

Multimodal data integration in cancer from real world and experimental data

Sohrab Shah, Chief Computational Oncology
Dept of Epidemiology and Biostatistics



Memorial Sloan Kettering
Cancer Center

COI:

Shah:

Stockholder in: Imagia Canexia Health
Consulting fees: AstraZeneca Inc

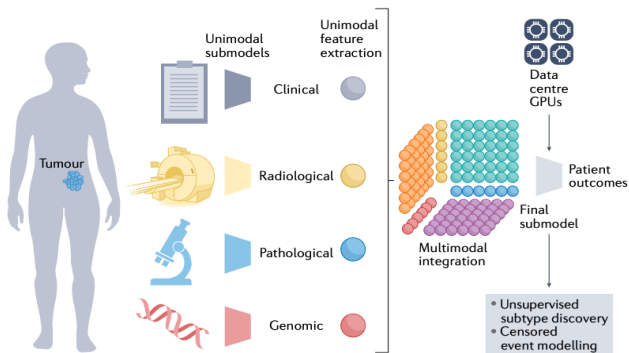
Frontier areas of research in Computational Oncology

Insights from Real World Data

Multimodal data integration

Cohort derivation and comparison

Patient stratification and predictive models



Boehm et al *Nat Rev Can* 2021

Vanguri et al *Nat Cancer* 2022

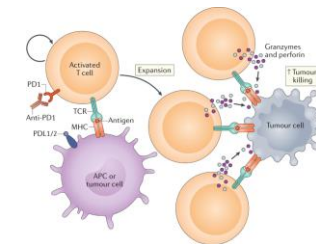
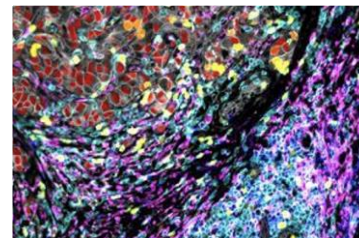
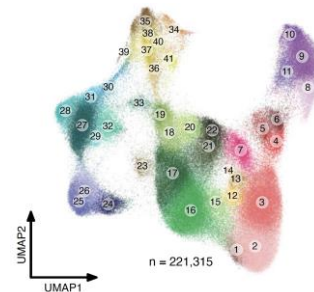
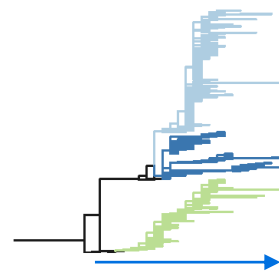
Boehm et al *Nat Cancer* 2022

Cancer evolution & drug resistance

Tumor microenvironments & Cellular phenotypes

Tumor-immune interactions

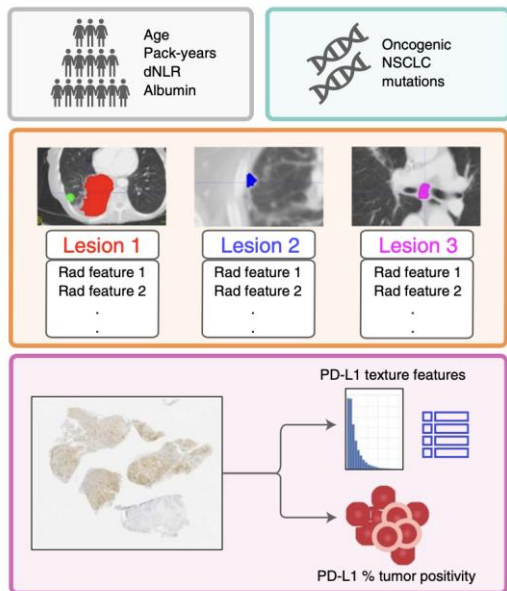
Multi-omic data integration



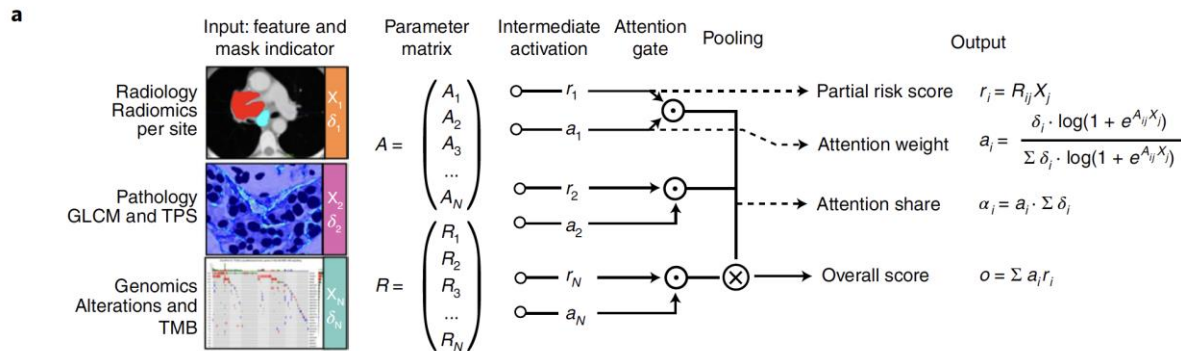
Warren Alpert
Center for
Computational
Pathology

Niki Schultz: Cancer Data Science Initiative

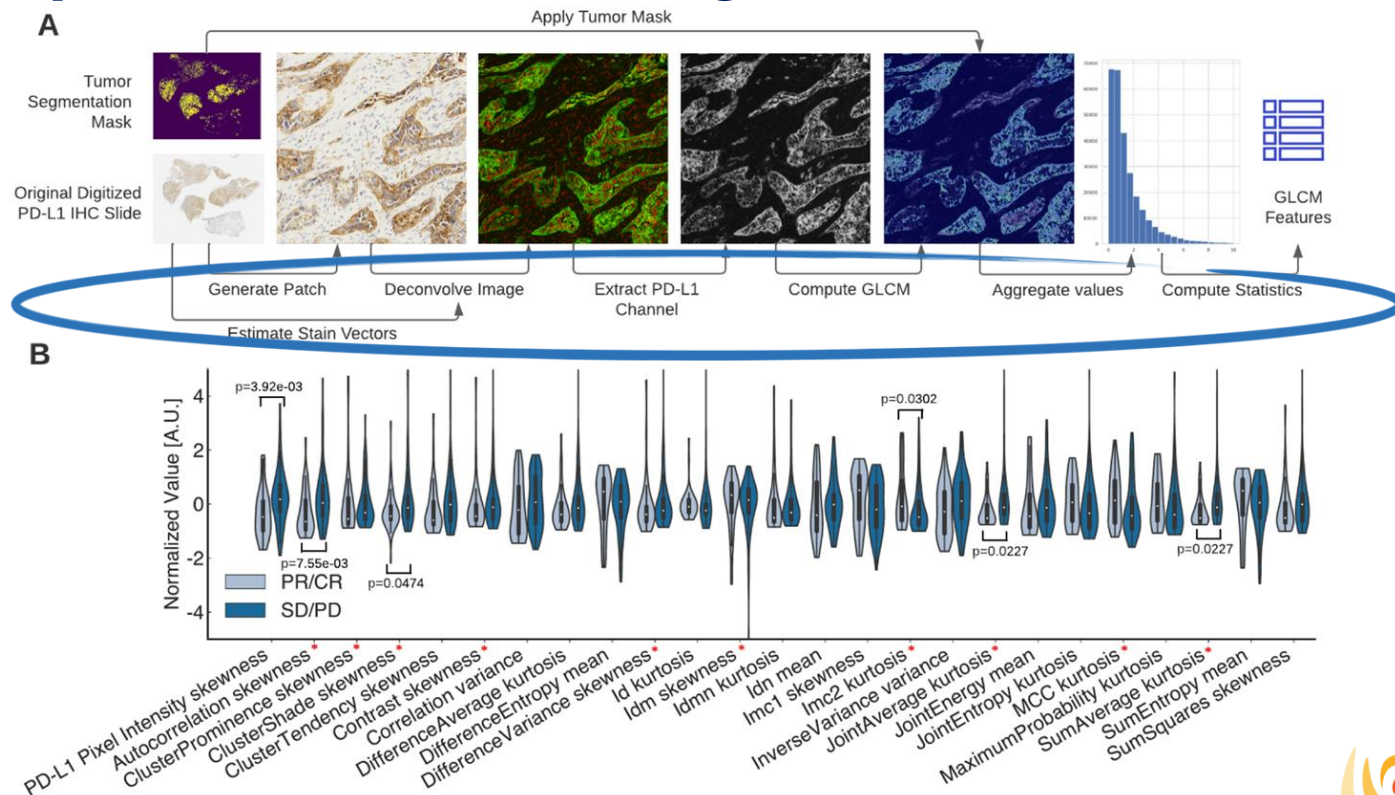
Proof of principle – improving prediction of response to ICB in lung cancer



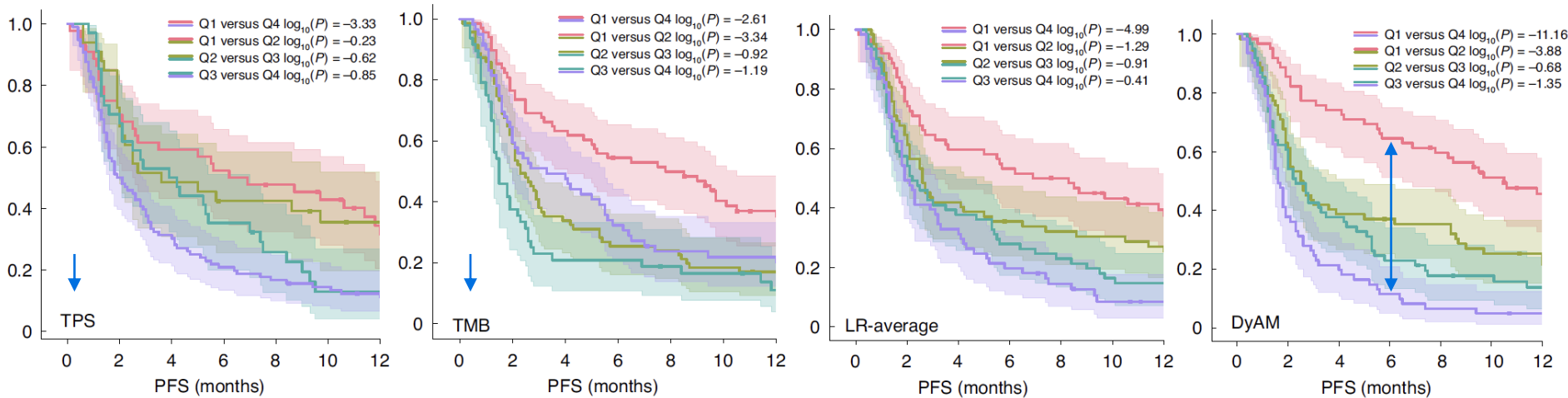
n=247 patients with advanced NSCLC and treated with anti PD-1/PD-L1



Proof of principle – improving prediction of response to ICB in lung cancer



Proof of principle – improving prediction of response to ICB in lung cancer

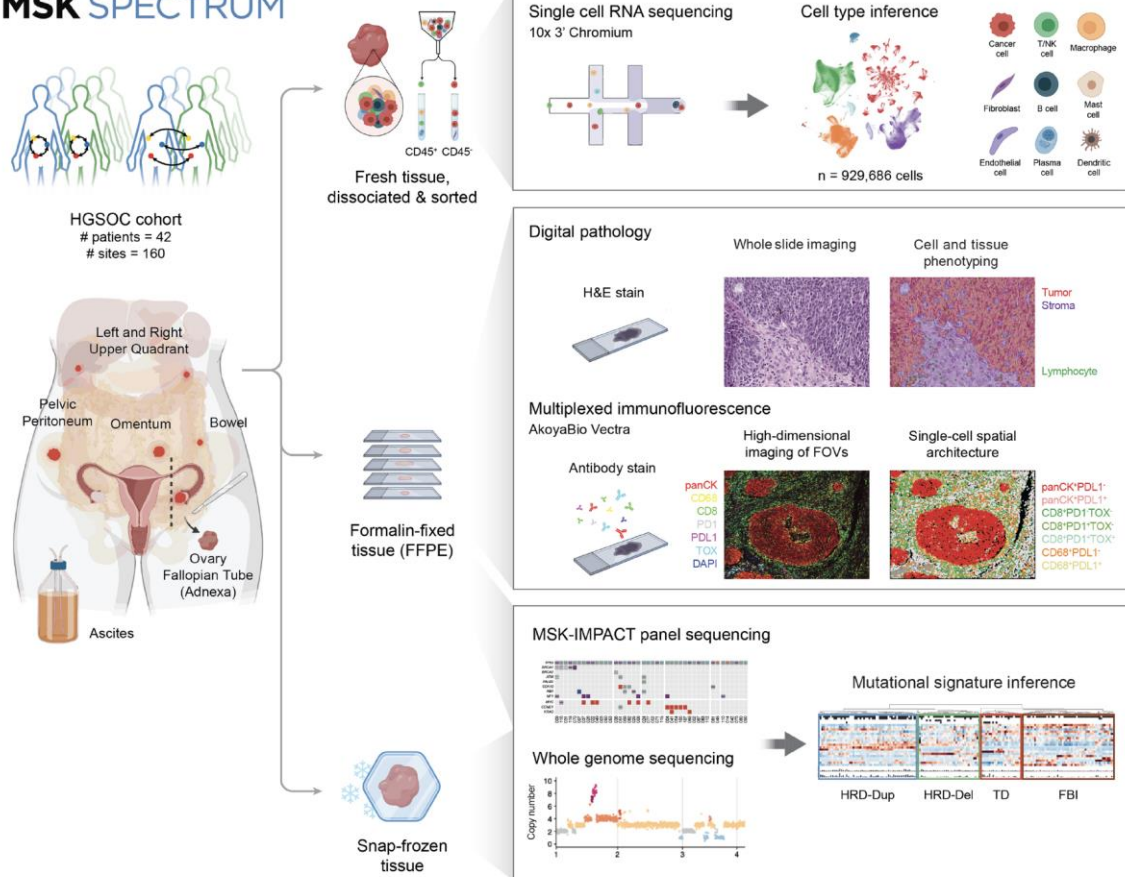


Q1																												
At risk	44	31	26	23	20	17	11	68	52	43	36	33	24	19	62	46	37	33	30	24	19	62	53	46	40	36	29	24
Censored	0	0	0	0	1	2	4	0	0	0	1	1	4	6	0	0	0	0	1	3	5	0	0	0	0	1	3	5
Events	0	13	18	21	23	25	29	0	16	25	31	34	40	43	0	16	25	29	31	35	38	0	9	16	22	25	30	33

Q4																												
At risk	97	48	29	20	16	13	9	60	35	28	19	13	12	11	62	30	19	12	8	4	2	62	23	12	7	4	3	2
Censored	0	1	1	1	1	2	3	0	1	1	1	2	2	2	0	1	1	1	2	3	5	0	1	1	1	1	1	2
Events	0	48	67	76	80	82	85	0	24	31	40	45	46	47	0	31	42	49	52	55	55	0	38	49	54	57	58	58

Multi-omic integration in ovarian cancer

MSK SPECTRUM



scRNA: 42 patients, 320 samples

Bulk WGS: 40 patients

mpIF: 35 patients, 1000 FOV

GYN DMT:

Nadeem Abu-Rustum

Carol Aghajanian

Surgery

Kara Long-Roche

Oliver Zivanovic

Yukio Sonoda

Dennis Chi

Ginger Gardner

Vance Broach

Elizabeth Jewell

Jennifer Mueller

Carol Brown

Mario Leitao

Pathology

Lora Ellenson

Robert Soslow

GYN Med Onc

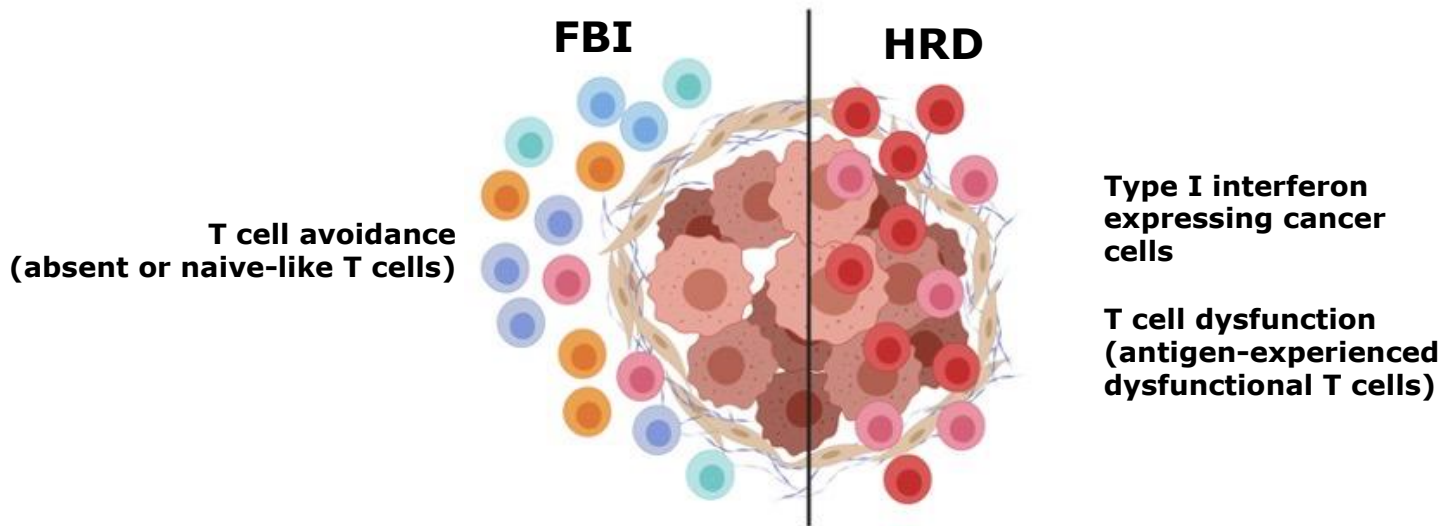
Dmitriy Zamarin**

Claire Friedman

Ying Liu

An emerging model for how mutational processes impact immuno-reactivity

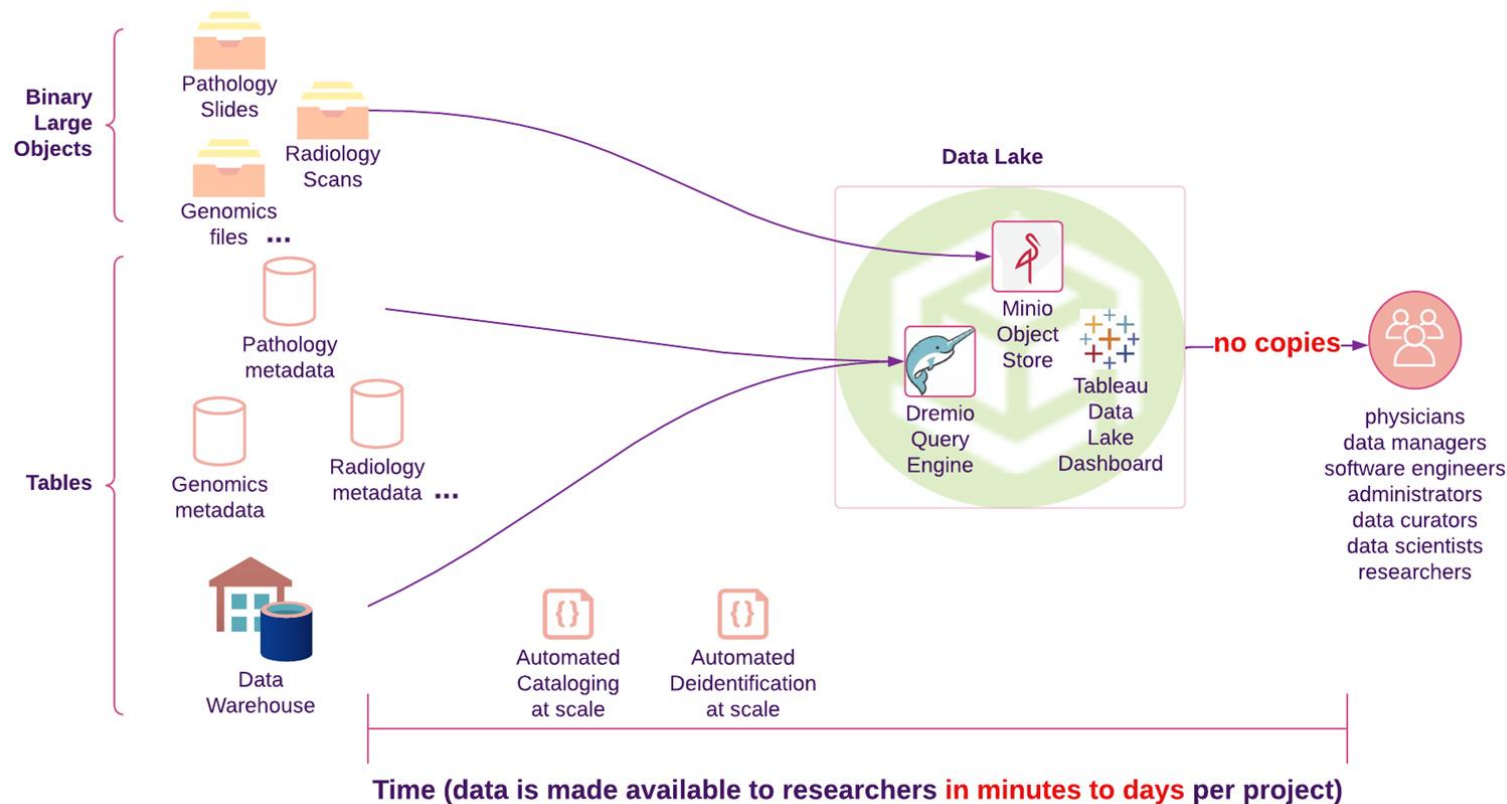
Vázquez-García I, *Nature*. 2022 Dec;612(7941):778-786.
PMID: 36517593



- Multi-omic views to answer central questions in disease biology
- Mutational processes lead to distinct cell-intrinsic signaling pathways
- Mutational processes associate with immune cell phenotypic states
- Mutational processes associate with distinct tissue architectures

How to scale?

Engineering process & scientific data management system



Lessons learned

1. Multimodal data integration is a powerful approach to BOTH real world and experimental data interpretation
2. Multimodal data integration is a TEAM sport requiring multidisciplinary collaboration: clinical expertise <> computational rigor
3. The path to integrative diagnostics will be informed by large-scale retrospective multimodal studies, but this field is early in its development
4. Retrospective data wrangling is a major bottleneck to progress
5. Investment in data engineering provides scale, automation and standards
6. Reproducibility is required but hard
7. Talent is a bottleneck – recruiting accomplished data scientists to work on the challenges we face in cancer