Breakout Session 2: Track B

ASTOR: Alliance Standardized Translational 'Omics Resource

Dr. Benjamin Vincent
Associate Professor, Division of Hematology, University of North Carolina, Lineberger Comprehensive Cancer Center

A-STOR
Alliance Standardized Translational Omics Resource

James S. Blachly, MD
Associate Professor
Division of Hematology
Ohio State University
Comprehensive Cancer Center

Daniel G. Stover, MD
Associate Professor
Division of Medical Oncology
Ohio State University
Comprehensive Cancer Center

Benjamin Vincent, MD
Associate Professor
Division of Hematology
University of North Carolina
Lineberger Comprehensive Cancer Center

A-STOR: Background

- Challenges for clinical trial high dimensional data
  - Correlative 'omics are an integral part of clinical trials
  - NCTN/GDC/dbGAP are definitive repositories
  - Data are decentralized & often deposited years after sequencing

- **ASTOR: Goals**
  - Stable, secure, **scalable storage** for multi-omics data from Alliance trials
  - **Accessibility** to approved investigators for more rapid and robust analysis including secondary analyses and meta-analyses
  - **Rapid correlative turnaround** enhances trial design, grants, and publications
  - **Data harmonization** to facilitate meta-analyses
  - **Portal for analyses**
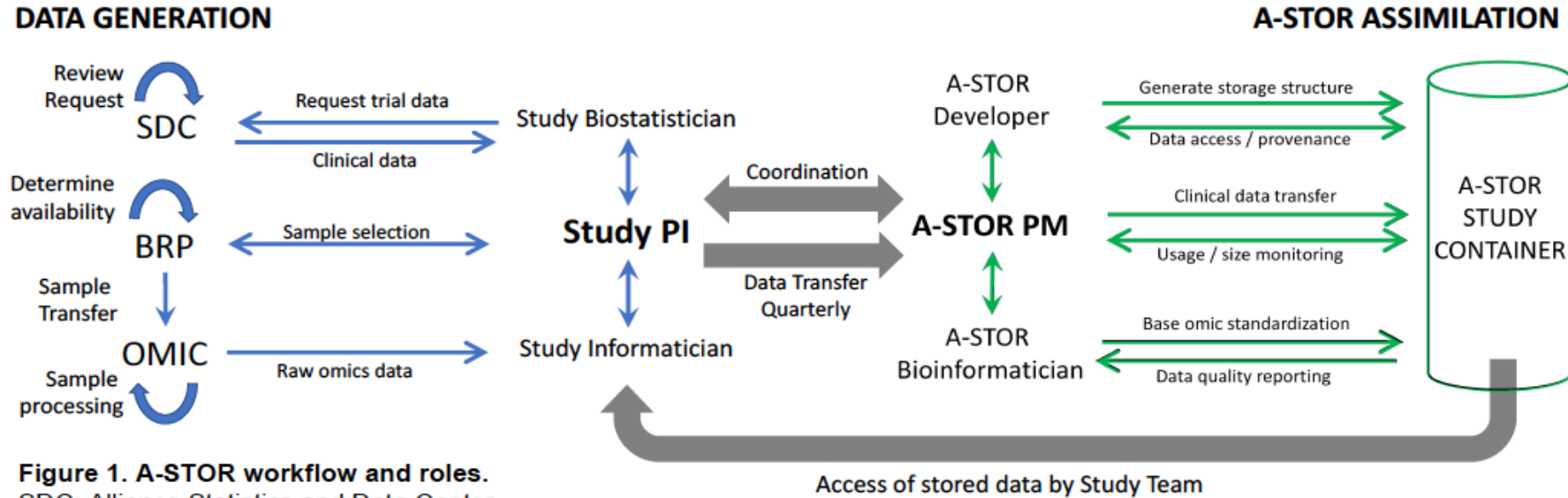
A-STOR: Workflow & Structure

Figure 1. A-STOR workflow and roles.
SDC: Alliance Statistics and Data Center
BRP: Alliance Biorepository   OMIC: Designated Sequencing Core
PM: Project Manager   PI: Principal Investigator

Governance: ASTOR 'Manual of Operations'

- Regulatory: Confirmation of consent, deidentification, HIPAA compliance

- Approvals: MOU to deposit data, User rights/responsibilities

- Data use/standards: FAIR principles, inter-operability, end-to-end provenance

A-STOR: Workflow & Structure

Consolidated directory structure: 'Put it all together'

- Diverse data types: DNAseq, RNAseq, SNP, path images, ctDNA
  - Flexible storage: Raw (fastq), aligned (BAM), processed (vcf, maf), variant results
- De-identified clinical data: Minimal (not full) clinical data + study docs

A-STOR: AI/ML Readiness

Overarching Goals Addressed:

- **Stable, secure, scalable storage for multi-omics data from Alliance trials**

- **Accessibility to approved investigators for more rapid and robust analysis**

  - Support primary study team with analyses and meta-analyses
  - Secondary analyses by Alliance investigators before 1° publication

- Correlative turnaround enhances trial design, grants, and publications

- **Data harmonization to facilitate meta-analyses**

- Portal for analyses

ASTOR AI/ML Readiness

- **AIM 1: To expand and optimize ASTOR to facilitate AI/ML research**
  - **Expand studies in ASTOR: 2300pts → >17,000 pts | 10Tb data → >100Tb data**
  - **Focus on digital pathology images**

Trial ID	Cancer	Unique Patients (n)	SNP	DNAseq	RNAseq	DigPath	Other
AFT-05	Breast	2558		X	X	X	
CALGB 40101	Breast	2015	X			X	
CALGB 9741	Breast	2005				X	
CALGB 80405	GI	1512	X	X	X	X	X (NanoString)
CALGB 80702	GI	1040	X				X (ctDNA)
A031201	GU	989	X				
A011106	Breast	857		X	X	X	
CALGB 90401	GU	827	X				
A0131501	GU	702		X	X	X	
CALGB 40502	Breast	661	X		X	X	X (CTC)
CALGB 70604	Myeloma	656	X				
CALGB 50303	Lymphoma	550				X	
A031202	GU	545		X			X (ctDNA)
AFT-38	Breast	518		X			X (ctDNA)
AFT-04	Breast	488			X	X	
CALGB 90601	GU	456	X				
CALGB 40603	Breast	389		X	X	X	
CALGB 80303	GI	374	X				
CALGB 40601	Breast	265		X	X	X	
A031902	GU	56		X	X		
TOTAL		**17,463 Unique Patients**					

ASTOR AI/ML Readiness

- **AIM 1: To expand and optimize ASTOR to facilitate AI/ML research**
  - **Expand studies in ASTOR: >17,000 patients, >100Tb data**
  - **Focus on digital pathology images**
    - Significant interest from academic and commercial partners
    - **Leadership by Mark Watson/BSSR – vision to scan slides results in powerful data**
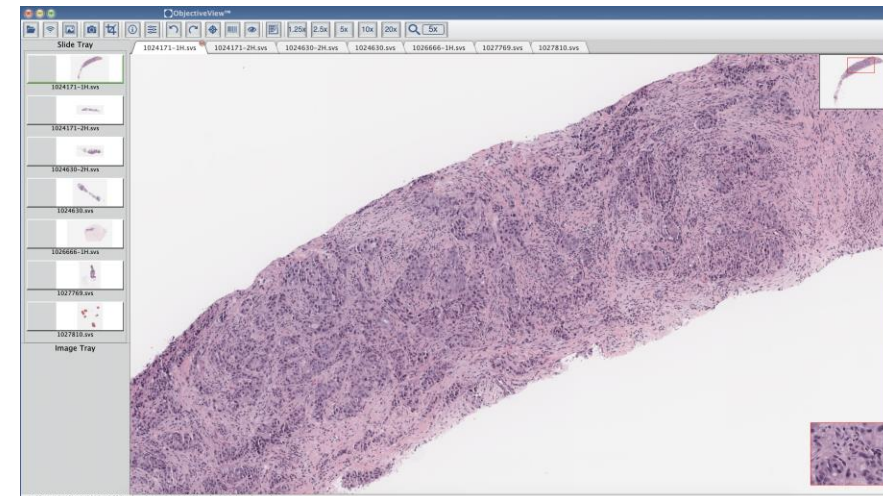
By end 2025, ASTOR plans to host >45,000 digital H&E images

ASTOR AI/ML Readiness

- **AIM 2: Test and implement Alliance Data Dictionary ExtedeR (ADDER) for trial clinical and adverse event data harmonization**
  - Develop a unified clinical & AE data data dictionary and data dashboard
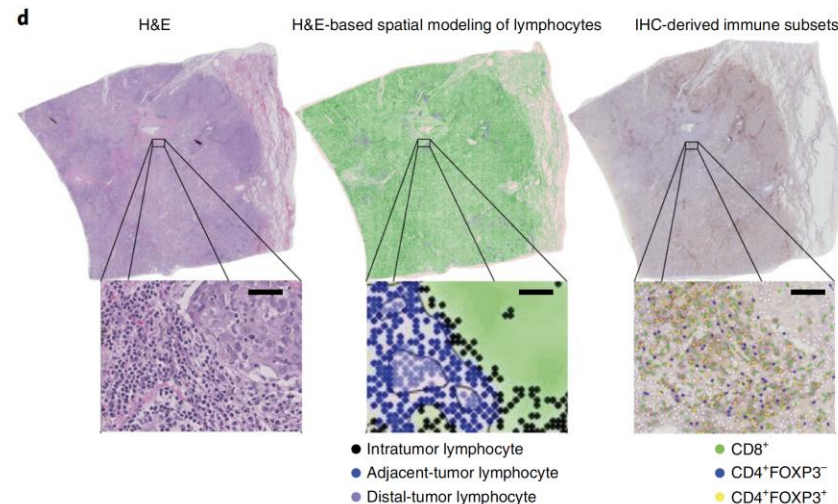

Rapid Clinical DATA HARMONIZATION
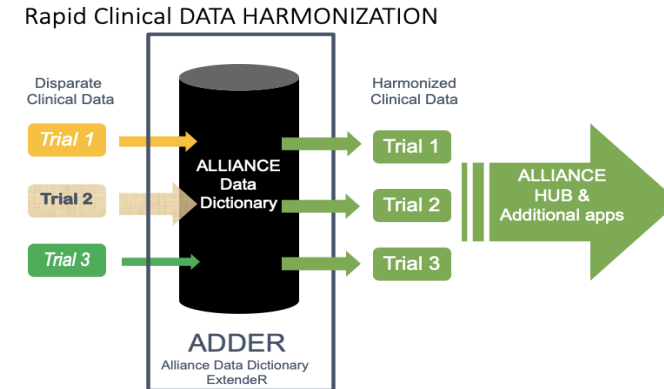
Example: ASTOR Data Dictionary v1.1

DATA ELEMENT Common Name (Based on	DATA CLASS	DATA ELEMENT Common Name	Provenance / Source Text?	Definition	CDASH/SDTM Variable	CDASH Note	mCODE Data Element
title	PROTOCOL/TRIAL	Short Title	SDMC	Study title	TS_TSVAL_TITLE	Study identifier not title or protocol number	
protocol_number (PK)	PROTOCOL/TRIAL	Protocol Number	SDMC	Alliance Protocol Number	STUDYID		
disease_site			SDMC	PENDING			
start_date	PROTOCOL/TRIAL	Start Date	SDMC	Protocol activation date	TS_TSVAL_SSTDAT		
end_date	PROTOCOL/TRIAL	End Date	SDMC	Last Patient Last Visit,	TS_TSVAL_SENDAT		
principal_investigator_name	PROTOCOL/TRIAL	PI	SDMC	Alliance Study Chair	TS_TSVAL_SCHAIR		
navigator_status	PROTOCOL/TRIAL		BBR				
ppid (PK)	**PARTICIPANT**	**PPID**		Alliance Participant ID	USUBJID		patient-9
sex	PARTICIPANT	Sex	SDMC	Sex, as determined by the	DM_SEX		patient-3
ethnicity	PARTICIPANT	Ethnicity	SDMC	Ethnicity, as entered in OPEN	DM_ETHNIC		patient-14
country	PARTICIPANT	Country	SDMC	Country of Residence, as ente	SC_SCORRES_CNTRYRES		
race	PARTICIPANT	Race	SDMC	Race, as entered in OPEN	DM_CRACE		patient-13
trial_protocol_number (FK)				Link patient to trial			
id (PK)				Uuid to link encounter to patient?			
height_inches	PARTICIPANT	Height at Registration	SDMC	Participant Height at	VS_VSORRES_HEIGHT		
weight_lb	PARTICIPANT	Weight at Registration	SDMC	Participant Weigh at	VS_VSORRES_WEIGHT		
registration_date	PARTICIPANT	RegistrationDate	SDMC	Participant (First)	DS_DSSTDAT_REGRAND		
trial_arm	PROTOCOL/TRIAL	Arm	SDMC	Participant (First)	CARM		
disease_code	PARTICIPANT	Disease Code					
nci_medra_disease_code	PARTICIPANT	NCI MedRA Disease Code	SDMC	NCI MedRA Disease Code at	SUPPMH_QVAL_MHDSXCD		
ecog_performance_status		ECOG PS at Registration	SDMC	ECOG Performance Status at I	RS_RSORRESC_ECOG101		patient-17
participant_mrn (FK)				Link patient to trial encounter			

ASTOR AI/ML Readiness

- **AIM 3: Pilot implementation of ASTOR and ADDER to develop a multi-modal ML-based neoadjuvant chemotherapy response predictor using pooled data from two phase III clinical trials.**
  - TNBC Neoadjuvant Chemotherapy: CALGB 40603 and BrighTNess
  - Integration of RNAseq and digital pathology

- AI Algorithm Deployment & Unit Testing
  - **Digital Pathology Algorithms:**
    - Geospatial immune variability algorithm (Abdul-Jabbar, et al *Nat Med* 2020)
    - Open source, QuPath-based tumor infiltrating lymphocyte

d

H&E   H&E-based spatial modeling of lymphocytes   IHC-derived immune subsets

- Intratumor lymphocyte
- Adjacent-tumor lymphocyte
- Distal-tumor lymphocyte
- CD8+
- CD4+FOXP3-
- CD4+FOXP3+

ASTOR AI/ML Readiness

- **AIM 3: Pilot implementation of ASTOR and ADDER to develop a multi-modal ML-based neoadjuvant chemotherapy response predictor using pooled data from two phase III clinical trials.**
  - TNBC Neoadjuvant Chemotherapy: CALGB 40603 and BrighTNess
  - Integration of RNAseq and digital pathology

- AI Algorithm Deployment & Unit Testing
  - **Digital Pathology Algorithms:**
    - Geospatial immune variability algorithm (Abdul-Jabbar, et al *Nat Med* 2020)
    - Open source, QuPath-based tumor infiltrating lymphocyte
  - **RNAseq**:
    - Harmonization/reprocessing of RNAseq via standard STAR-Salmon pipeline
    - Deployment of immune deconvolution algorithms + immune signatures
  - **AI/ML Implementation:** Implementation of dockerized platforms on Ohio Supercomputer Center 'Ascend' cluster cluster devoted entirely to intensive GPU processing (NVIDIA A100 80GB GPUs)
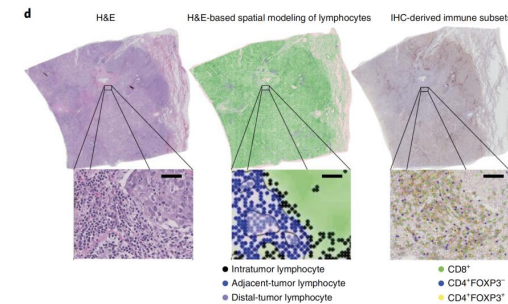
ASTOR AI/ML Readiness: Deliverables

- **Example Approved Projects:**
  - **Aim 1 (Multi-trial Data)**: Kalari (Mayo MN) – RNAseq predictor for TNBC; Steckline (Kansas) – DDIR and neoadj chemo response; Rajagopal (NCI) – germline:somatic interaction; Magbanua (UCSF) – CTCs and RNAseq to predict breast cancer outcomes
  - **Aim 2 (ADDer):** Lustberg (Yale) – Meta-analysis of CIPN prediction
  - **Aim 3 (Image Analysis)**: Vater (Ohio State) – Image/RNAseq TILs classification; Chumsri (Mayo FL) – TLS and outcomes; Howard (UChicago) – AI-based pathology image subtyping and predictors
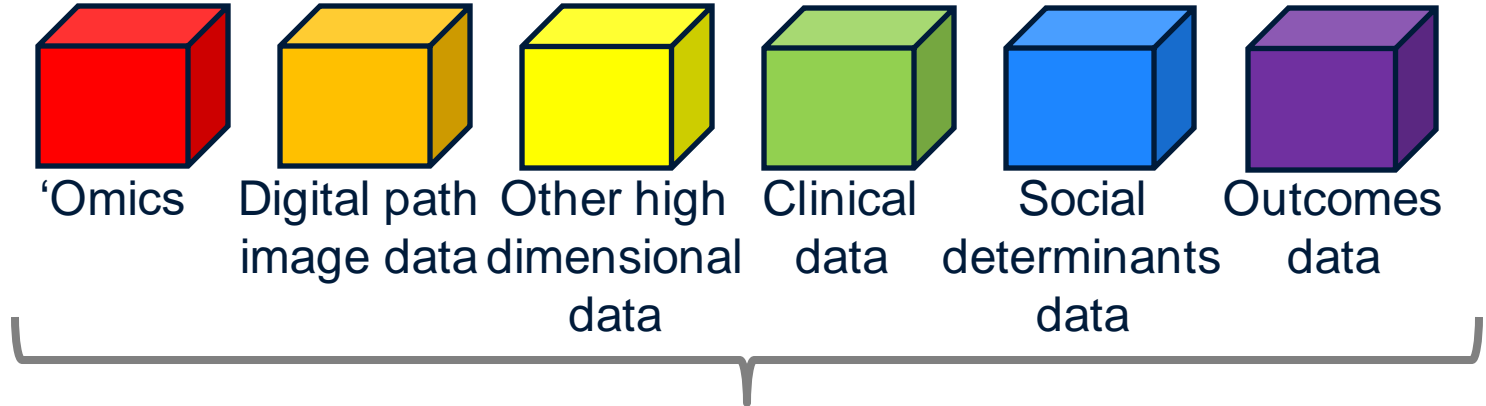
- **Publications:**
  - Thompson KJ, Leon-Ferre RA, Sinnwell JP, Zahrieh DM, Suman VJ, Metzger FO, Asad S, Stover DG, Carey L, Sikov WM, Ingle JN, Liu MC, Carter JM, Klee EW, Weinshilboum RM, Boughey JC, Wang L, Couch FJ, Goetz MP, Kalari KR. Luminal androgen receptor breast cancer subtype and investigation of the microenvironment and neoadjuvant chemotherapy response. *NAR Cancer*. doi: 10.1093/narcan/zcac018. PMID: 35734391
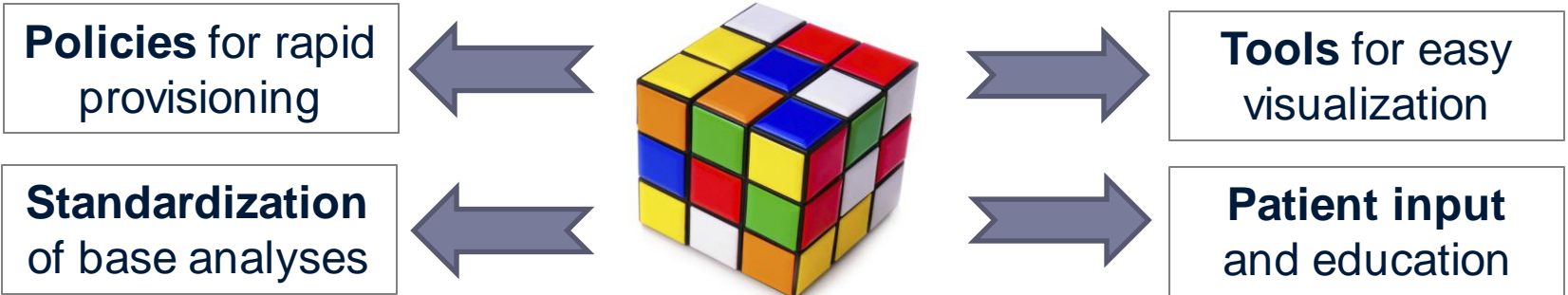
ASTOR: Innovation & Future Directions

- **First NCI NCTN cooperative group** to establish and operationalize a secure, cloud-based HIPAA-compliant repository for clinical and high dimensional data.

- **Harmonization** of 'omic <u>and</u> clinical/adverse event data across clinical trials.

- Coordination of **multiple distinct high dimensional data elements** for each patient (e.g. RNAseq, DNAseq, and digital pathology) facilitates AI/ML analyses beyond single modality efforts.

- **Deployment of tools to facilitate user accessibility** for non-informaticians and informaticians, including cBioPortal instances for individual trials.

- **Future Directions/Next Steps**:
    - Continue expansion of diverse data hosted and supported
    - User interfaces: cBioPortal, OpenSlide, immunogenomic tools
    - User feedback and input on trajectory

Future of Alliance Translational Research

'Omics
Digital path image data
Other high dimensional data
Clinical data
Social determinants data
Outcomes data

ASTOR: Unified, expandible, secure

Policies for rapid provisioning

Standardization of base analyses

Tools for easy visualization

Patient input and education

Thank You!

A-STOR
Dan Stover
James Blachly
Benjamin Vincent
Justin Auerbach
Kristyn Glover
Shawn Striker

Collaborators
ITSC: Chuck Perou
Pathology: Michael Linden

Alliance/AFT
Eva Galanis
Monica Bertagnolli
TRP
Fraser Symmans
Yujia Wen
Biorepostiory
Mark Watson
Linda McCart
SDMC
Ann Oberg
Nick Larsen

Funding

PATIENTS.
Advocates
Patients who guide our research
Patients who donated samples

Daniel.Stover@osumc.edu
James. Blachly@osumc.edu

@StoverLab   @JamesBlachly