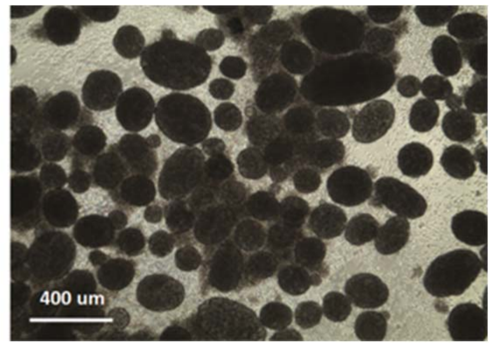
Ability to vary ligand density



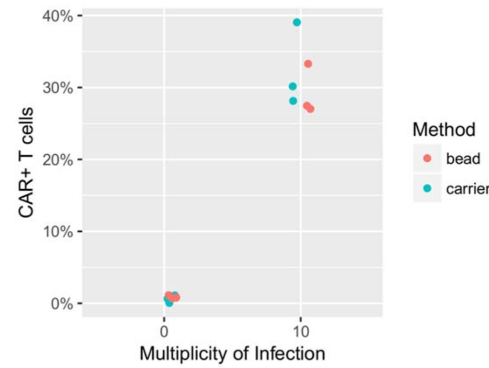
Uniform Coating (FITC-Biotin)



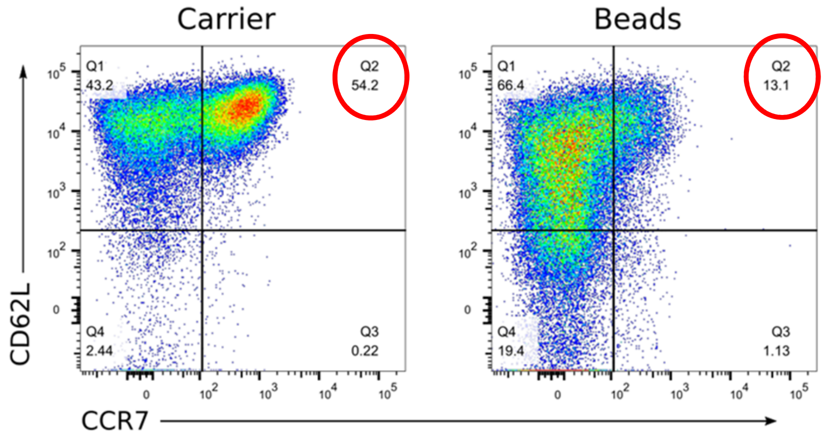
High Density Culture



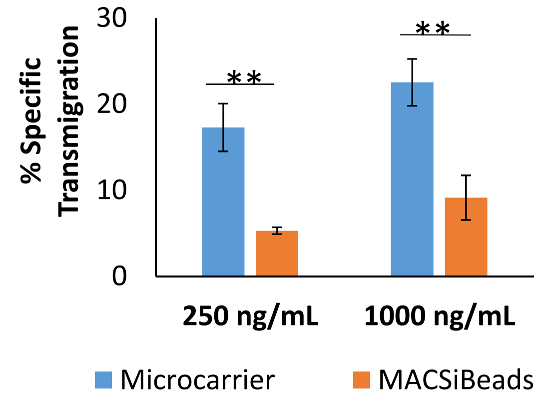
CAR Transduction Efficacy is Similar



Representative % of CD62L+CCR7+ Central Memory T cells



CCL21 chemotaxis assay: More potentially LN homing T cells



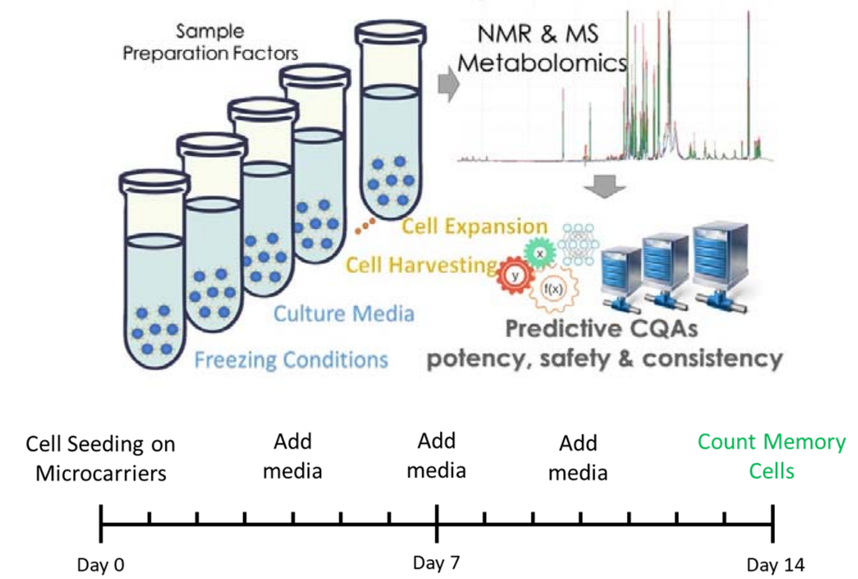
Nathan J. Dwarshuis,, Krishnendu Roy

T cell Characterization Project Aims

Development of a **workflow** to **enable multi-omics characterization and unbiased modeling** of the **end-product** and **early, predictive-signatures of quality** during manufacturing

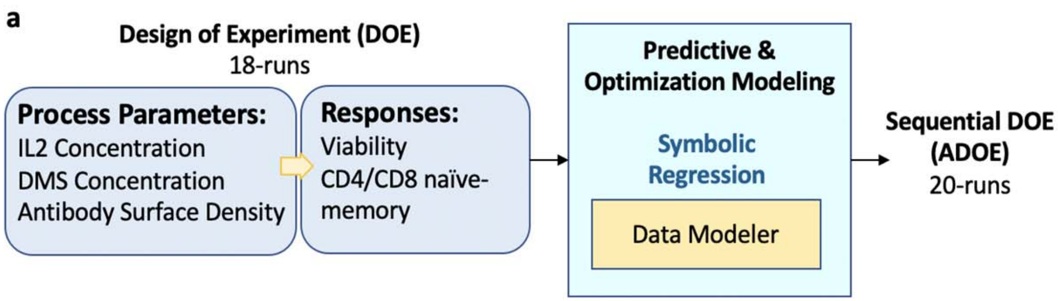
Understanding ***variance***

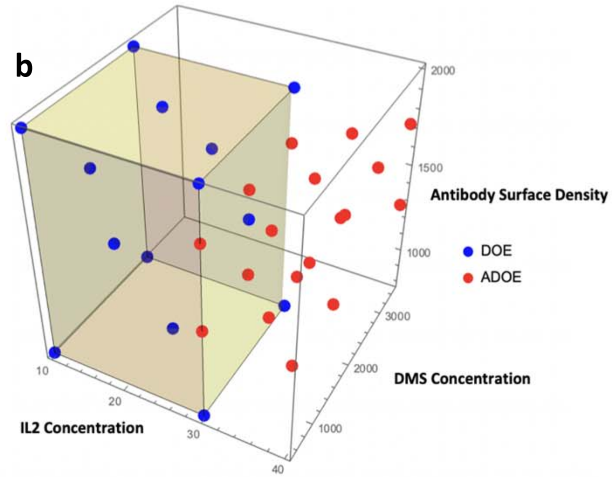
Establishing **CQAs & CPPs** that are predictive of ***potency***, ***safety*** and ***consistency***



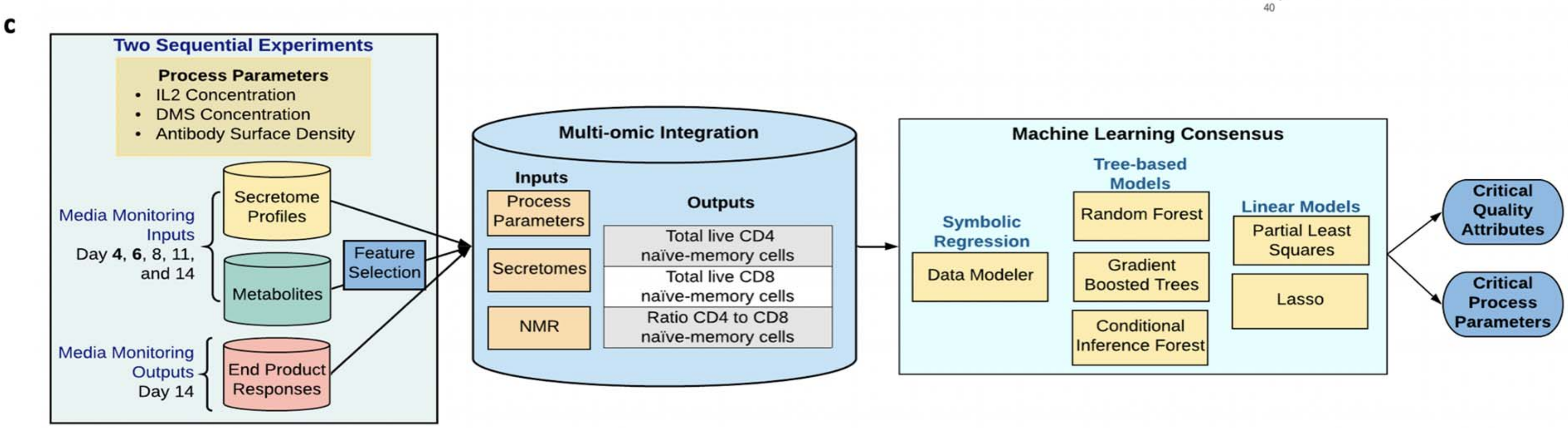
Key achievements

Phase 1: Experimental Process Parameters Optimization





Phase 2: Multi-Omics Integrative Approach for Early Predictive Signatures

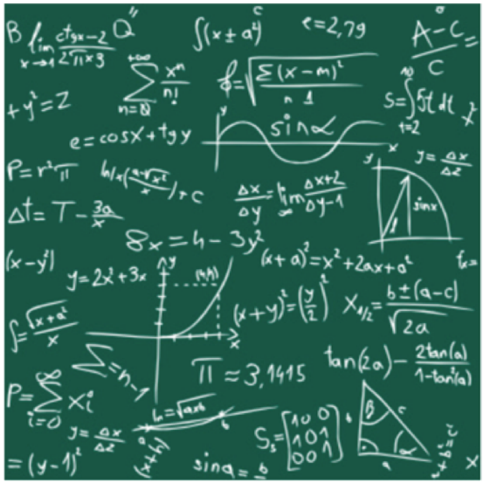


Data Challenges

High complexity –
Heterogeneous –
Unknown behavior
Volume of data

Modeling Challenges

Apriori model structure (linear vs non-linear)
Standalone vs interactive effects
Prediction performance vs overfitting
Interpretability
Computational infrastructure



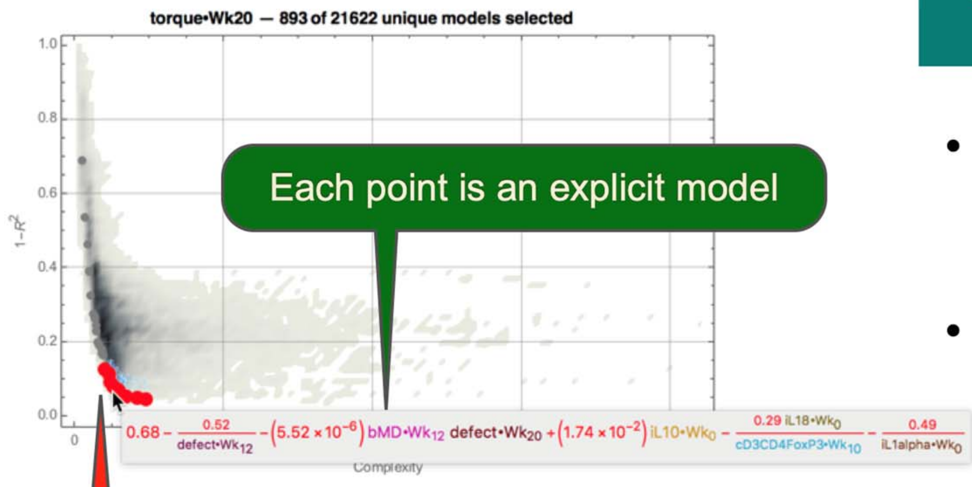
Modeling Options Are Determined By

What's Known & What's Unknown

		Model Structure		
		Known	*Known*	*Unknown*
		Linear	*Non-Linear*	*????*
Driving Variables	*Known*	Linear Regression	Non Linear Regression Parameter Estimation	Neural Networks SVM Random Forests Symbolic Regression
	Unknown	***Symbolic Regression***		

Evolution handles Symbolic Regression model development

- Evolutionary computation automatically generates novelty from available data
- Solutions are found by simulating natural evolution & selection
- Rather then biological species being created, mathematical expressions are created
- Novelty is generated by the competition for high fitness of the mathematical expressions during the simulated evolution.
- Explicit algebraic, interpretable models are generated.
- Symbolic regression is an augmented intelligence hypothesis generator and optimizer.



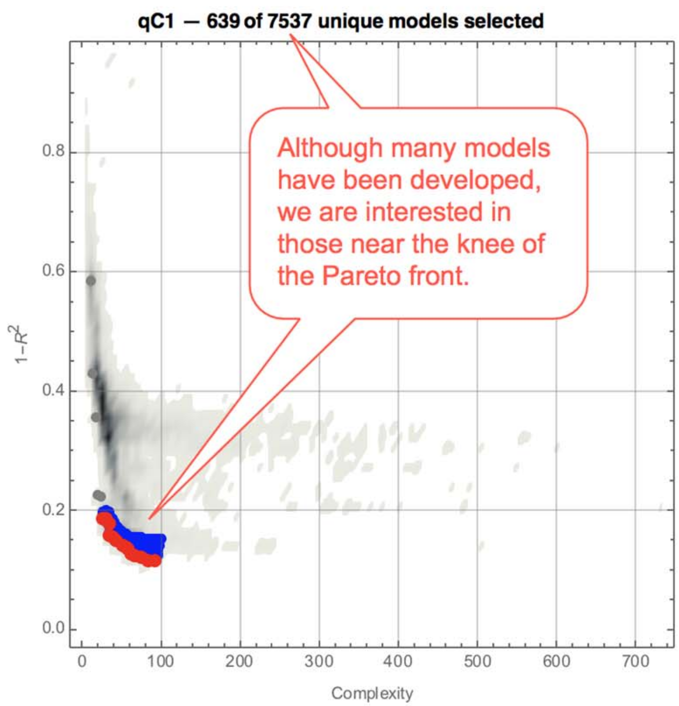
Models at the “knee”of the Pareto front provide best balance of complexity and accuracy

Select models with optimal performance & complexity

The Pareto Front

- Pareto front solutions are the best “bang-for-the-buck”
- Identifies trade-off between competing objectives of accuracy vs. complexity
- Unwarranted complexity is punished automatically
- Select models with optimal performance and low complexity

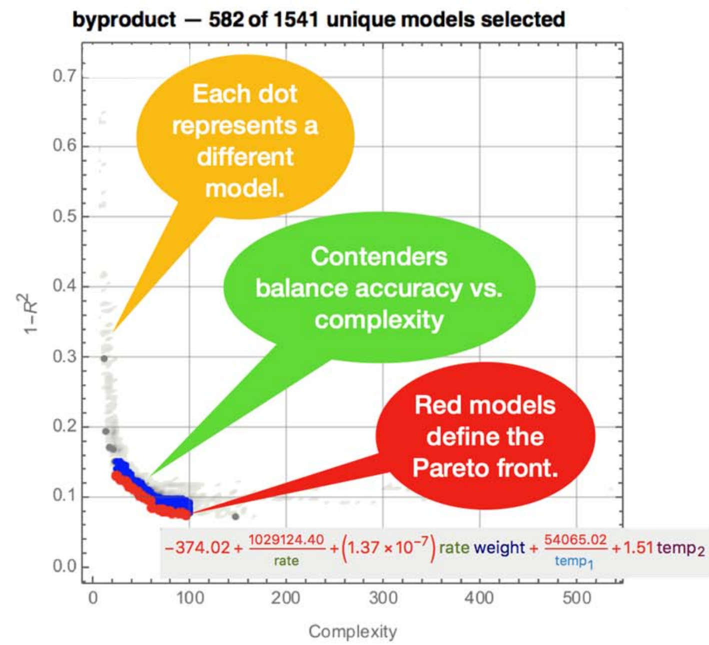
You can learn a lot from a model set …



- Modeling Potential
- Complexity vs Accuracy Trade-off
- Number of Variables Required
- Variable Presence
- Variable Combinations
- Variable & Combination Distributions
- Metavariables

Big Data vs. Big Insight

- Data sizes
  - 2-10,000 variables
  - 10-1,000,000 records
- Automated hypothesis generation & refinement
  - Develop explicit algebraic models
  - Reward simplicity & accuracy
  - Focus on the good and simple models
- Many models are contenders
  - Exploit contenders for insight & guidance
- Iterate towards final model set
  - Focus on most impactful/useful variables
  - Implement trustable model from final good & simple models



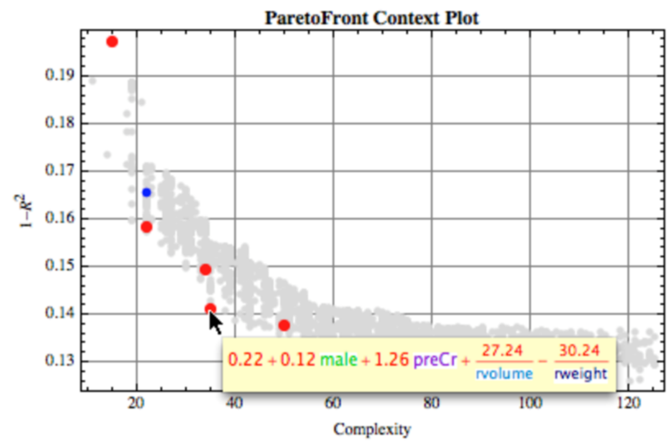
We have to let the data determine the proper tradeoff of complexity vs. accuracy

Ensembles are Trustable Models

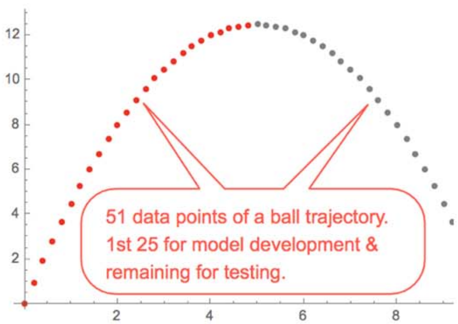
- Ensemble Creation
  - Start with interesting models
    - reasonable accuracy & complexity
    - desirable variables and variable combinations and dimensionality
  - Automatically chosen to maximize diversity of error residuals
- Ensembles
  - Agree where constrained by data
  - Diverge when exposed to novel parameter conditions
  - Guide decision making
  - Enable active learning & design of experiments

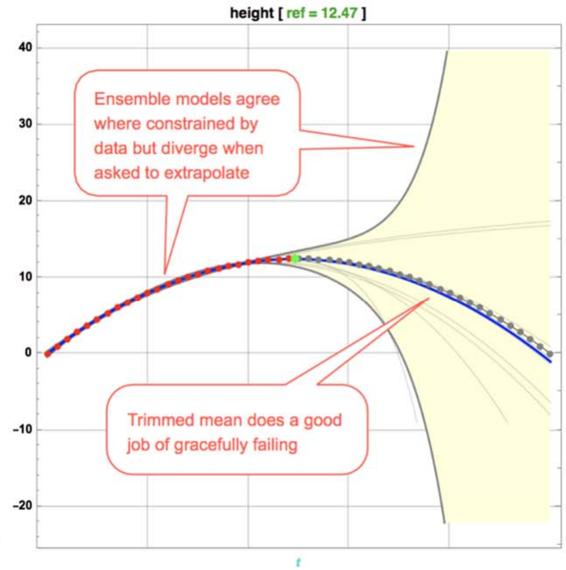
cr7POD

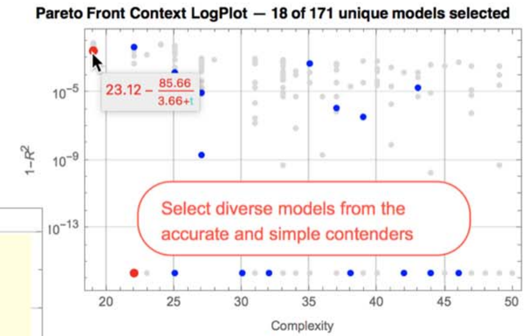
	Complexity	$1-R^2$	Function
1	15	0.197	$0.61 + 1.08\ \text{preCr}^2$
2	22	0.159	$0.24 + 0.16\,(7.48 + \text{male})\ \text{preCr}$
3	22	0.166	$0.40 - 0.27\,(-3.55 - \text{male})\ \text{preCr}$
4	34	0.150	$0.32 + 1.29\ \text{preCr} + \frac{18.06\ \text{male}}{\text{rvolume}} - \frac{18.78}{\text{rweight}}$
5	35	0.141	$0.22 + 0.12\ \text{male} + 1.26\ \text{preCr} + \frac{27.24}{\text{rvolume}} - \frac{30.24}{\text{rweight}}$
6	50	0.138	$5.63\times10^{-2} + (2.04\times10^{-3})\ \text{bw} + 0.12\ \text{male} + 1.21\ \text{preCr} - (4.44\times10^{-4})\ \text{rvolume} + \frac{0.13\ \text{rweight}}{\text{rvolume}}$

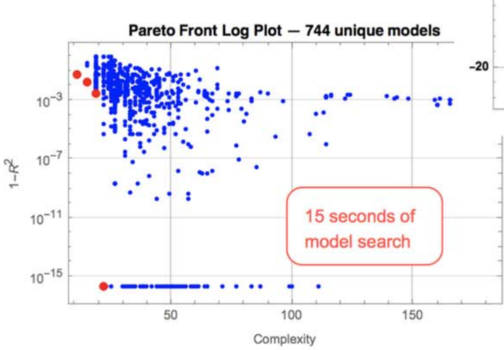


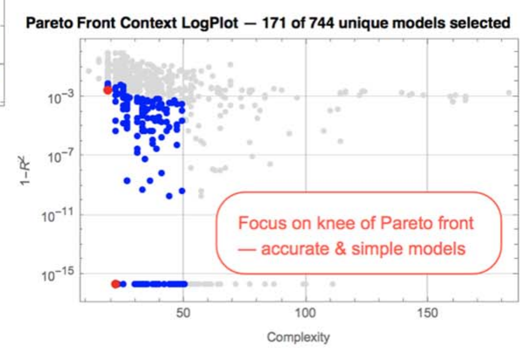
Ensemble in Extrapolation



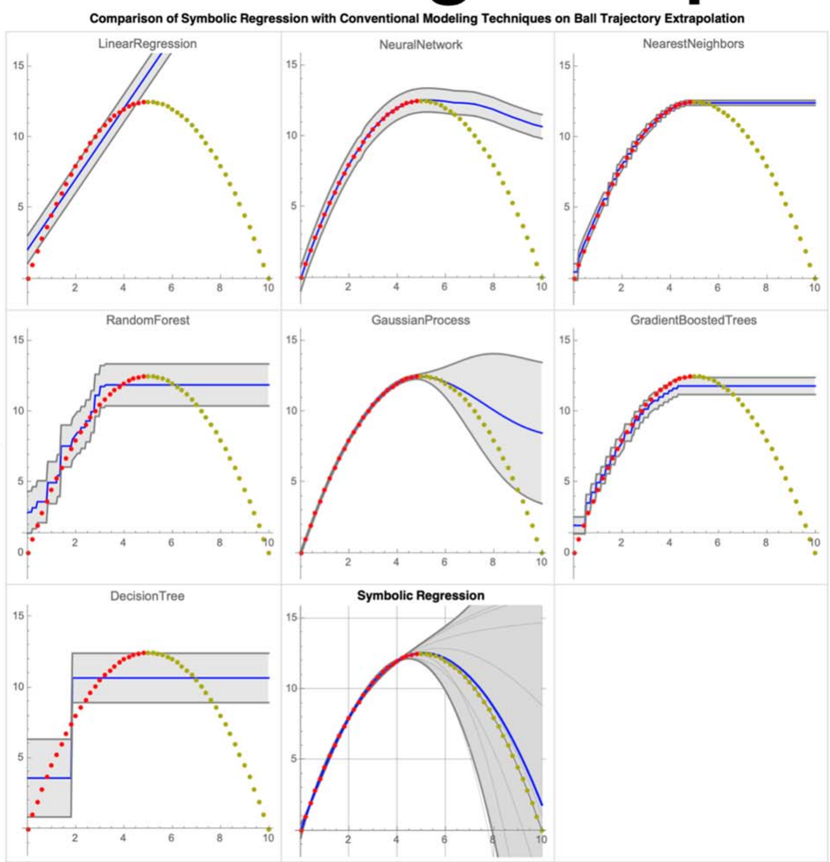




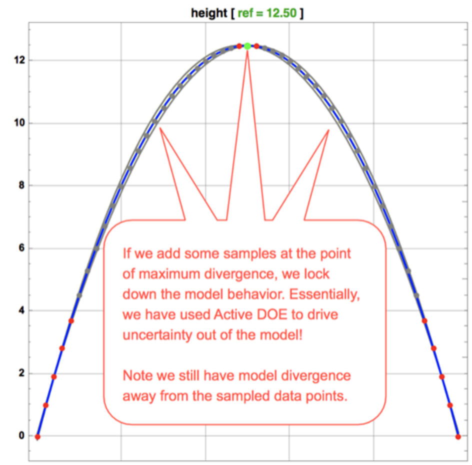


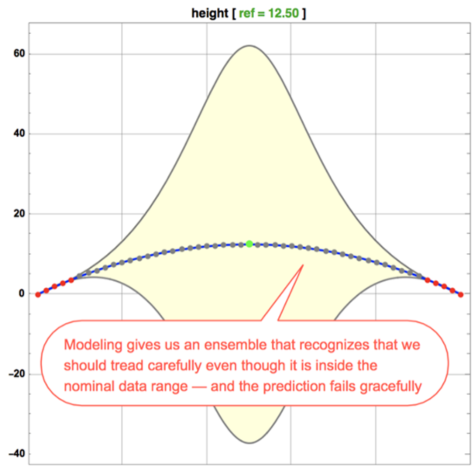


Machine Learning Comparisons



Active DOE



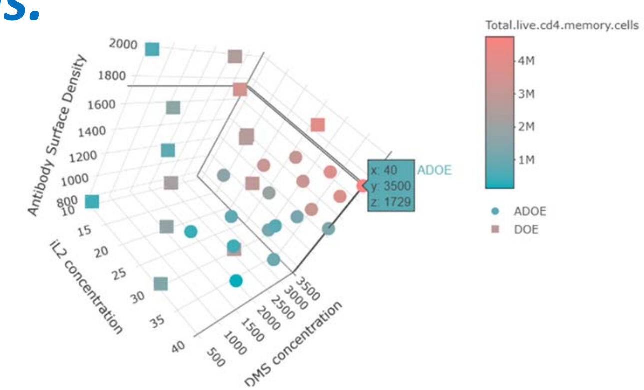


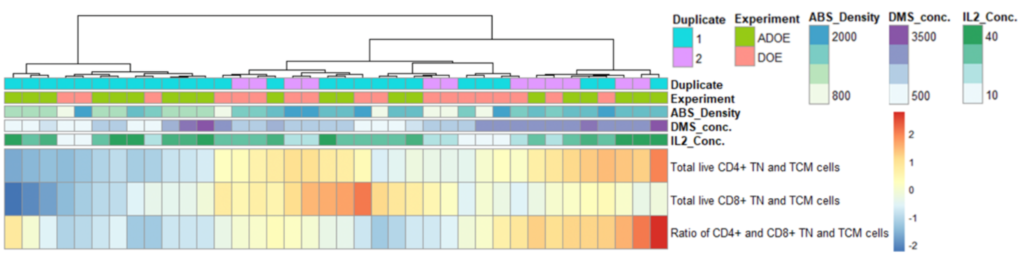
- Extrapolation can happen within the nominal data range
- Trustable models are especially important with multivariate models
- Trustable models are the foundation of **Active DOE** — targeting the next experiment to be the most informative

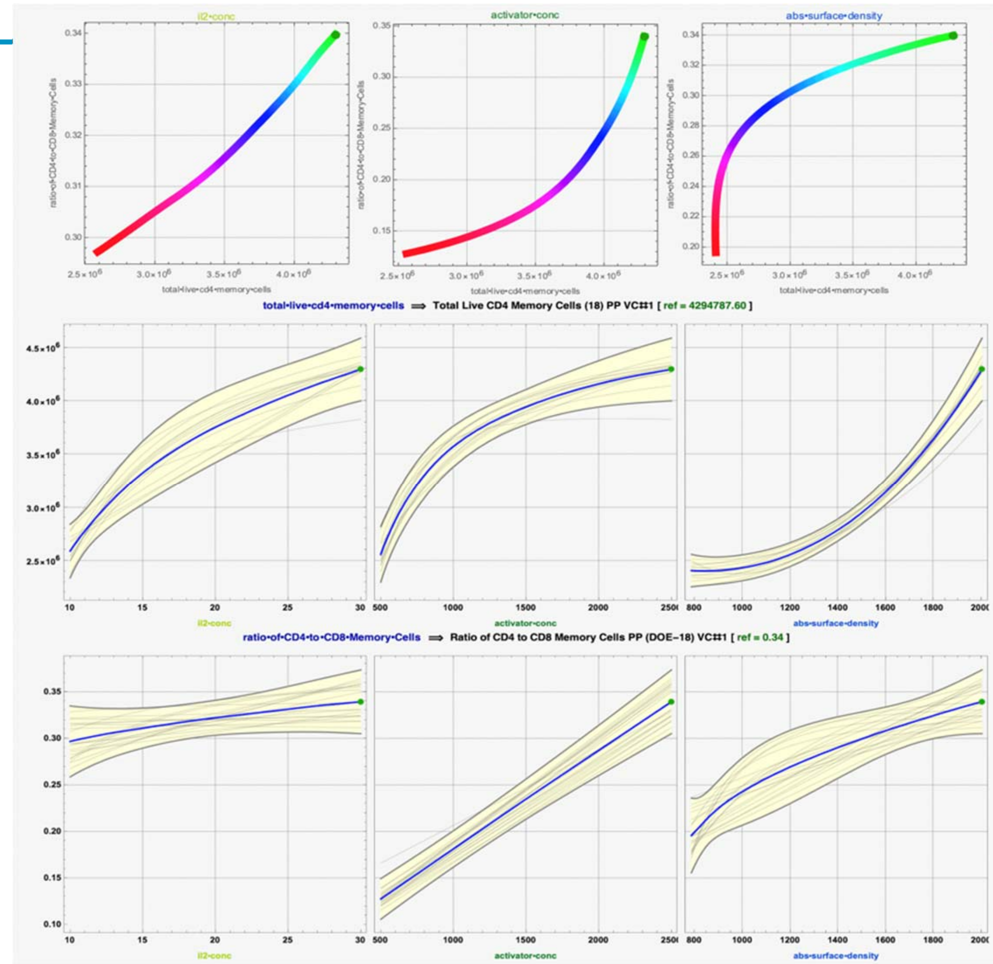
Significant Achievements

*"**Design of Experiments (DOE)** identified optimum conditions for **maximizing memory T cells.**"*

– NSF 2020 Report



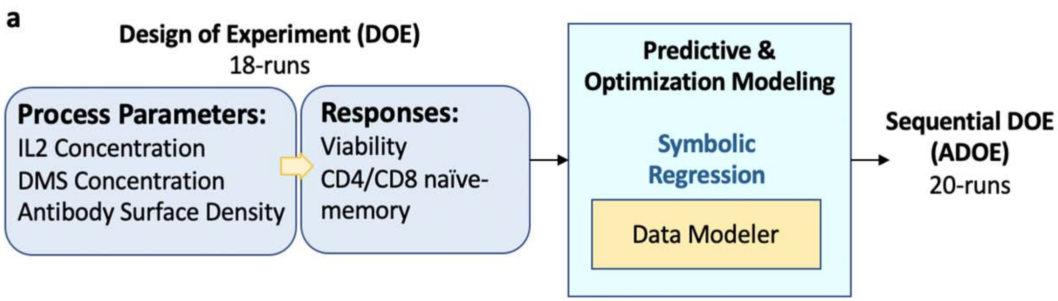




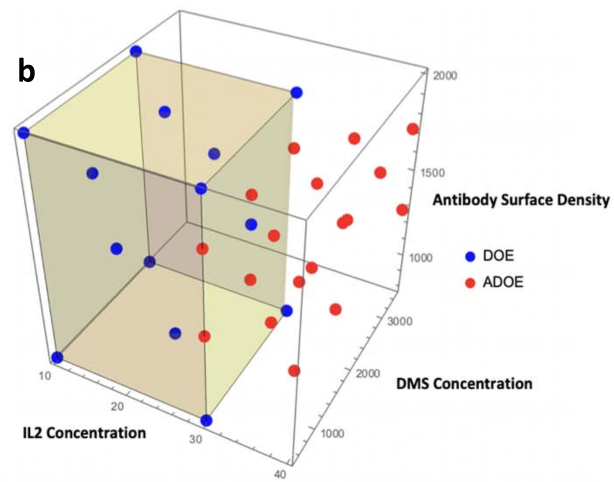
Key achievements

Phase 1: Experimental Process Parameters Optimization

a



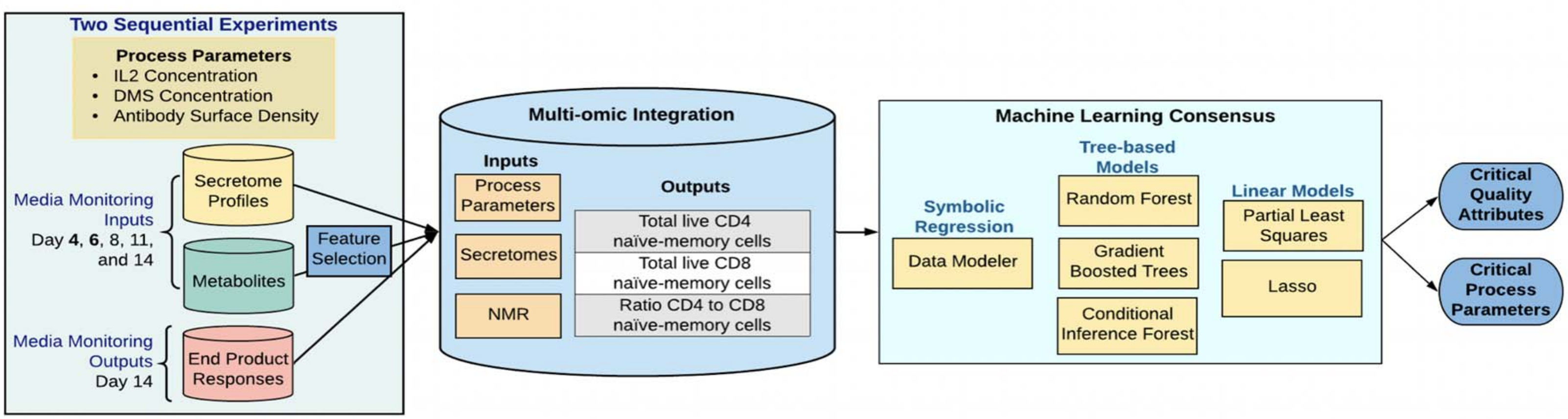
b



Data Challenges
High complexity –
Heterogeneous –
Unknown behavior
Volume of data

Phase 2: Multi-Omics Integrative Approach for Early Predictive Signatures

c



Modeling Challenges
Apriori model structure (linear vs non-linear)
Standalone vs interactive effects
Prediction performance vs overfitting
Interpretability
Computational infrastructure

Predictive Performance

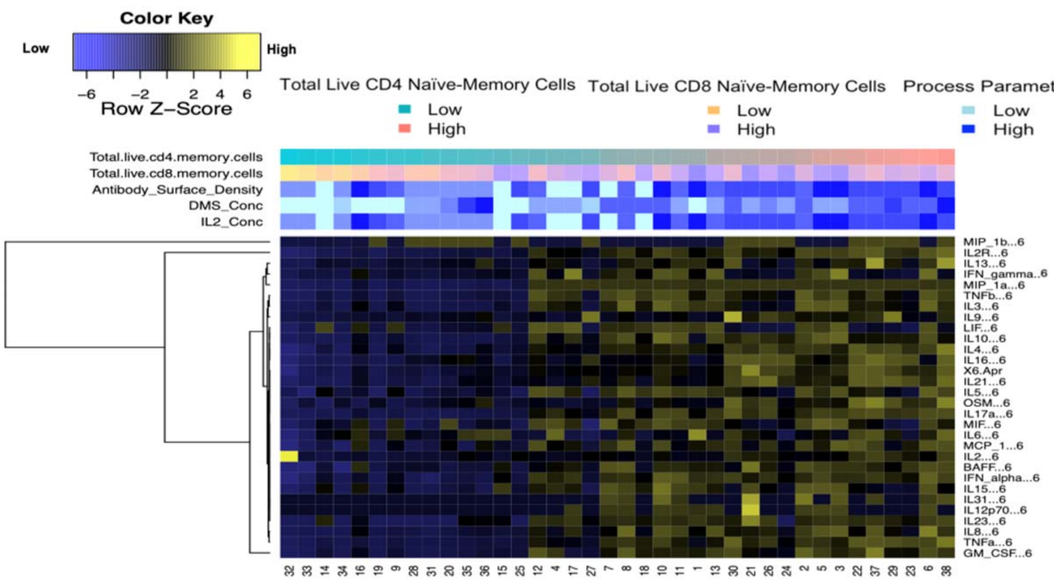
- Machine Learning Comparison
  - Conditional Inference Forest
  - Random Forest
  - Gradient Boosted Trees
  - Symbolic Regression

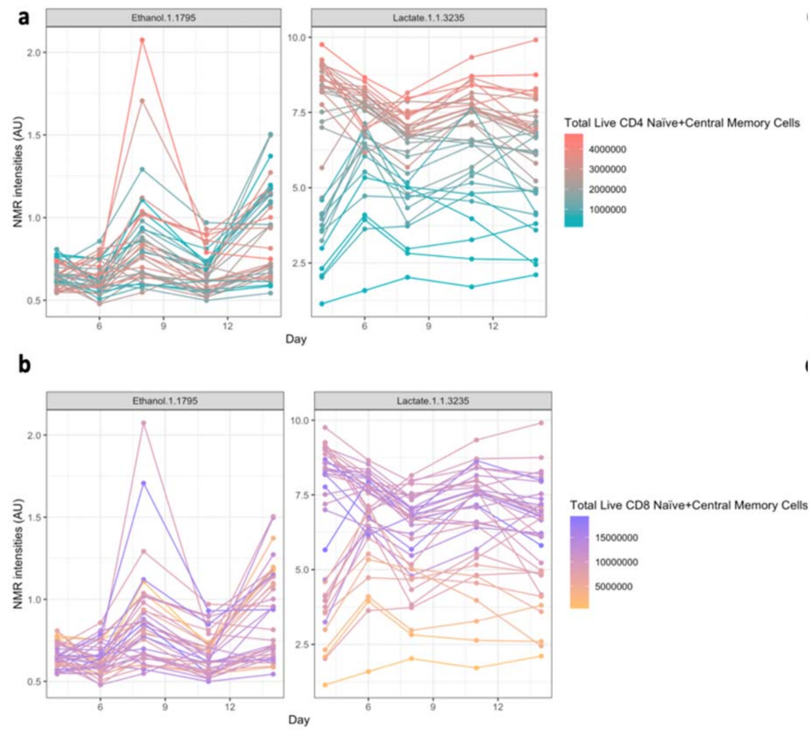
Single and Multi-omics Analysis - R^2 Prediction Performance

Endpoint Responses		CIF	RF	GBT	SR	Lasso	PLSR
Multi-Omics	Total live CD4+ T_N and T_{CM} cells	84%	81%	96%	99%	90%	91%
	Total live CD8+ T_N and T_{CM} cells	47%	39%	86%	93%	80%	61%
	Ratio CD4+/CD8+ T_N and T_{CM} cells	75%	77%	88%	98%	90%	82%
Single-Omics (Process Parameters & NMR features 4)	Total live CD4+ T_N and T_{CM} cells	79%	76%	93%	97%	-	-
	Total live CD8+ T_N and T_{CM} cells	47%	36%	73%	89%	-	-
	Ratio CD4+/CD8+ T_N and T_{CM} cells	76%	74%	85%	96%	-	-
Single-Omics (Process Parameters & NMR Day 6)	Total live CD4+ T_N and T_{CM} cells	64%	55%	90%	97%	-	-
	Total live CD8+ T_N and T_{CM} cells	43%	37%	83%	86%	-	-
	Ratio CD4+/CD8+ T_N and T_{CM} cells	40%	84%	69%	96%	-	-
Single-Omics (Process Parameters & Cytokines Day 6)	Total live CD4+ T_N and T_{CM} cells	84%	91%	93%	96%	-	-
	Total live CD8+ T_N and T_{CM} cells	44%	74%	87%	92%	-	-
	Ratio CD4+/CD8+ T_N and T_{CM} cells	76%	66%	88%	96%	-	-

Significant Achievements

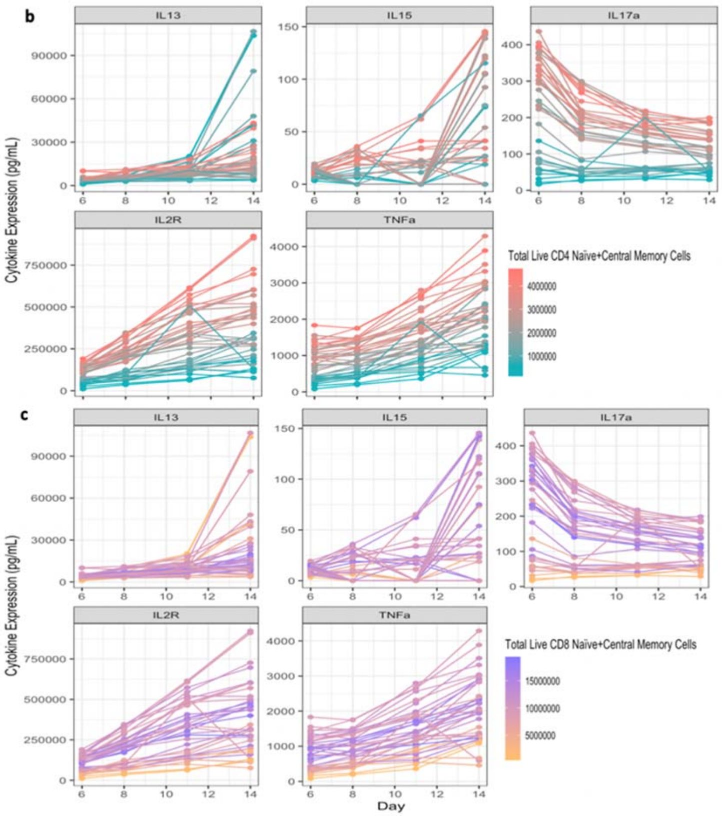
- "***Machine learning** was performed to correlate and identify critical process parameters and **early secretomic and metabolomics-features** that are **predictive of product quality**.*" – NSF 2020 Report

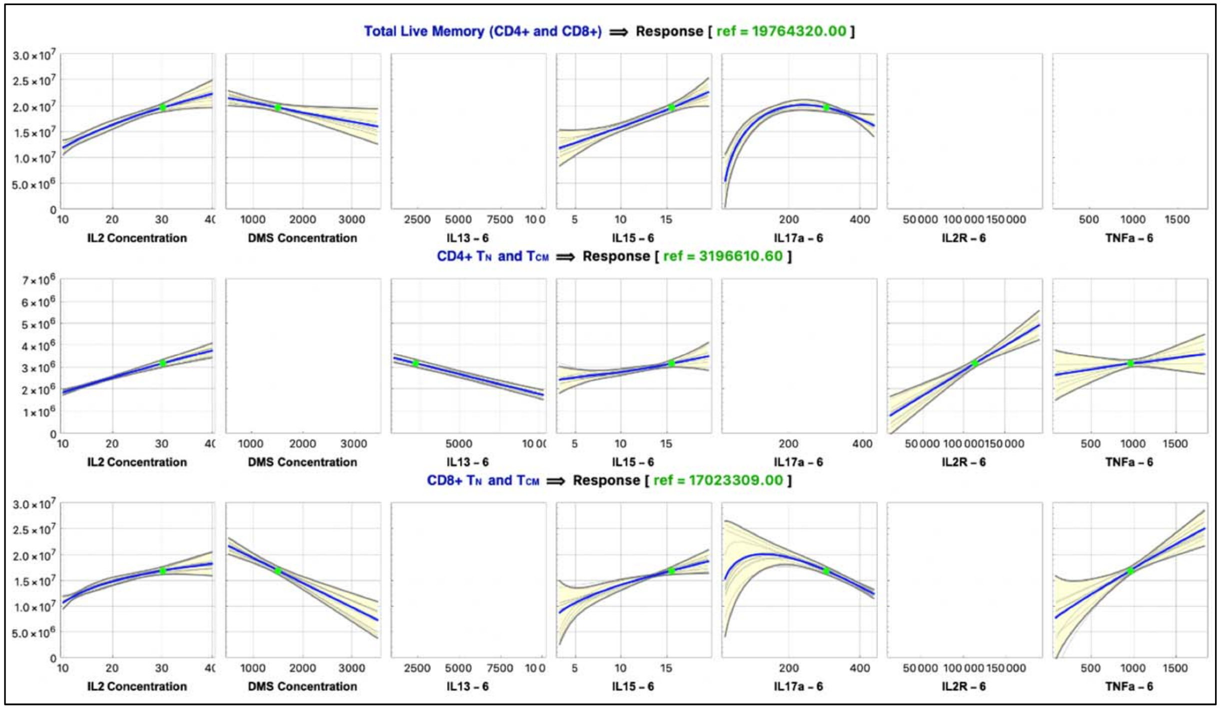




Early Prediction of Total T_N and T_{CM}

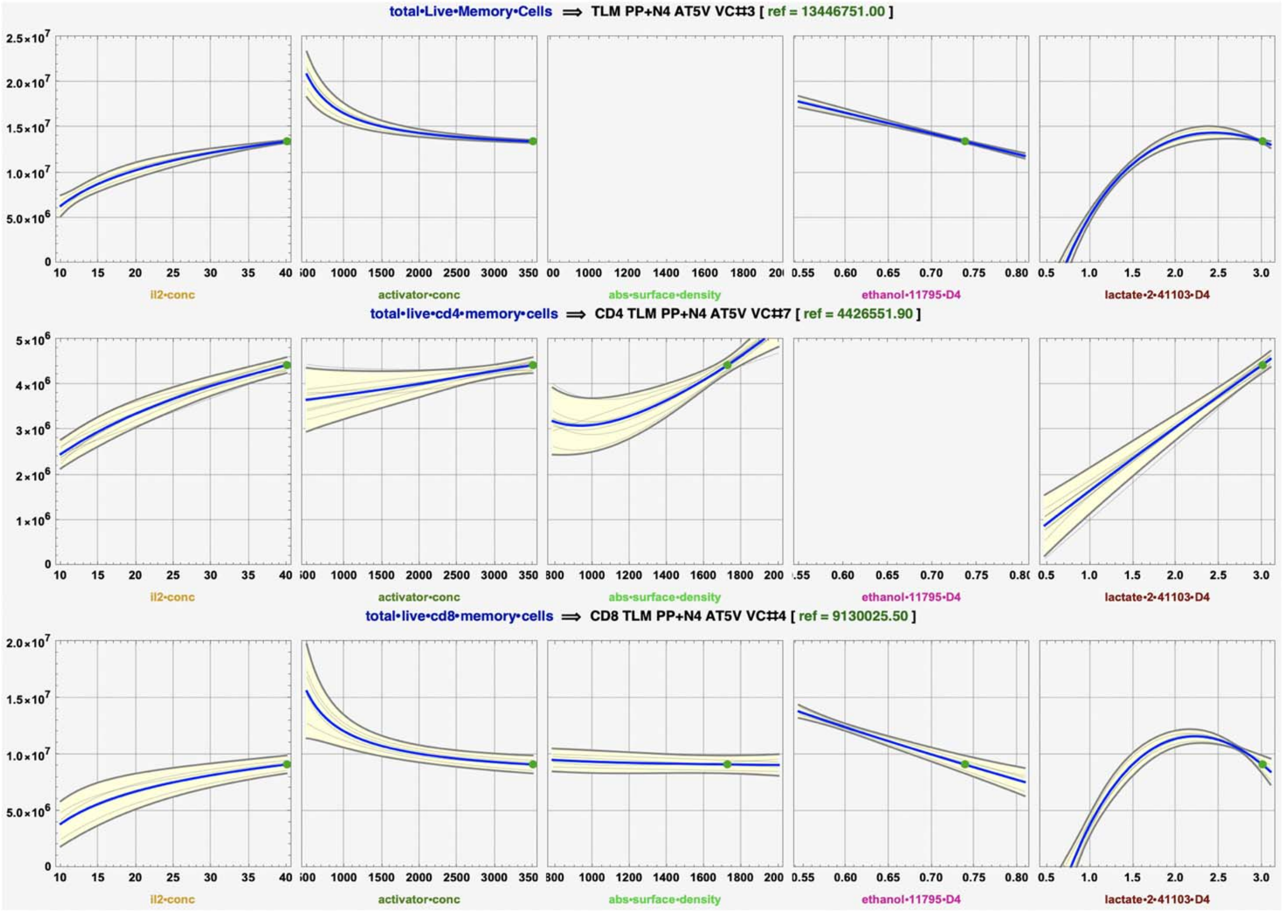
- Multi-Ohmic Prediction Profiles→ Process Parameters + Cytokines (Day 6)



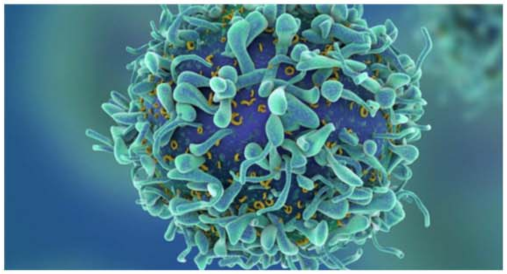


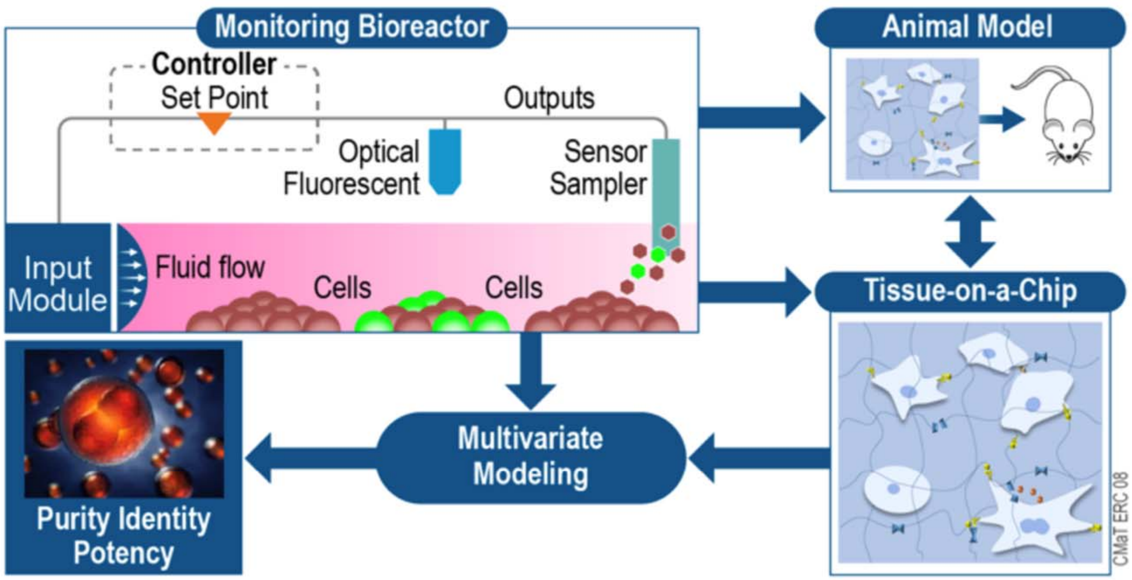
Early Prediction of Total T_N and T_{CM}

- NMR Media Analysis
  - Day 4



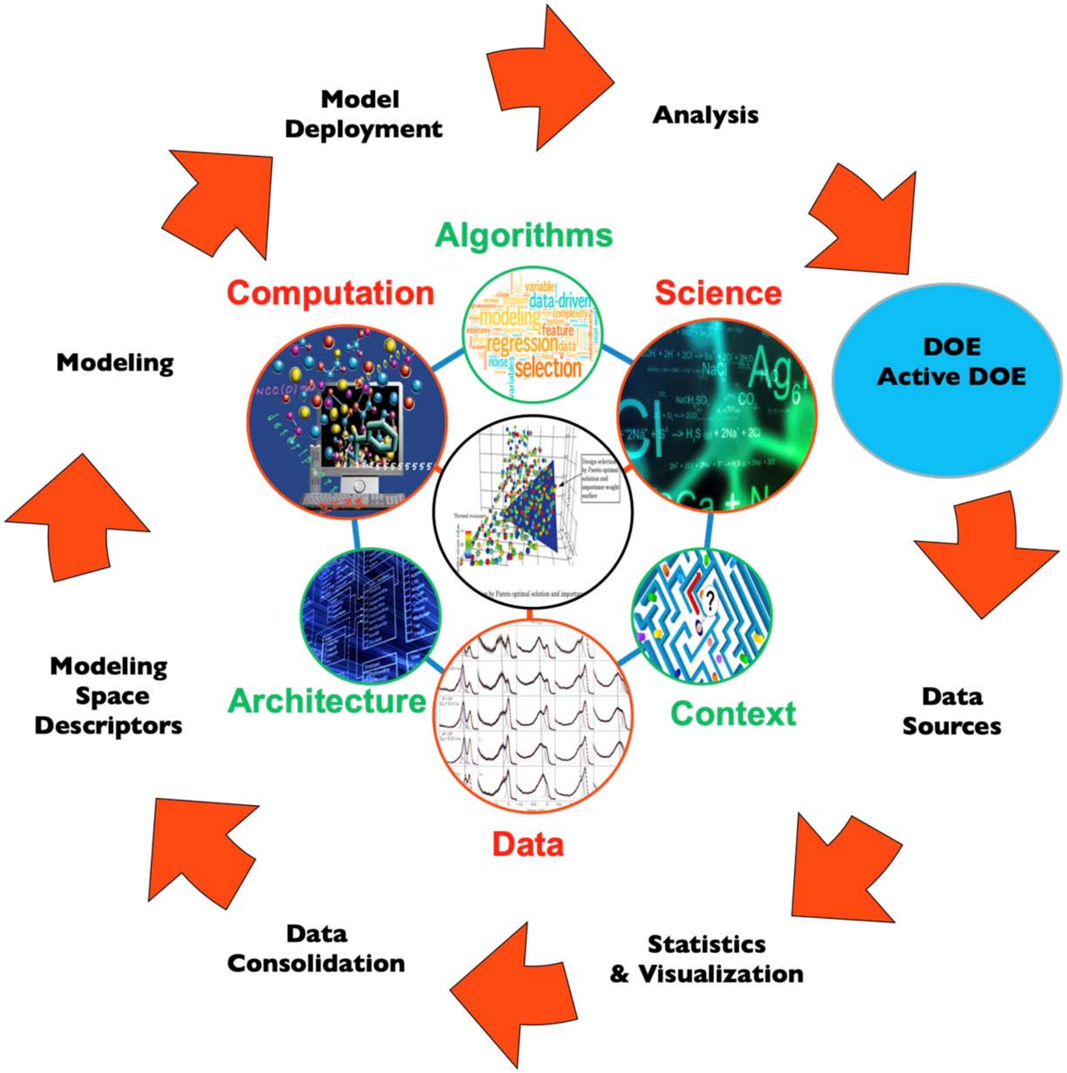
Dynamic Sampling Platform





- What are the key factors?
- How should we design the system?
- Where should we look for better solutions?

Informatics Critical for Systems Integration & Active Learning

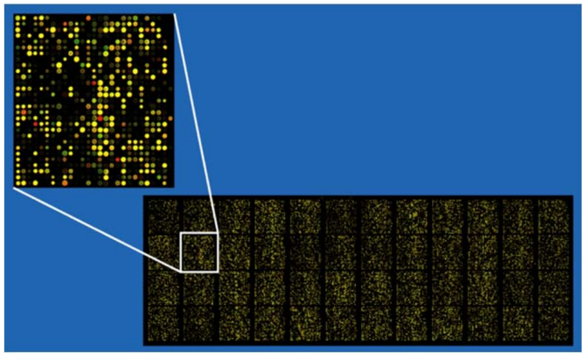


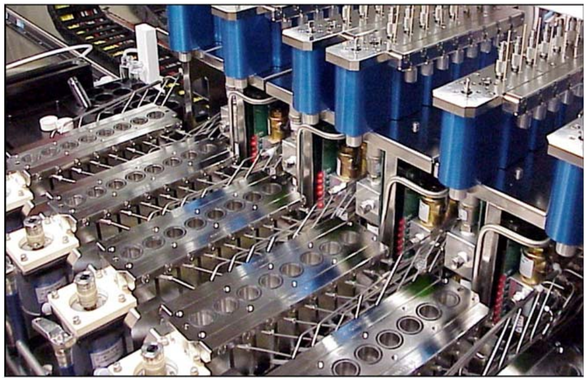
Moving Forward… What is Your Analysis Objective

Stage	Objective	Analysis
Discovery	New Insight/ Knowledge?	Identify new modes of action, therapies, formulations & process technologies
Development	How does it work?	Develop multi-ohmic system level structure/ property/ process relations and robust IP
Design	What & how to produce?	Design products, processes, testing and release protocols
Validation	What to control?	Identify and verify CQAs & CPPs
Optimization	How to optimize?	Determine optimal design, formulations, process conditions & yields Identify what to improve/ optimize based on clinical outcomes
Process Control	How to control?	Identify what to control & control limits Define and implement real time control and lot release
Supply Chain	How to maximize?	Determine optimal supply logistics plans to maximize working capital & patient outcomes
Commercialization	What costs? Pricing?	Determine costs and pricing
Strategy	Which scenario?	Examine potential scenarios using what-if exploration tools

What Questions Do You Need to Answer?

- Variable Selection & Relationships
  - Which variables matter?
  - What variable combinations are useful?
  - Are there important metavariables?
- Prediction
  - Can we accurately predict performance?
  - Can we utilize for real time control?
- Optimization & Deployment
  - Can we build robust emulators for what if and design?
  - Can we simultaneously optimize multiple KPIs?
  - Can we build active learning systems?
- Risk Management
  - Outlier detection & assessment?
  - What is our extrapolation trust metric?
- Insight & Understanding
  - Is this novel? Is it patentable?





Moving Forward…

- What data do you have?
- Exploit it! Good & bad!!!
- Lives can be saved!
- Act now!!!

Thank you!

Acknowledgements

CMaT: Valerie Y. Odeh-Couvertier, Nathan J. Dwarshuis, Maxwell B. Colonna, Bruce L. Levine, Arthur S. Edison, Krishnendu Roy, Wandaliz Torres-Garcia, Andres Garcia, Andrei Federov

This material is based upon work supported by the National Science Foundation under Grant No. EEC-1648035. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.