John Baldoni
GSK

Frontiers of Predictive Oncology and Computing

October 17-19, 2017

SUNY

New York City

Topics

- Scale of data available
- Challenge of accessing and using that data
- Some examples

ATOM Consortium

Accelerating Therapeutics for Opportunities in Medicine

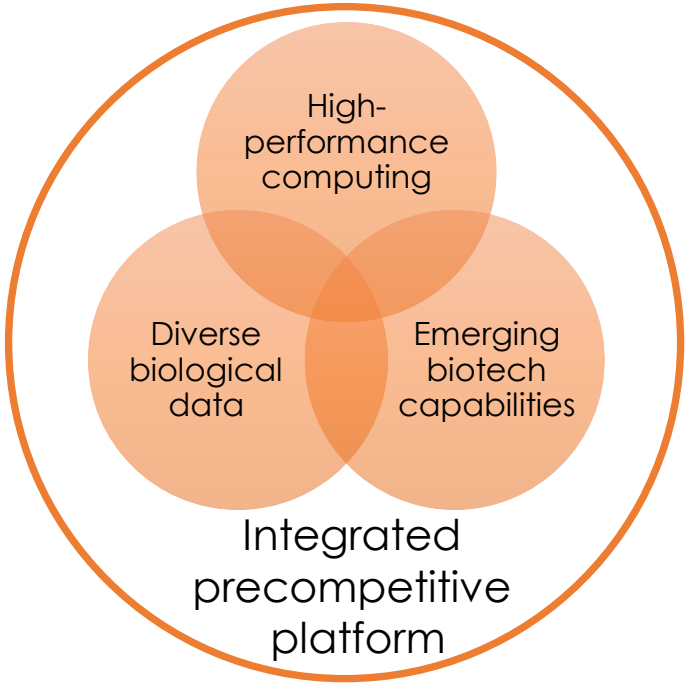
Mission

To accelerate the development of more effective therapies for patients

Vision

A new starting point:
Transform drug discovery from a slow, sequential, and high-failure process into a rapid, integrated, and patient-centric model

How?



3 year goal

Starting from a validated oncology target, deliver a patient-ready cancer therapy in <1 year, a process that currently takes an average of 6 years

ATOM Consortium Membership

Actively encouraging additional members who share in the vision

Elements of membership:

1. Augment knowledge and capabilities
2. Senior leadership participation
3. Data, equipment, and/or resources
4. Staff with expertise to work at the ATOM lab
5. Financial support for the consortium
6. Commitment to a new pre-competitive balance

Examples of contributions from founding members

gsk	Lawrence Livermore National Laboratory	Frederick National Laboratory for Cancer Research	UCSF	...
2M compounds and associated data from terminated projects and screening collections	Access to DOE supercomputers for purposes of carrying out Consortium related activities	Scientific expertise in cancer, computational chemistry and biology, data science, and predictive oncology	Access to UCSF facilities, expertise from UCSF faculty, cancer center	

Access to additional items

- **Study Reports on Safety/Tox, ~345 compounds - 5000+ reports**
  - **SAPAD table entries for compounds that were terminated because of animal or human safety issues, indicating the tissues and pathologies observed.**
- **Crystal structures (ligand/protein): 1,356 in review. Will include mostly targets along with a smaller number of off target related assays**
- **Human clinical reports for 68 compounds**
  - **EKG data and other outcomes will be valuable for linking larger collections of experimental data with human outcomes**

What is GSK doing to be more data enabled…?

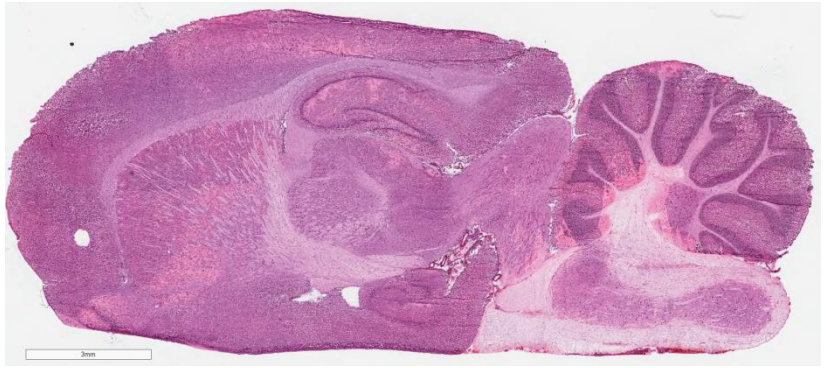
- **GSK has appointed a Chief Data Officer and a built a department to ensure that historic and future data are in a searchable format.**
- **We have built our internal compute and data transfer infrastructure to enable deep analytics, from target to clinic.**
- **We have hired people from traditional and nontraditional Pharma backgrounds to accelerate use of data in problem solving.**
- **We have put in place collaborations to jumpstart our efforts to use data and deep analytics to solve problems.**
- **We are a founding member of ATOM, a precompetitive govt, academic and pharma collaboration to share data for algorithm development.**

Examples…

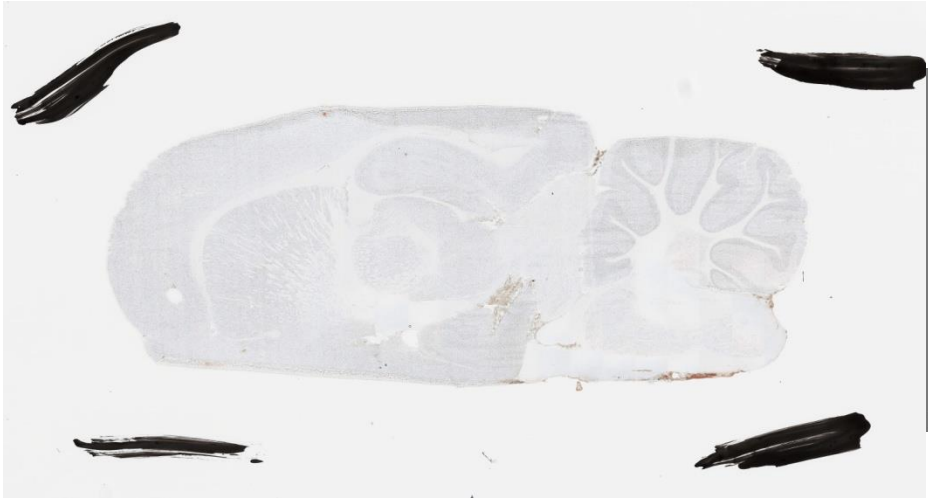
Objective: co-registration MALDI images and optical images

- **Align high resolution whole slide histology images (H&E stained) and MALDI (optical) image.**

- **Example below: rat brain tissue slices**

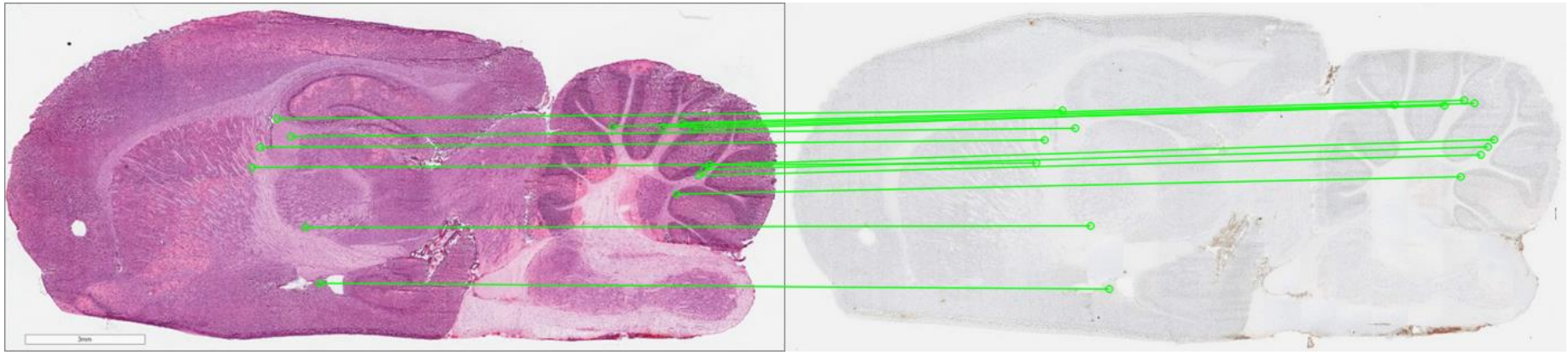


Whole Slide Histology Image



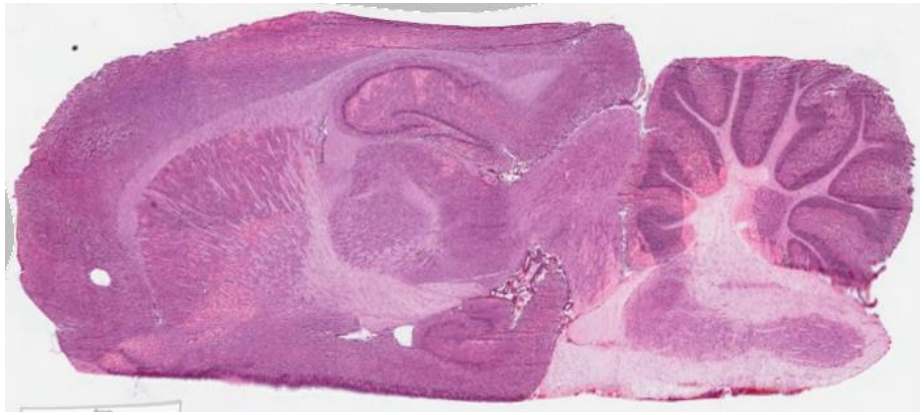
Optical image of MALDI slide

Rigid Transformation

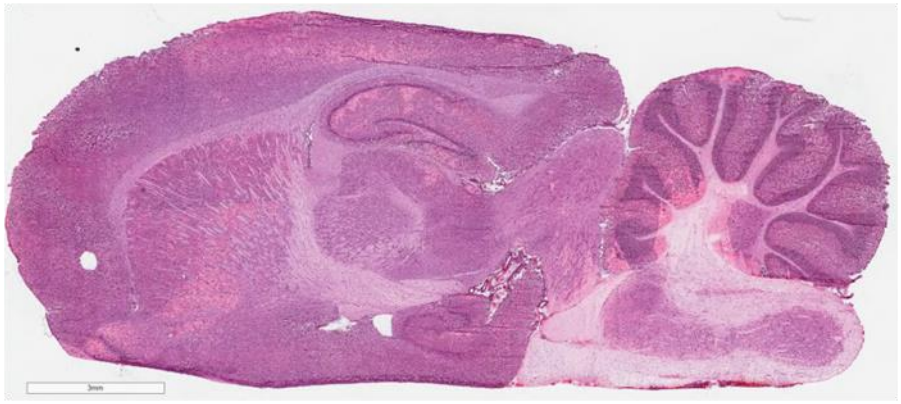


Non-rigid transformation

Optical Image of MALDI

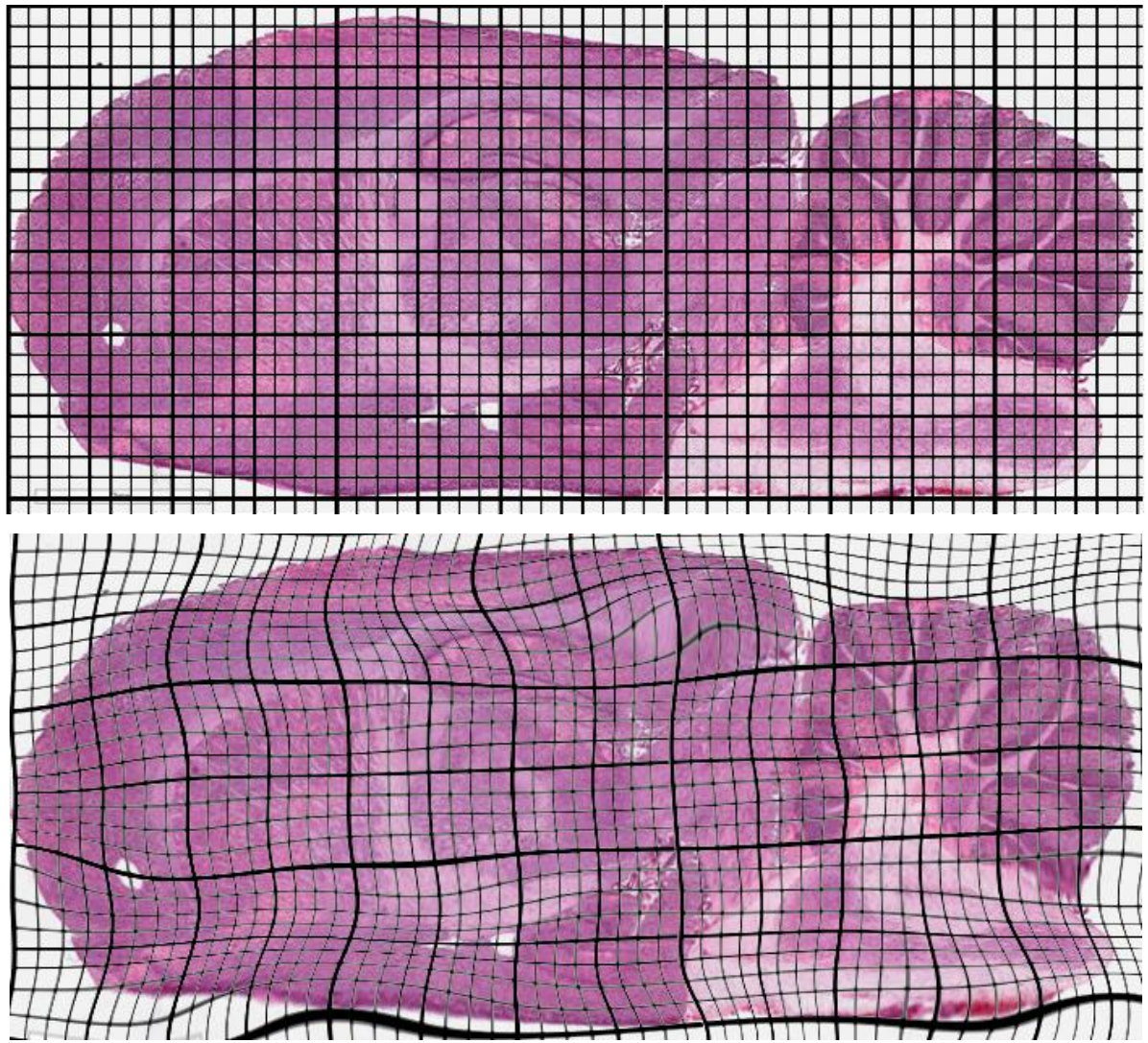


Non-rigid registered Image



Whole Slide Histology Image

Deformable transformation: visualization

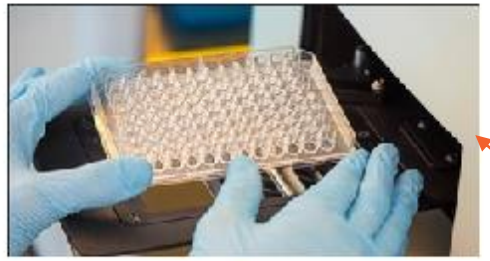


Objective: A broader interrogation of cell biology through automation

Compound

Cell Lines

$>10^6$ cells per screening plate



GSK scientist:
20-30 plates/day
Per plate: >25GB of image data

Image Analysis

Quantification of visual phenotypes enables:

- Design of cellular imaging assays
- Understanding mechanisms of action
- Identification of pathways
- Detection of toxicity

Two big enablers

Hardware, high throughput automation

Automated image analysis routines

Next phase

Data driven image feature extraction

Automated image analysis workflow

Automated phenotype characterization

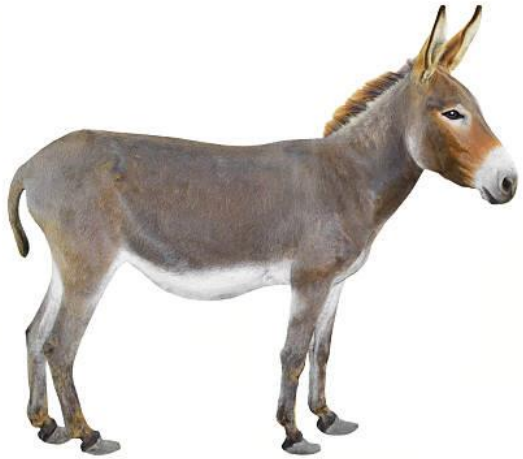
Abraham, V. C. et al. (2004). *Trends in biotechnology*, *22*(1), 15-22.

Dealing with unseen phenotypes

Minimize learning unimportant features instead of true discriminators

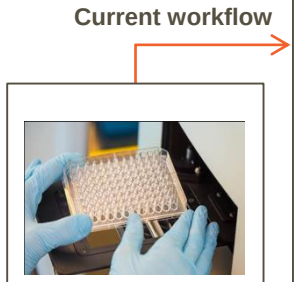
Training set

Unseen test set



Comparing workflows

Handling of complex images, increased objectivity and time saving



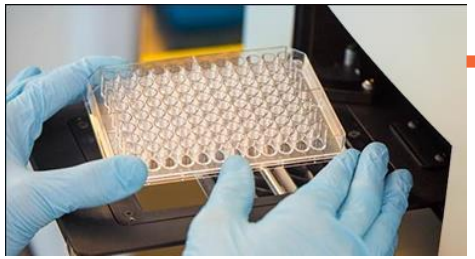
Current workflow

Proposed workflow

	Image complexity	Designing features	Processing	Results
Columbus/HCP	1-2 fluorescent channels	• Subjective human-defined features • >3 hour human effort	• ~5 hours per plate • Process scales with number of phenotypes	• Binary classification of phenotypes, cannot fingerprint A or B B or C
	>2 fluorescent channels • Infeasible			
Deep Learning	1-2 fluorescent channels	• Objective data-driven feature discovery • <1 hour human effort	• <30 min per plate • Independent of number of phenotypes	• Multiclass classification to generate fingerprint in single pass • Can identify known phenotypes with >99% accuracy A 12% B 82% C 6%
	>2 fluorescent channels	• Identical approach		

High Content Imaging – Deep Learning Workflow

Two proposed pipelines and use cases



Read in data from well plate

Image acquisition

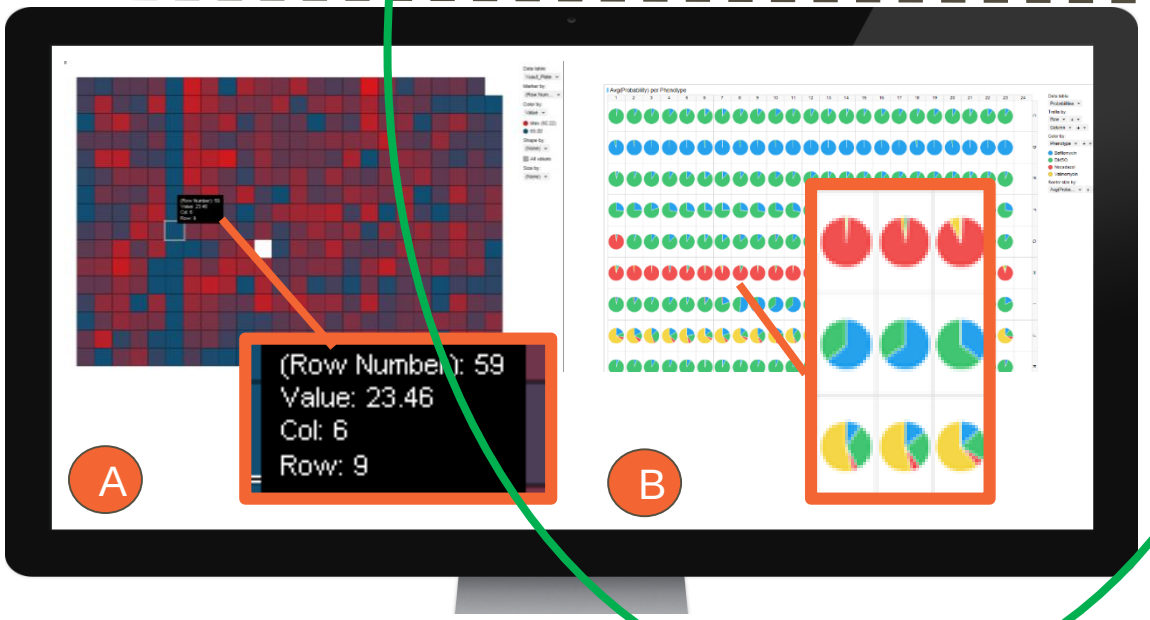
Simply provide a control plate to automatically train a model

A

Use a model trained on a library of phenotypes to perform phenotypic fingerprinting

B

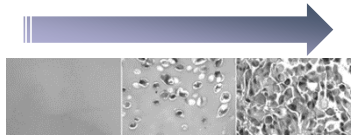
Image processing



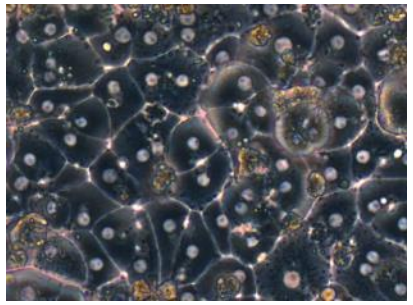
Output

Bioprinting ExVive™ Human Tissues

Translational human relevant models of disease

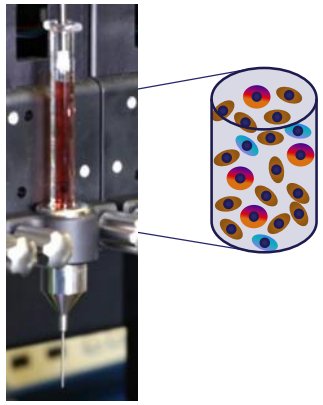


Up to 100% cellular



Human Cells

- Primary or iPS cells
- Normal or diseased

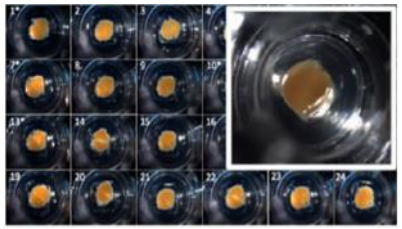


NovoGel® Bio-Ink

- Cell mixture
- Proprietary media
- Optional temporary matrix

NovoGen Bioprinter® Platform

- Biocompatible
- Multimodal
- Spatial control



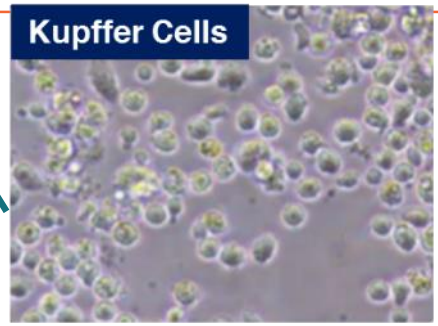
Human Tissues

- Reproducible
- Scalable
- 100% cellular

Bioprinting: The automated fabrication of a tissue through the spatially-controlled deposition of cells and/or cell-containing materials in user-defined, geometric patterns.

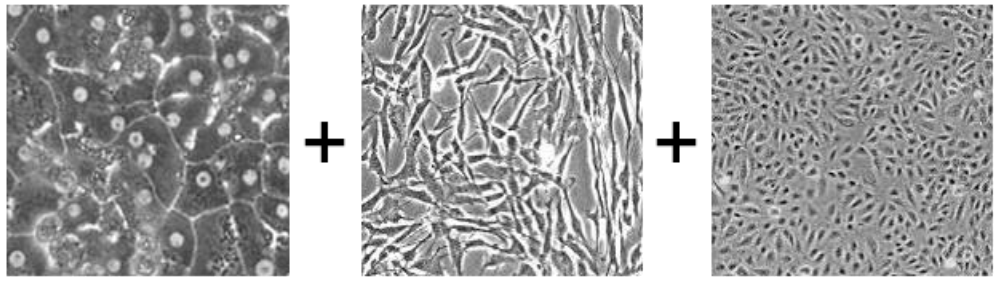
ExVive Human Liver Tissue

- Fully human, multicellular structure
- Large size (2mm^2 X 0.5mm; >10^6 cells)
- Compartmentalized architecture
- Sustained function and viability

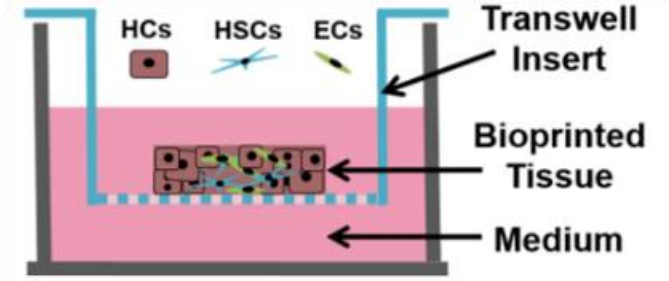


(+) Optional

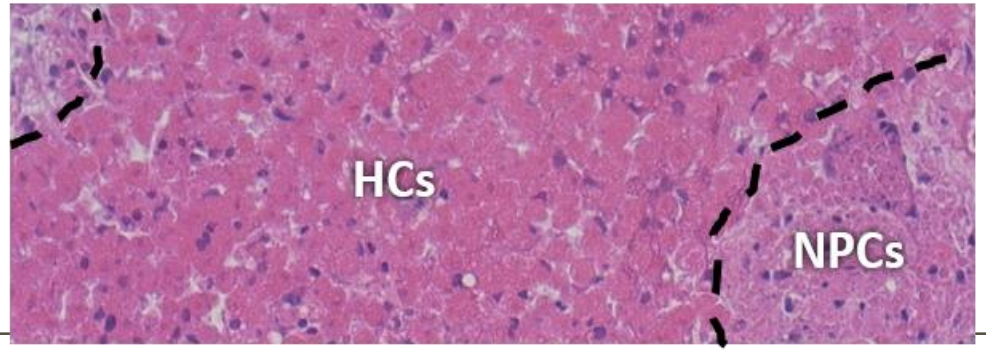
Hepatocytes + Stellate Cells + Endothelial Cells



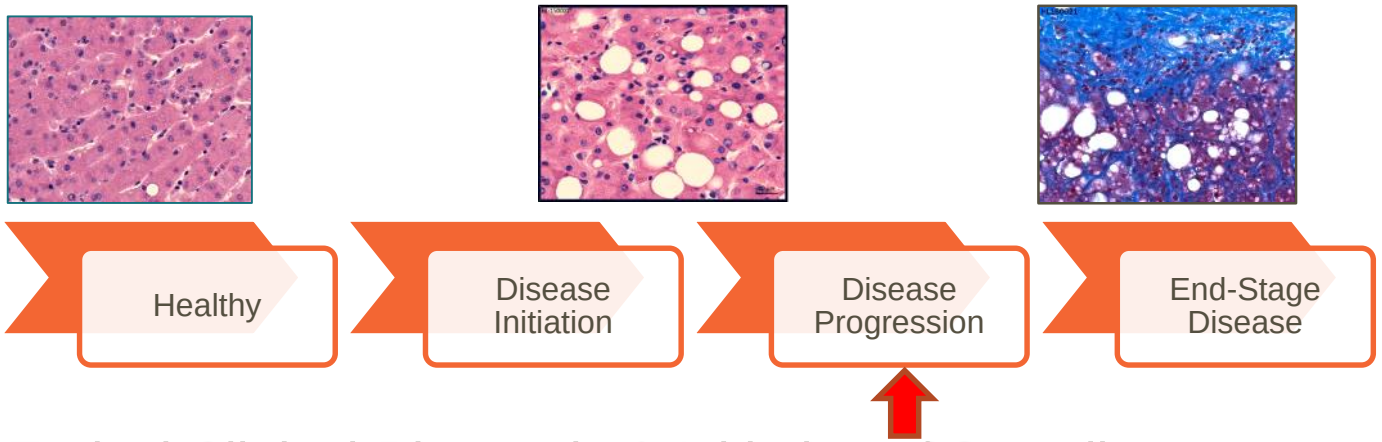
Transverse Cross Section



C



3D Human Liver Tissue Models may bridge translational gap in NASH with fibrosis



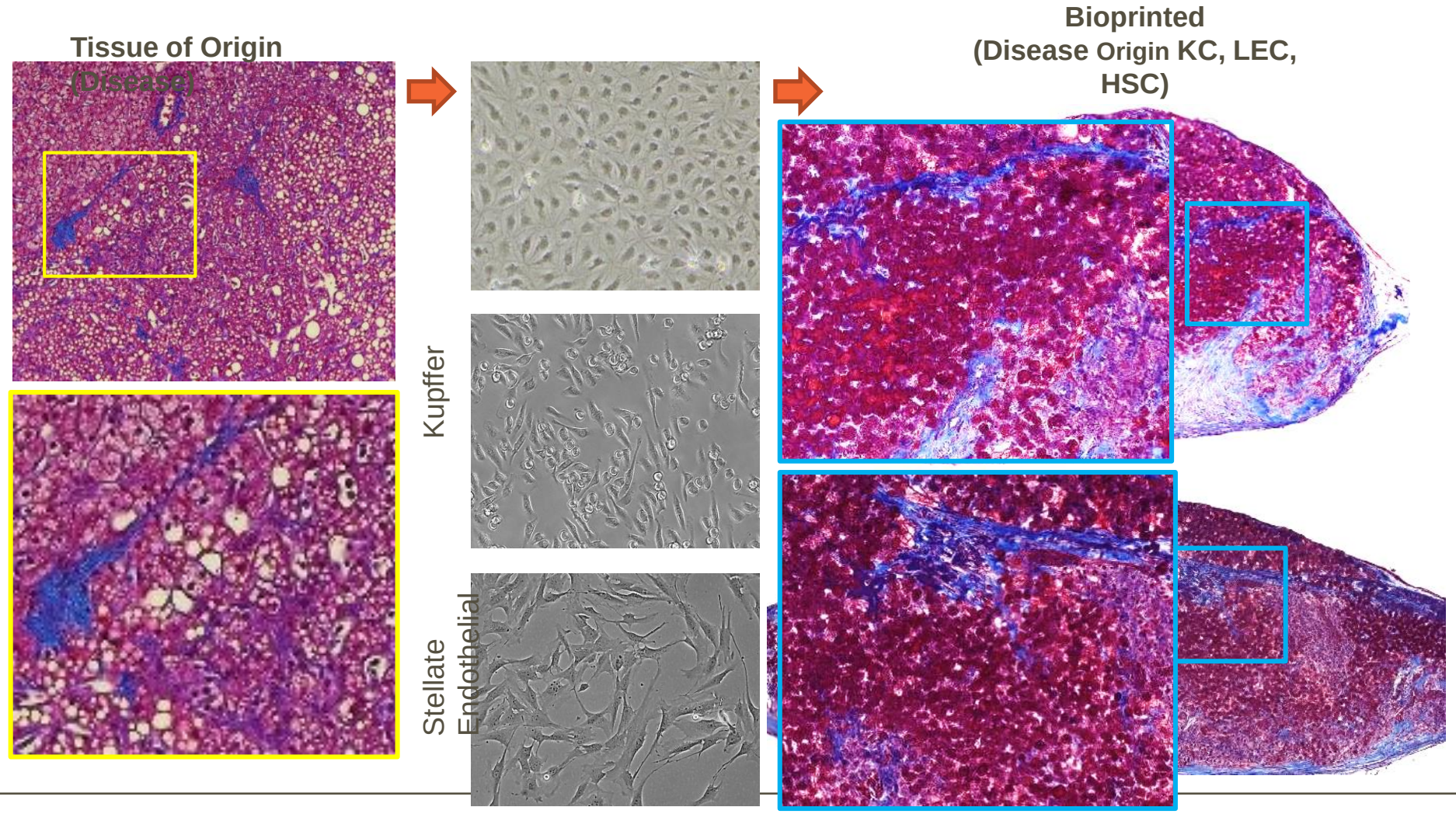
Typical Clinical Diagnosis & Initiation of Sampling

Window of discovery & intervention using clinical samples

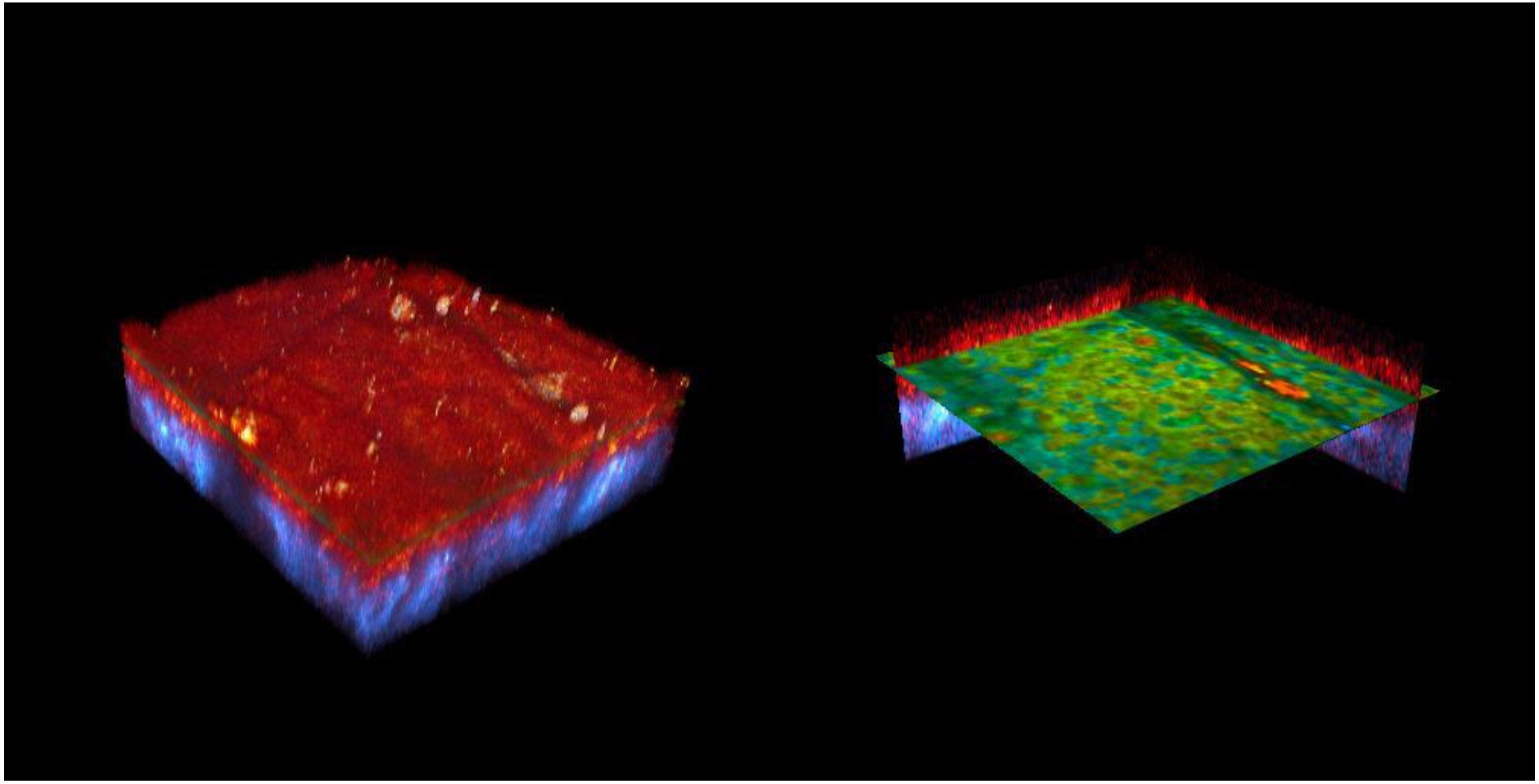
ExVive™ Human Liver can be analyzed at any stage of disease

- *Induced by high fat / high sugar 'diet'*
- *Driven by producing tissues from Disease-Origin Cells*

Tissue-of-origin features are recapitulated in bioprinted tissues

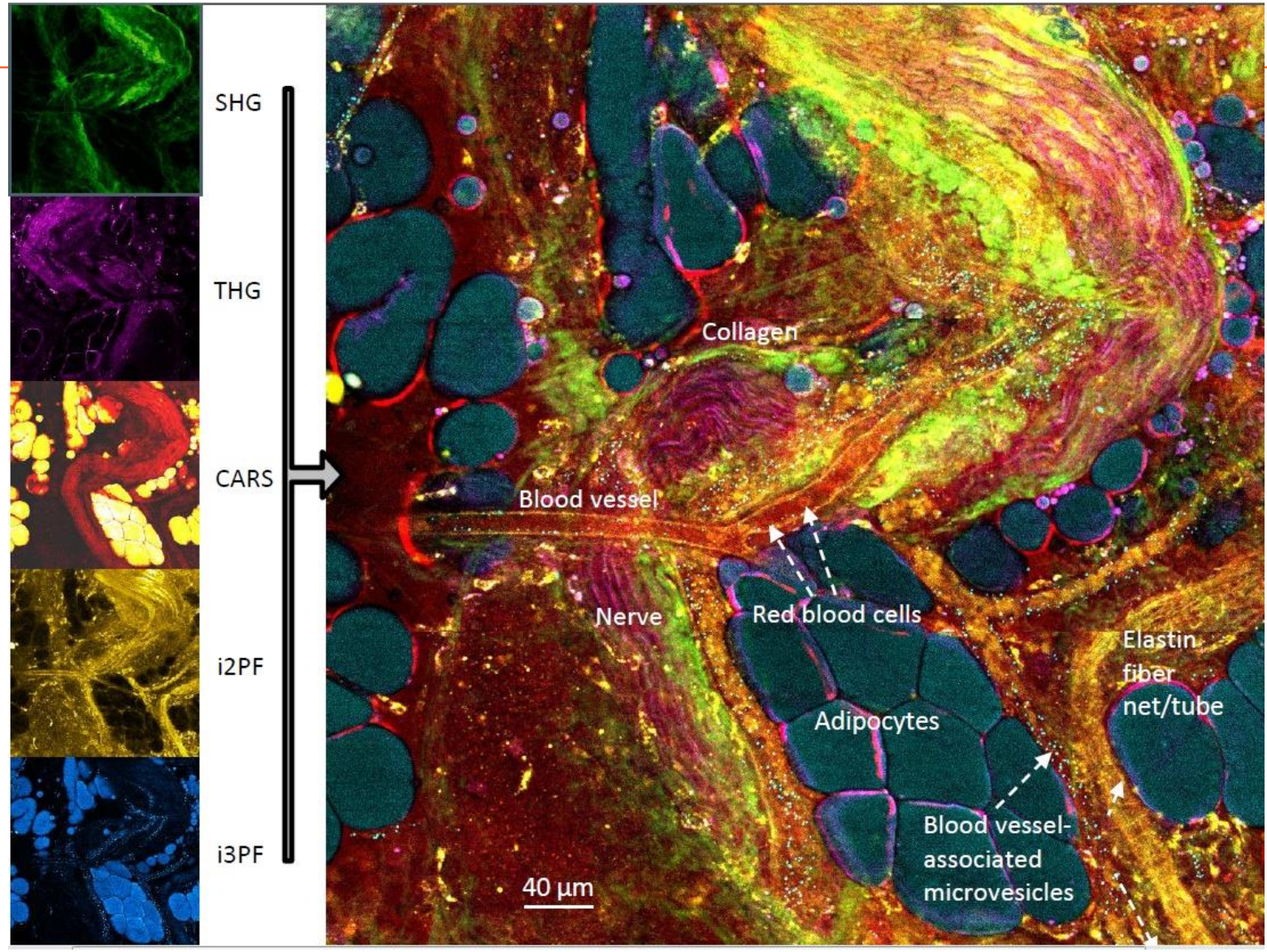


In-Vivo Optical Biopsy of Mouse



Bower AJ, Li J, Arp Z, Marjanovic M, Zhao Y, Chaney EJ, Boppart SA. Longitudinal in vivo tracking of adverse effects following topical steroid treatment . Experimental Dermatology, 25:362-367 2016.

Combination of Methods – The Power of Multi-Modal



Tu H, Liu Y, Marjanovic M, Chaney EJ, You S, Zhao Y, Boppart SA. Concurrence of extracellular vesicle enrichment and metabolic switch visualized label-free in the tumor microenvironment . Science Advances, 3:e1600675 2017.