

UAB Informatics Institute Biocomputing Research Infrastructure and Tools for Omics Data Analysis

Jake Y. Chen, PhD, FACMI, FAIMBE, FAMIA

Professor of Genetics, Computer Science & Biomedical Engineering

Chief Bioinformatics Officer, Informatics Institute

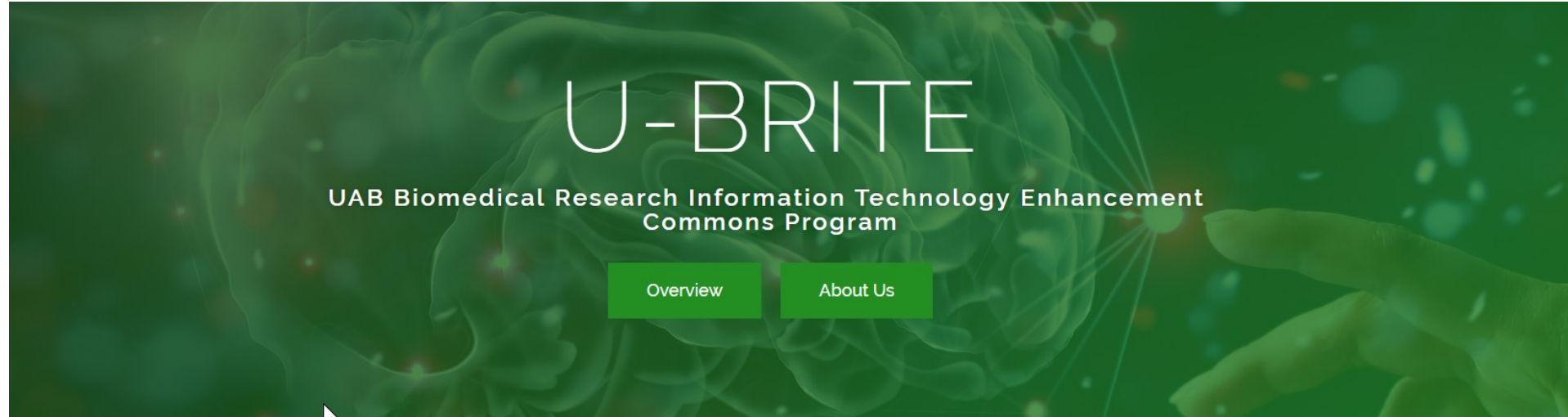
The University of Alabama at Birmingham, USA

jakechen@uab.edu | aimed-lab.org

Schedule

- **10-10:40am** **U-BRITE**
 - Jake Y. Chen, PhD, UAB Informatics Institute
 - Zhandos Sembay, MS, UAB Informatics Institute
- **10:40-11:40am** **PAGER**
 - Zongliang Yue, PhD, UAB Informatics Institute
- **11:40am-noon** **SEAS (Cohort analysis)**
 - Thanh Nguyen, PhD, Department of Biomedical Engineering

UBRITE: Enabling community-based team data science



U-BRITE is a biomedical data science infrastructure developed at the UAB Informatics Institute to:

- Provide HIPAA-compliant, and FAIR-compliant biomedical bioinformatics **software tools and APIs** on novel research computing infrastructure.
- Help researchers better manage and analyze genomic medicine **data sets** in a new “translational research commons” environment.
- Facilitate interdisciplinary research teams to perform “**open, virtual science**”.

The new paradigm of computing

“inverting the model of genomic data access”



U-BRITE is a community-based team data science platform (<http://ubrite.org/>)

U-BRITE is a biomedical data science infrastructure developed at the UAB Informatics Institute

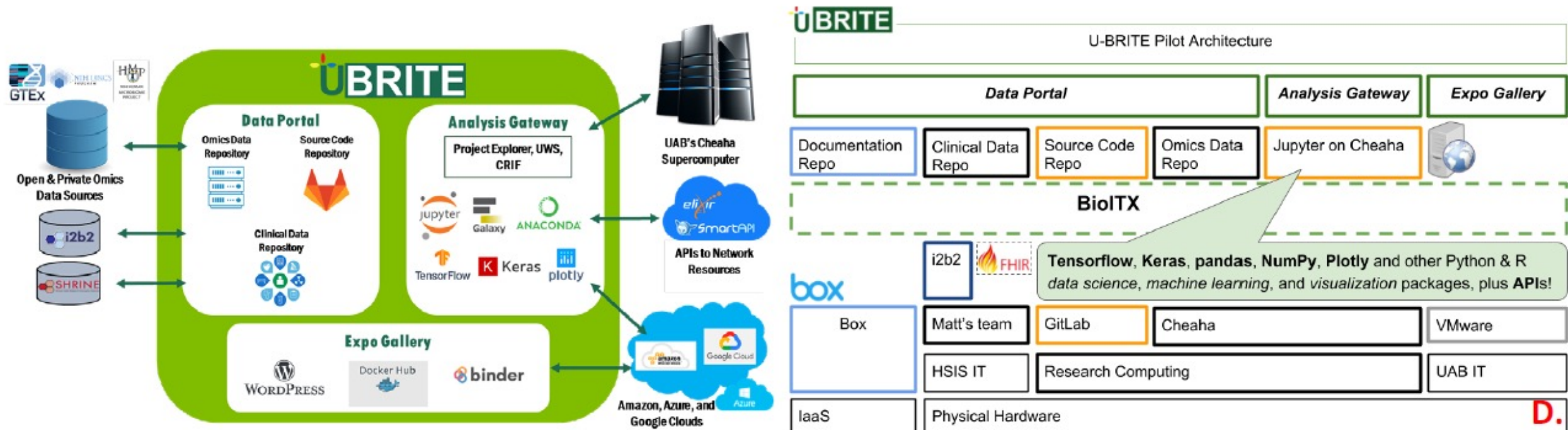



Figure 4: An overview of U-BRITE design to enable team-based translational science. The left shows how U-BRITE relates to UAB informatics systems/data. The right is an architecture showing components.

UBRITE is a data/analysis/communication commons environment 3-in-1

- Community web portal (<http://cancer.ubrite.org>)
- Provide data commons for coding, documentation
- Provide analysis commons
- Provide workforce training by hosting data science hackathon (<http://covid.ubrite.org>)
- Hosting DREAM challenge contest for AI/ML
- Providing Jupyter web notebook based R/python application development
- Develop data-driven R-Shiny/Py-Streamlit Web APP

Demo

U-BRITE as a team data science gateway to HPC locally at UAB



OverviewExpo gallery ▾Analysis Gateway ▾Data Portal ▾Learning ResourcesGet Started

ENSEMBL

Shared Data Directory

Omics Data Repository

Source Code Repository

Clinical Data Repository

Documentation Repository

About

"Ensembl is a database and genome browser for enabling research on vertebrate genomes and integrate a diverse collection of large-scale reference data to create a more comprehensive view of genome biology. Our extensive data resources include evidence-based gene and regulatory region annotation, genomic tracks, and a comprehensive suite of tools, infrastructure and programmatic access methods ensure uniform data analysis and distribution for all supported species. Together, these provide a comprehensive solution for large-scale and targeted genomics applications alike."

Excerpted from paper at [doi: 10.1093/nar/gkw1104](https://doi.org/10.1093/nar/gkw1104).

rc.uab.edu/pun/sys/shell/ssh/default/

GmailYouTubeMapsOther bookmarksReading list

Warning: Permanently added the ECDSA host key for IP address '172.20.0.29' to th
list of known hosts.

Disk Quota Report

data/user/jakechen : 977.97GB of 11811.16GB

jakechen@login004 /]\$ ls

bin bootstrap.cgi data efi gpfs home.local lib local mnt opt root run scratch share srv tftpboot usr

root cm dev etc home home.orig lib64 media old-scratch proc rstore sbin scratch.old sm.log sys tmp var

jakechen@login004 /]\$ dir

bin bootstrap.cgi data efi gpfs home.local lib local mnt opt root run scratch share srv tftpboot usr

root cm dev etc home home.orig lib64 media old-scratch proc rstore sbin scratch.old sm.log sys tmp var

jakechen@login004 /]\$ cd share

jakechen@login004 share]\$ ls

apps

jakechen@login004 share]\$ cd /data

jakechen@login004 data]\$ ls

m project rc scratch temp-scratch temporary-scratch user

jakechen@login004 data]\$ cd project

jakechen@login004 project]\$ ls

VM1_TTS_CFD Grant_scRNAseq NLSB alstate ccts gitlab-data kimberly-lab lu

trora_lab HARPLab PerioperativeDataScience amaraLab cfdlab gitlablfs kinglab ma

BRL ICOS RA2_DREAM atlab chonglab grytzlab kjleelab ma

IRC_NBI KHOmics SharafLab bamman-lab circRNA-Analysis hartmanlab kobielab ma

ryoEM KJ_projects Szaflarski_DoD_NODDI bhattlab ctstackhlab hcgs kutschlab me

UTTALAB KnightResearch UVM_COVID bioinf datascienceteam hsght lahtilab me

towns-Lab LV_lab U_BDS bioitx daylab iqml lasseigne_lab mo

rosophila_Melkanilab MRIPhantom Welner biokdd-lab dti kanalab lcdl mr

MemRE MillerLab aghi boldlab galaxy kes leallab mu

BM_PDX NCRlab allendorferlab bridgeslab gitlab-container-registry kimberly-cser lubinlab ne

jakechen@login004 project]\$ cd ubrite/

jakechen@login004 ubrite]\$ ls

ancer-hackathon covid19-hackathon covid19-public general-public pilot-hackathon pipelines shared-data shared-data.sha256sum

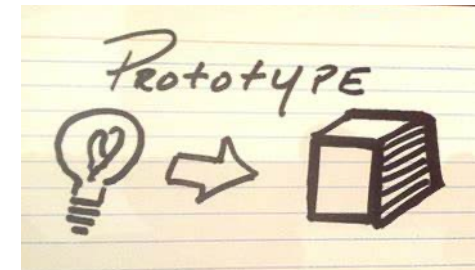
jakechen@login004 ubrite]\$

Modern team data science: *interdisciplinary, agile, and translational*

1. Ideas & Team Formation



2. Prototyping



3. Pilot Studies



4. System Building



5. Real-world Solutions



*Feedback,
decomposition, testing,
refinement, integration,
and redeployment*




RA2 DREAM Challenge

Automated Scoring of Radiographic Joint Damage



Radiographic joint damage



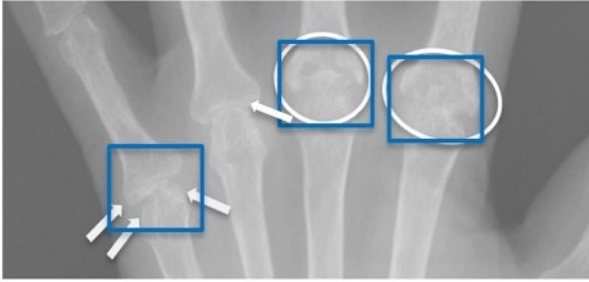
-  Normal
-  Bone erosion
-  Joint space narrowing

Three sub challenges

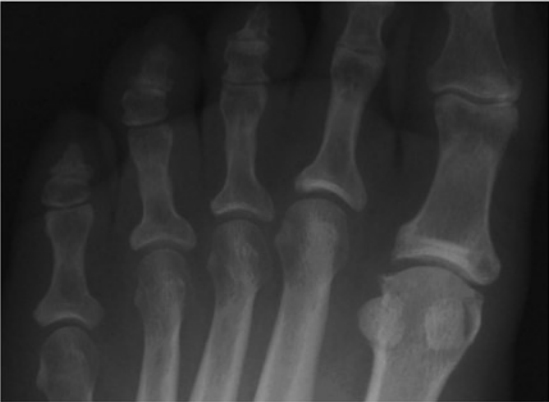
- # 1: Predict overall RA damage in target joints of the hands and feet
- # 2: Predict total joint space narrowing scores in target joints of hands and feet
- # 3: Predict joint erosion scores in target joints of hands and feet

Supplemental Figure

	UAB586 (total 11)				UAB504 (total 195)			
	mcp-5	mcp-4	mcp-3	mcp-2	mcp-5	mcp-4	mcp-3	mcp-2
Erosion Score	0	0	0	1	3	1	5	5
Narrowing Score	0	0	0	0	2	0	4	4



	UAB272 (total 7)						UAB634 (total score45)					
	mtp-5	mtp-4	mtp-3	mtp-2	mtp-1	mtp-ip	mtp-5	mtp-4	mtp-3	mtp-2	mtp-1	mtp-ip
Erosion	0	0	0	0	0	0	7	3	5	4	1	0
Narrowing	0	0	0	0	0	0	0	0	0	0	0	0



UAB272 (total 7)						UAB085 (total 179)					
ulna	lunate	radius	nav	mul	mc1	ulna	lunate	radius	nav	mul	mc1
0	0	0	0	0	0	3	5	0	3	2	0

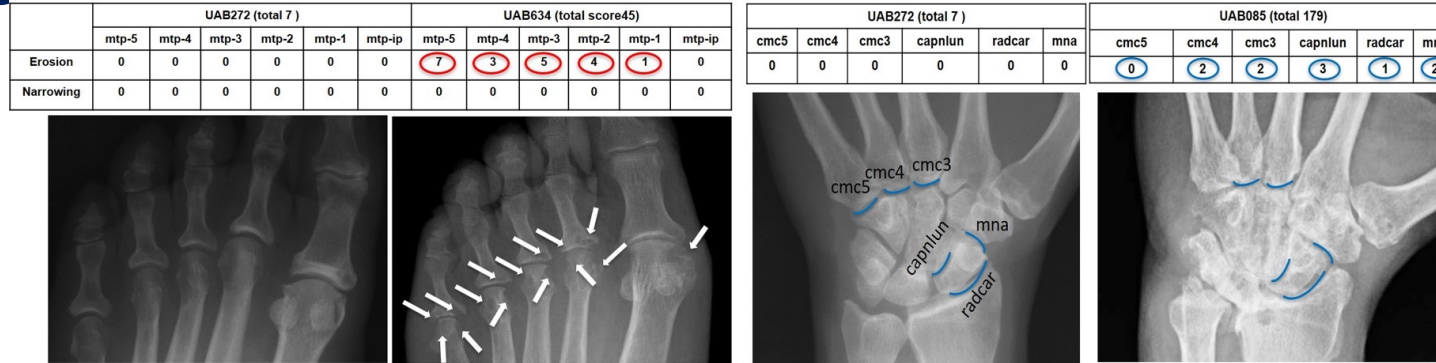


UAB272 (total 7)						UAB085 (total 179)					
cmc5	cmc4	cmc3	capnlun	radcar	mna	cmc5	cmc4	cmc3	capnlun	radcar	mna
0	0	0	0	0	0	0	2	2	3	1	2



Training Machine Learning and AI models on U-BRITE

Assessing community-submitted AI projects



B

RA2-DREAM Challenge Design

