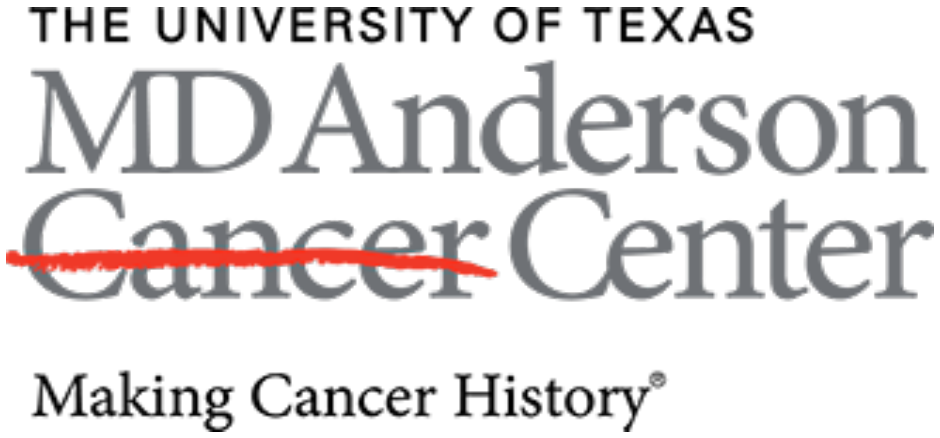
Applications of HPC for Prediction of Liver Cancer Treatment Response

Lunch and Learn

David Fuentes

Department of Imaging Physics

February 2019



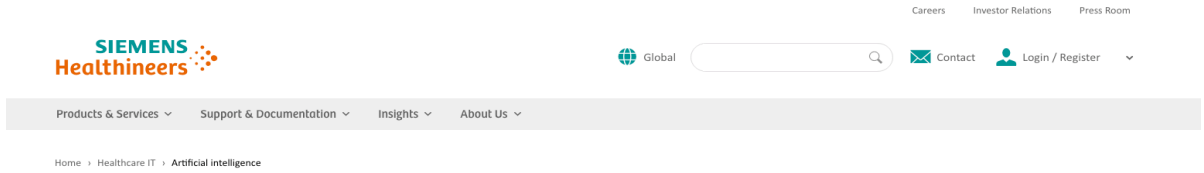
Outline

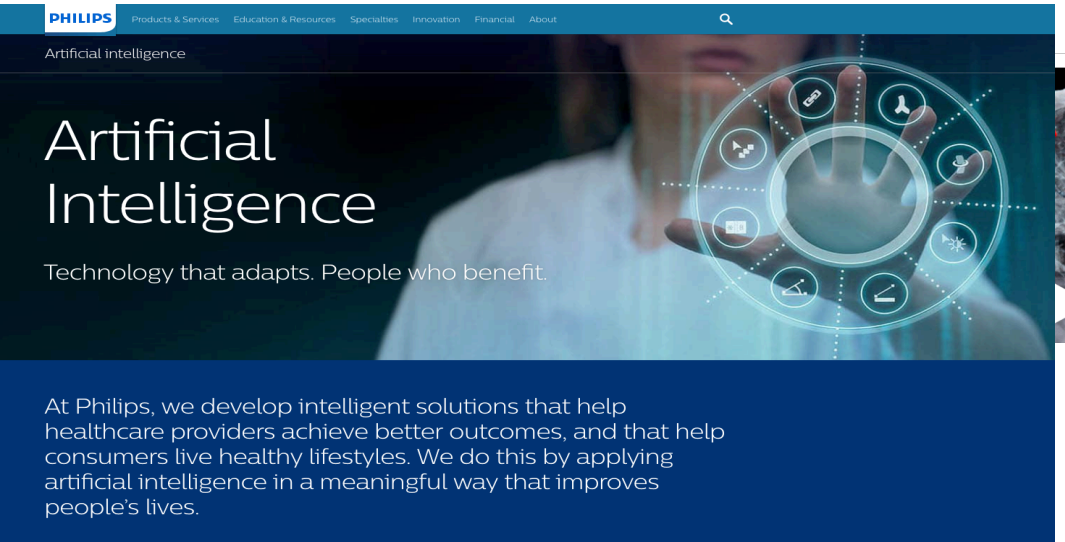
- Utilization of HPC
  - Hepatocellular Carcinoma (HCC)
    - AI
    - FEM
    - MD

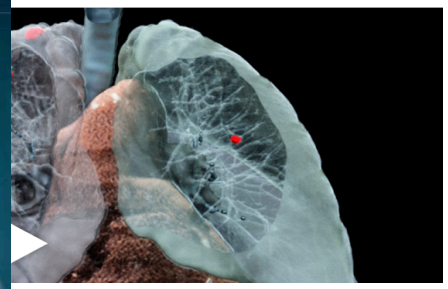
Industry adoption of AI

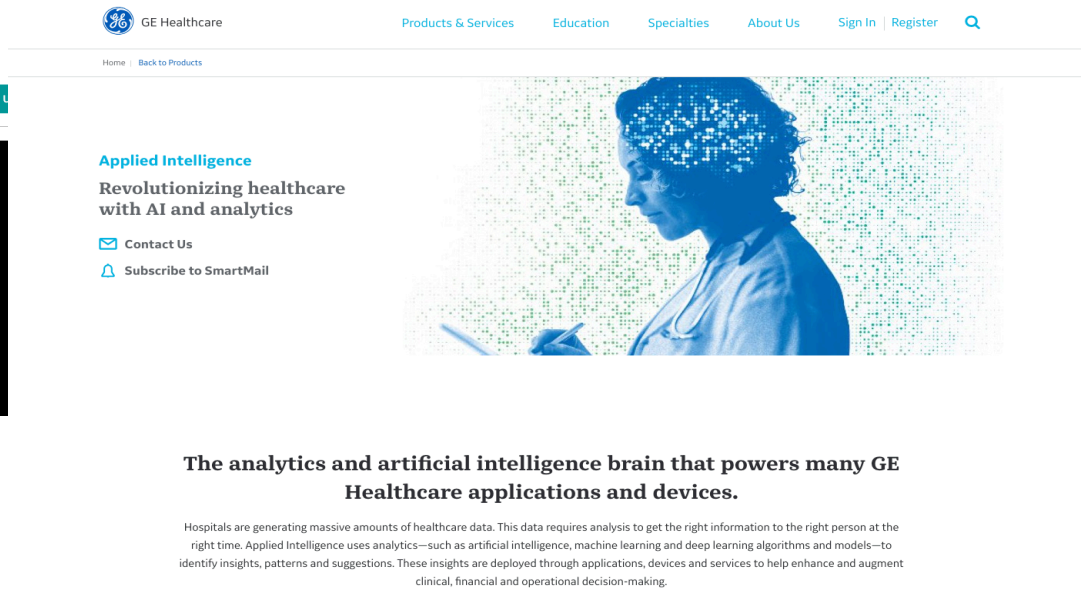
- Profound adoption of AI in healthcare industry
  - Fully automated!
- Business model: semi-automate large datasets to reduce cost

→ localize → scan → push to db → preprocess → analyze → model zoo → outcome predictions to dash board









Motivation

- Hepatocellular carcinoma (HCC)
  - >600,000 new cases diagnosed globally per year
  - ~ 1 new case every minute
  - Challenging treatment decisions

- Liver detoxifies blood
  - Alcohol, parasites, hepatitis → liver damage → inflammation → liver disease/cirrhosis → reduced liver function

- Liver disease commonly asymptomatic until life-threatening disease has developed
  - HCC treatment decisions challenging and must balance (1) preservation of liver function and (2) treatment of disease
  - 20% eligible for curative therapy

The Lancet, 379, Forner, A., Llovet, J. M. & Bruix, J. Hepatocellular carcinoma, 1245–1255, 2012

Multiple Staging approaches to guide treatment

- Semi-quantitative Staging models

Dhir, M., et al. (2016). A Review and Update of Treatment Options and Controversies in the Management of Hepatocellular Carcinoma

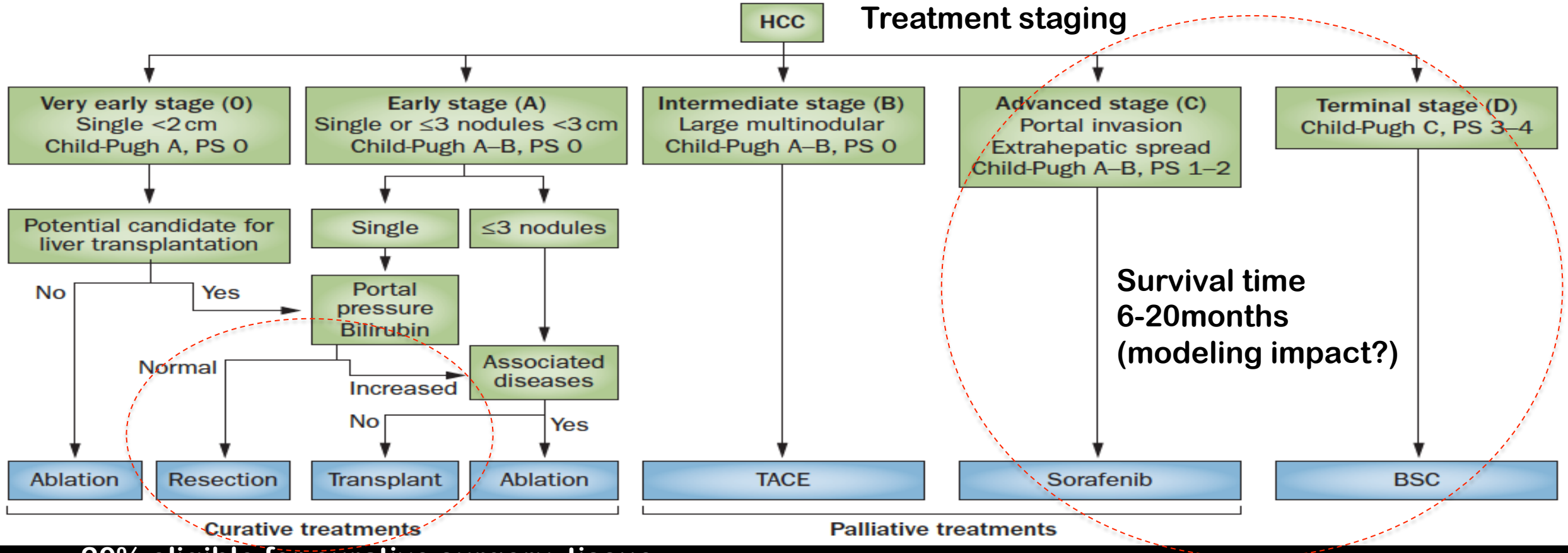
TABLE 1. Summary of Various Staging Systems Used for HCC

Classification	Type	Number of Subtypes	Subtypes	Tumor Staging Criteria	Liver Function	Health Status
Okuda stage	System	3	Stage I, II, III	Tumor size (<50% vs >50% liver involvement)	Bilirubin Albumin Ascites	—
French	Score	3	A: 0 points B: 1–5 points C: ≥ points	Portal invasion AFP	Bilirubin Alkaline phosphatase	Karnofsky
CLIP	Score	7	0, 1, 2, 3, 4, 5, 6	Tumor morphology (</>50% liver involvement) Portal vein thrombosis AFP	Child-Pugh stage	—
BCLC	Staging	5	0: very early A: early B: intermediate C: advanced D: end stage	Portal invasion Metastases Morphology Okuda	Child-Pugh stage Portal hypertension Bilirubin	Performance status test
CUPI	Score	3	Low risk: score ≤1 Intermediate: 2–7 High: ≥8	TNM AFP	Ascites Bilirubin Alkaline phosphatase	Symptoms
TNM	System	4	Stage I, II, III, IV	Number of tumors Vascular invasion Metastases	Fibrosis	—
JIS	Score	6	0,1,2,3,4,5	TNM stage by LCSGJ	Child-Pugh stage	—
ER	System	2	ER wild-type ER variant	Estrogen receptor	—	—

BCLC indicates Barcelona-Clinic Liver Cancer staging; CLIP, cancer of the Liver Italian Program; CUPI, Chinese University Prognostic Index; ER, estrogen receptor; JIS, Japanese Integrated Staging; LCSGJ, liver cancer study group of Japan. Adapted from Pons et al HPB 2005.[4]

Multiple Staging approaches to guide treatment

- Barcelona Clinic Liver Cancer (BCLC) Curative therapies reserved for ‘good’ liver function (child pugh A)

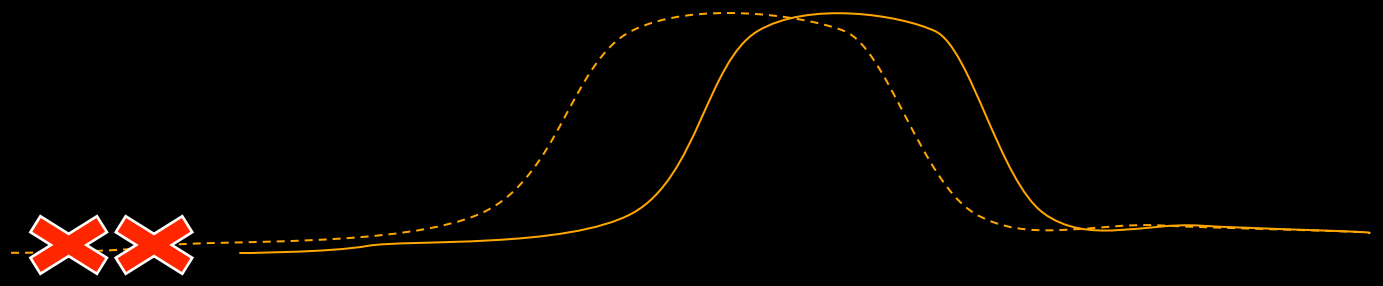


20% eligible for curative surgery, tissue extraction for gold standard comparison

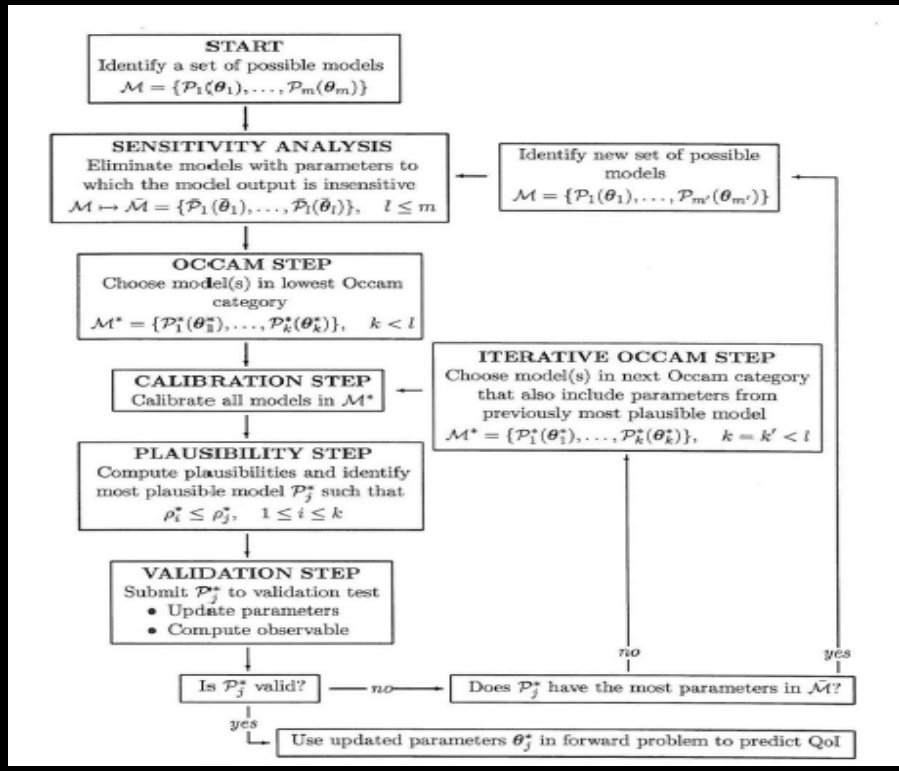
The Lancet, 379, Forner, A., Llovet, J. M. & Bruix, J. Hepatocellular carcinoma, 1245–1255, 2012

Patient Selection

- Thought experiment: How can math improve therapy ?
  - Traditional survival analysis stratifies responders and non responders.
  - This is equivalent to identifying patients who will not respond well to improve the mean in a statistically equivalent sense.
    - Allows physicians to make actionable choices.



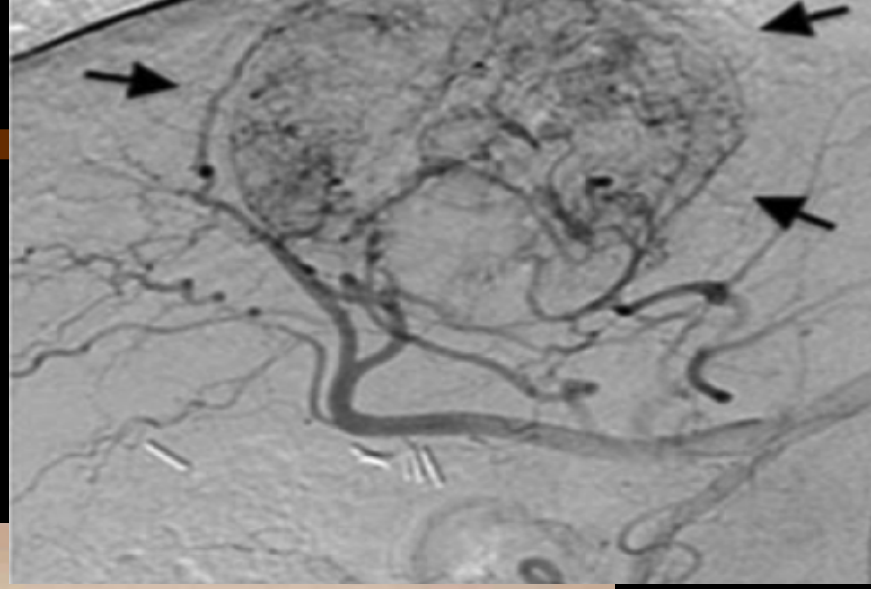
Oden, Farrell 2015



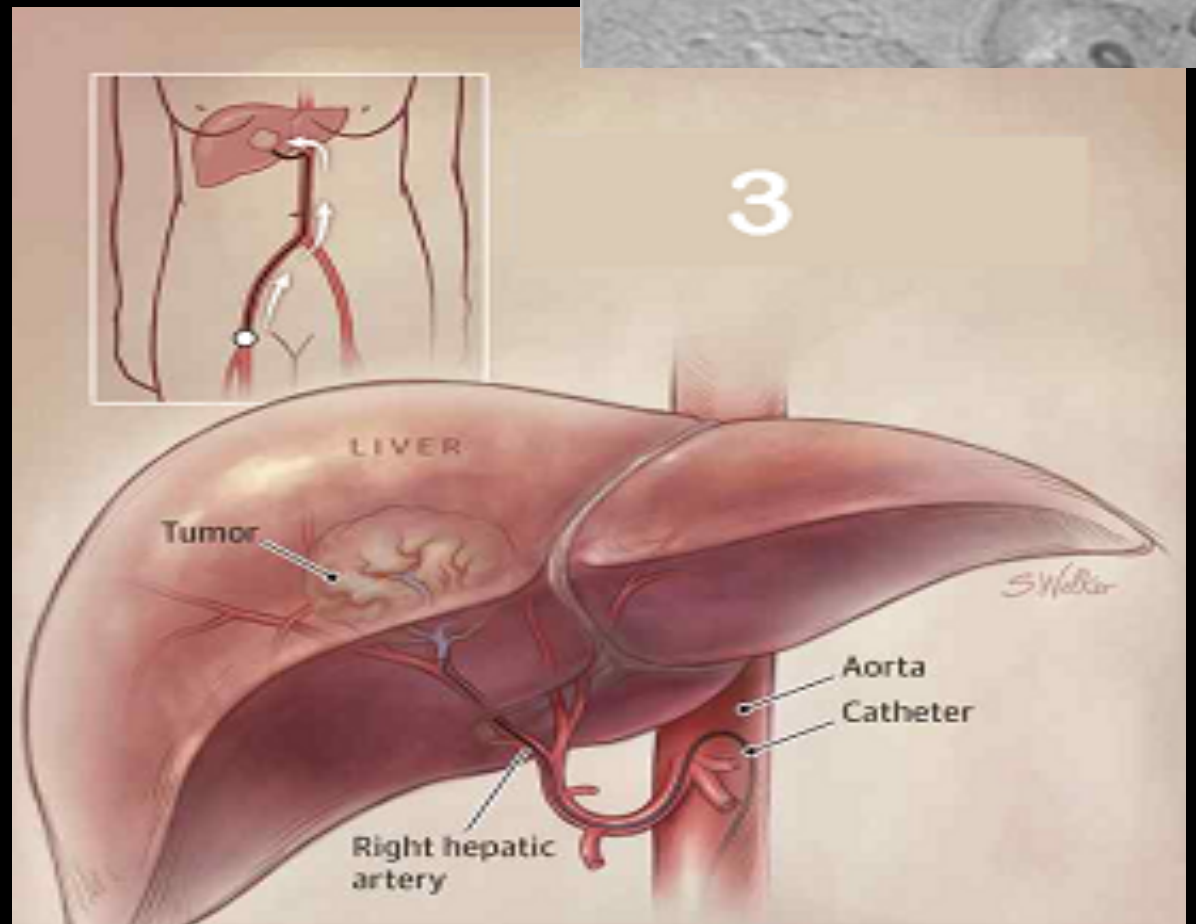
Treatment Overview

- Curative therapy: resection and liver transplant

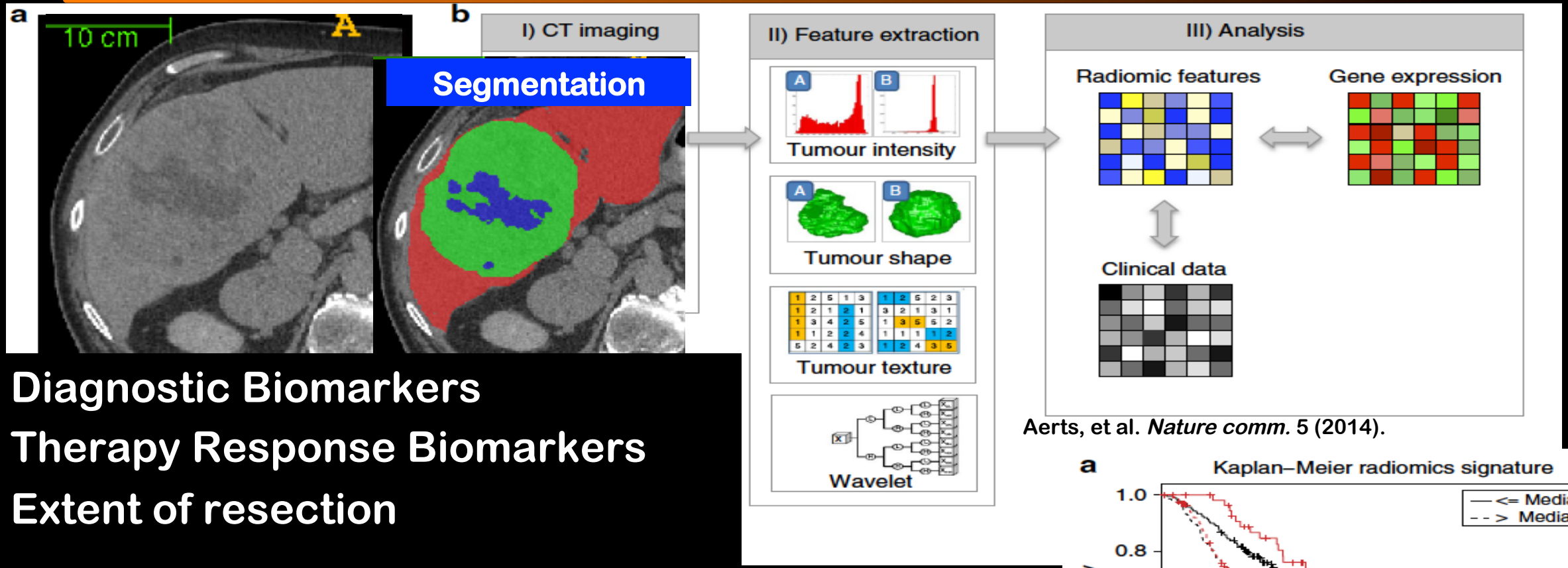
- Otherwise, Survival = 6-20months
  - Ablations does not work for lesions > 3cm
  - TACE does not work for lesions > 7cm
    - Catheter navigated from femoral artery to feeding vessels of tumor → inject chemo and material to stop blood flow to tumor
  - Y90 is used for lesions > 7cm
  - Role of RT is unclear



West, JAMA Oncology 2015

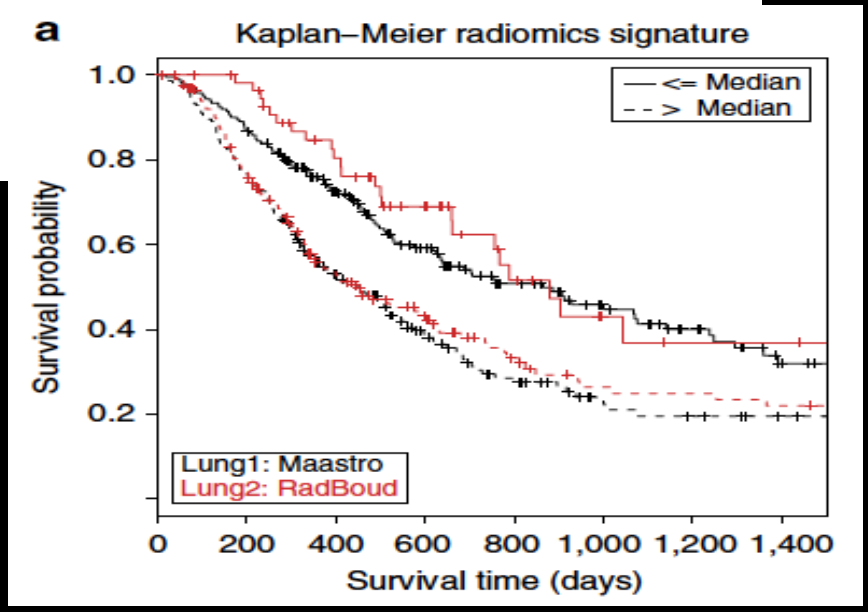


Radiomics, Quantitative Imaging, Computer Aided Detection, etc



Aerts, et al. *Nature comm.* 5 (2014).

- Diagnostic Biomarkers
- Therapy Response Biomarkers
- Extent of resection



Background

- Manual Segmentation, CT
  - Tumor size, arterial enhancement, GLCM homogeneity correlated with response
- Semi-Automated, MR
  - 78% prediction accuracy
- Patient characteristics (N=105)
  - TACE as the sole first-line or initial bridging therapy
  - Availability of multiphasic contrast-enhanced CT images obtained at baseline (on average 3 weeks before TACE) with no image artifacts
  - Clinical endpoint: TTP based on mRECIST

CLINICAL STUDY

Check for updates

Predicting Treatment Response to Intra-arterial Therapies for Hepatocellular Carcinoma with the Use of Supervised Machine Learning—An Artificial Intelligence Concept

Aaron Abajian, MD, Nikitha Murali, BA, Lynn Jeanette Savic, MD, Fabian Max Laage-Gaupp, MD, Nariman Nezami, MD, James S. Duncan, PhD, Todd Schlachter, MD, MingDe Lin, PhD, Jean-François Geschwind, MD, and Julius Chapiro, MD

Prediction of Therapeutic Response of Hepatocellular Carcinoma to Transcatheter Arterial Chemoembolization Based on Pretherapeutic Dynamic CT and Textural Findings

Hyun Jeong Park[1]

OBJECTIVE. The objective of our study was to assess the value of CT texture analysis

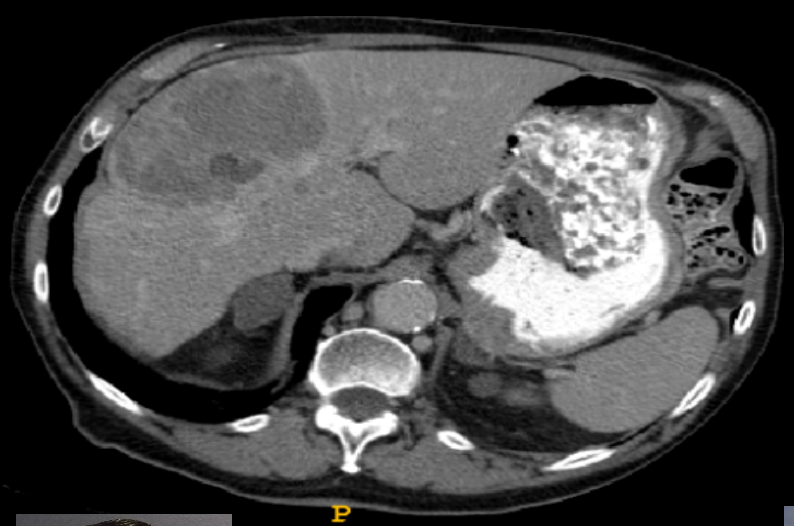
Step 1 – Segmentation

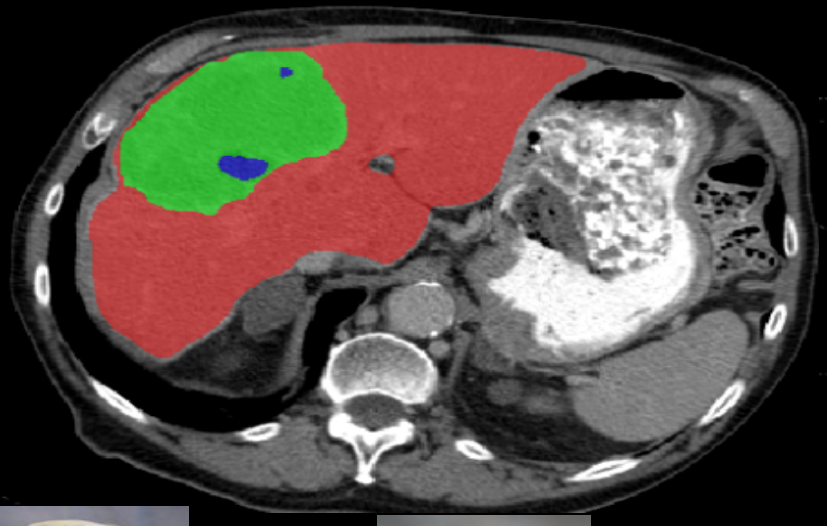
- **Training data**

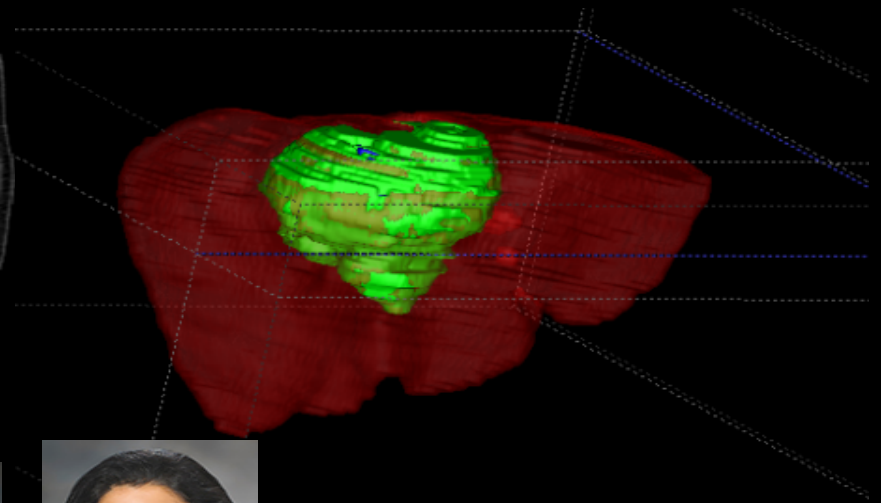
Study cohort of 105 patients with HCC. Each patient with pre-TACE liver CT and subsequent follow up CTs.

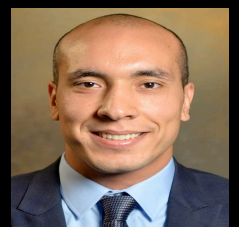
Manual segmentation of all lesions with repeats. Segmented portions included background liver, viable tumor, and necrotic tissue.

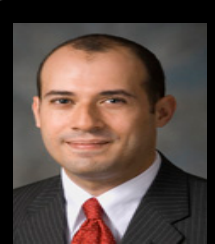
Random Forest (RF) classifier trained on the manually segmented livers.







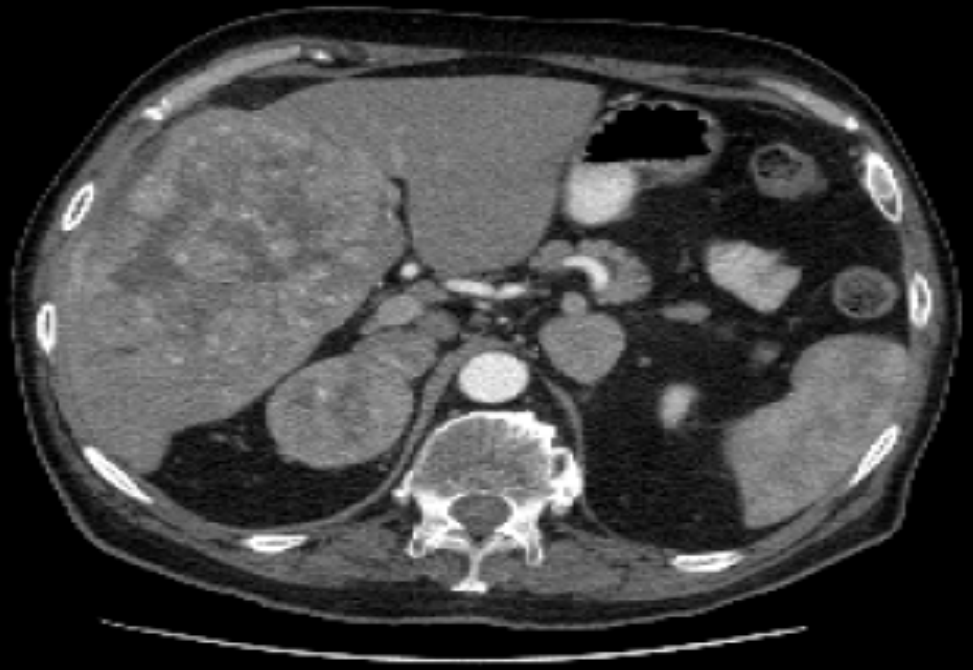


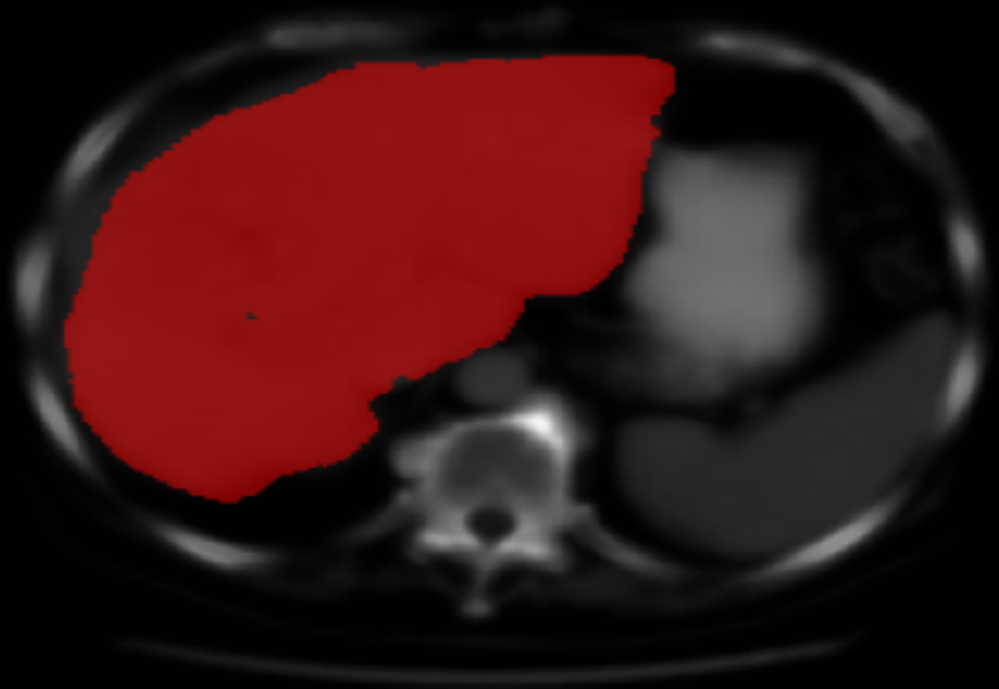


Pipeline – segmentation

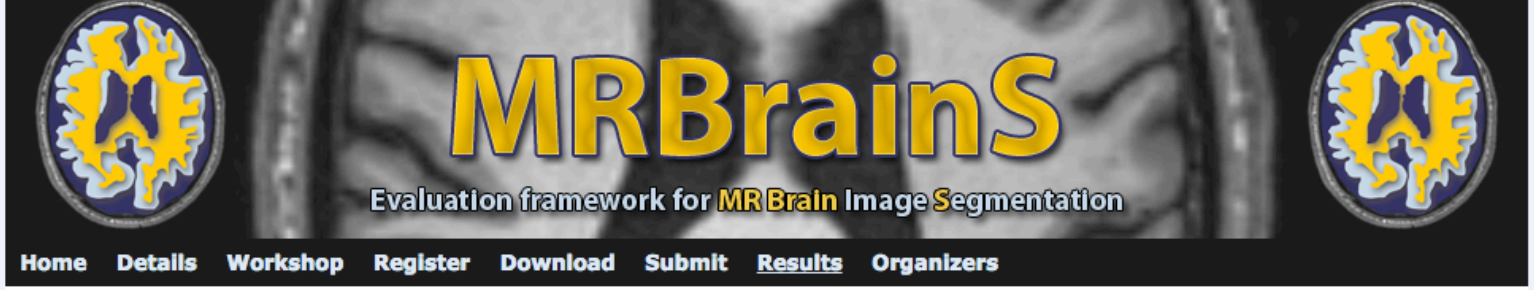
Publically available neural network models used to generate the initial liver mask.

Trained RF model used to automatically segment viable, necrotic tumor and vessels.





Automatic Liver and Tumor Segmentation of CT and MRI Volumes Using Cascaded Fully Convolutional Neural Networks
Christa P F, Ettlinger F et al arXiv:1702.05970v2 [cs.CV] 23 Feb 2017



Results

When teams **submit** their segmentation results, the evaluation results will be sent to the team contact person by e-mail and will be listed below.

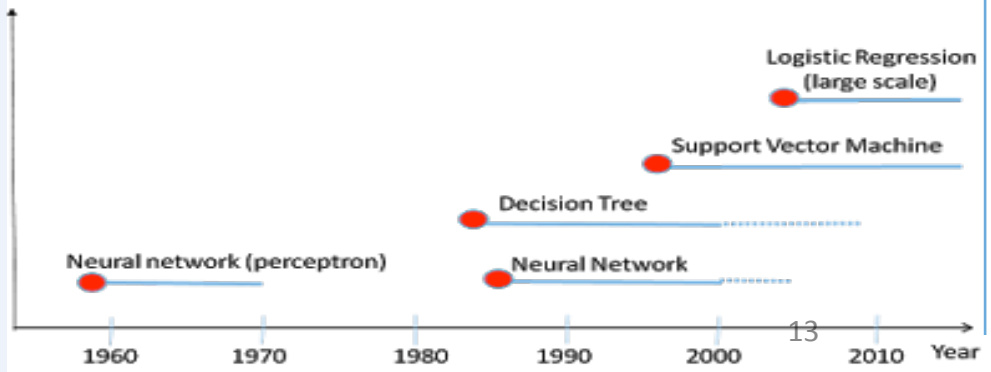
Rank	Team name	Submission name	Date	Sum Scores	Sequences used	Speed	Doc
1	TailHot	Multi-modality aggregation network[3]	13-04-18	39	T1; T1_IR; FLAIR	~13 sec	
2	WTA2	3D Cascade convolutional architecture - Method 2[2]	23-05-18	47	T1; T1_IR; FLAIR	~2 min	
3	LIVIA_ETS	HyperDenseNet[2]	06-02-18	64	T1; T1_IR; FLAIR	~4 min	
4	CU_DL2	3D Deep Learning; voxnet2	28-06-16	74	T1; T1_IR; FLAIR	~2 min	
5	CU_DL	3D Deep Learning; voxnet1[3]	16-06-16	77	T1; T1_IR; FLAIR	~2 min	
6	MSL-SKKU	Deep Convolutional Neural Network	19-06-17	82	T1; T1-IR; FLAIR	~1.5 min	
7	LRDE	Fully Convolutional Network	20-12-16	83	T1	~2 sec	
8	MDGRU	Multi-Dimensional Gated Recurrent Units[3]	27-07-16	106	T1; T1_IR; FLAIR	~2 min	
9	FBI/LMB Freiburg	U-Net (3D)	01-05-16	111	T1-1mm; T1-IR; FLAIR	~2 min	
10	PyraMiD-LSTM2	NOCC with rounds[3]	23-05-16	113	T1; T1-IR; FLAIR	~2 min	
11	AOC	Atlas of Classifiers	24-12-17	126	T1	~6 sec	
12	IDSIA	PyraMiD-LSTM	05-06-15	131	T1; T1_IR; FLAIR	~2 min	
13	STH	Hybrid ANN-based Auto-context method[2]	03-06-16	146	T1; T1-IR; FLAIR	~ 5 min	
14	ISI-Neonatology	Multi-stage voxel classification	31-05-14	151	T1	~1.5 hours	
15	UNC-IDEA	LINKS:Learning-based multi-source integration	09-02-15	154	T1; T1_IR; FLAIR	~3 min	
16	BCH_CRL_IMAGINE	3D patch-wise DenseNet and Patch Fusion[2]	24-05-18	178	T1; T1-IR; FLAIR	~2 min	
17	MNAB2	Random Forests	21-02-14	180	T1; T1_IR; FLAIR	~25 min	
18	KSOM GHMF	ASeTs: MAP-Based with Manifold learning	13-05-14	182	T1; T1_IR; FLAIR	~23 min	
19	THUity	Modified U-Net[2]	21-07-18	183	T1; T1_IR; FLAIR	~40 sec	
20	WTA	3D Cascade convolutional architecture - Method 1[3]	15-05-18	202	T1; T1_IR; FLAIR	~5 min	
21	vicorob UdG T1_F	MSSEG using T1 + FLAIR (T1-IR skull)	14-01-16	204	T1; IR; FLAIR	~2 min	
22	VBM12	VBM12_r738 with WMHC=2	07-10-15	206	T1	~6 min	
23	BIGR2	Multi-Feature SVM Classification	26-09-13	207	T1; T1_IR; FLAIR	~35 min	

Evolution of Algorithms in Crowd Sourcing Challenges

2013 – 2014: Classical SVM, RF methods at top

2015: Deep learning

Current: Ensemble of NN, Differentiable programming



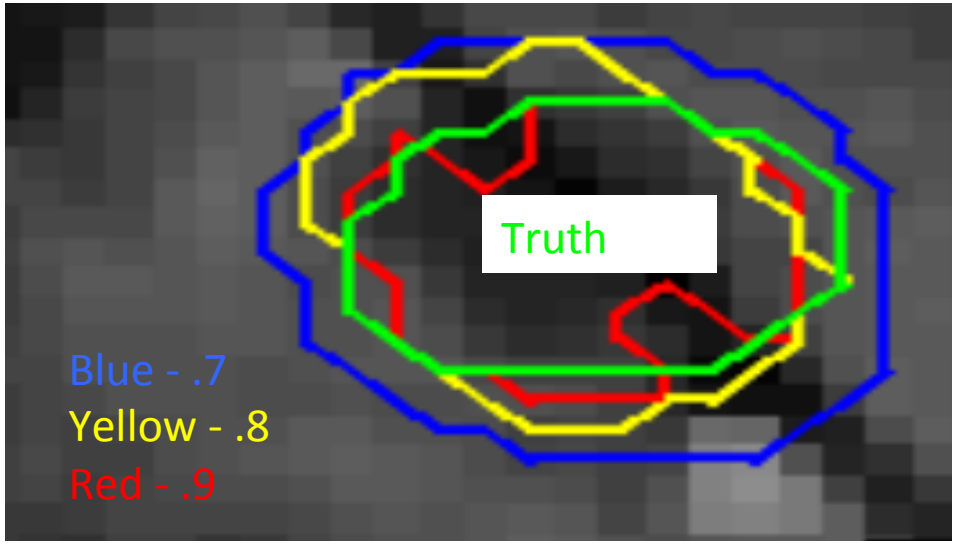
Motivation - Deep Learning

- 2014, DSC = .7

A Bottom-Up Approach for Automatic Pancreas Segmentation in Abdominal CT Scans

Amal Farag, Le Lu, Evrim Turkbey, Jiamin Liu and Ronald M. Summers

Imaging Biomarkers and CAD Laboratory, Department of Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bld. 10. Rm. 1C224D; Bethesda, MD



- 2015, DSC = .8

Jun 2015

DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation

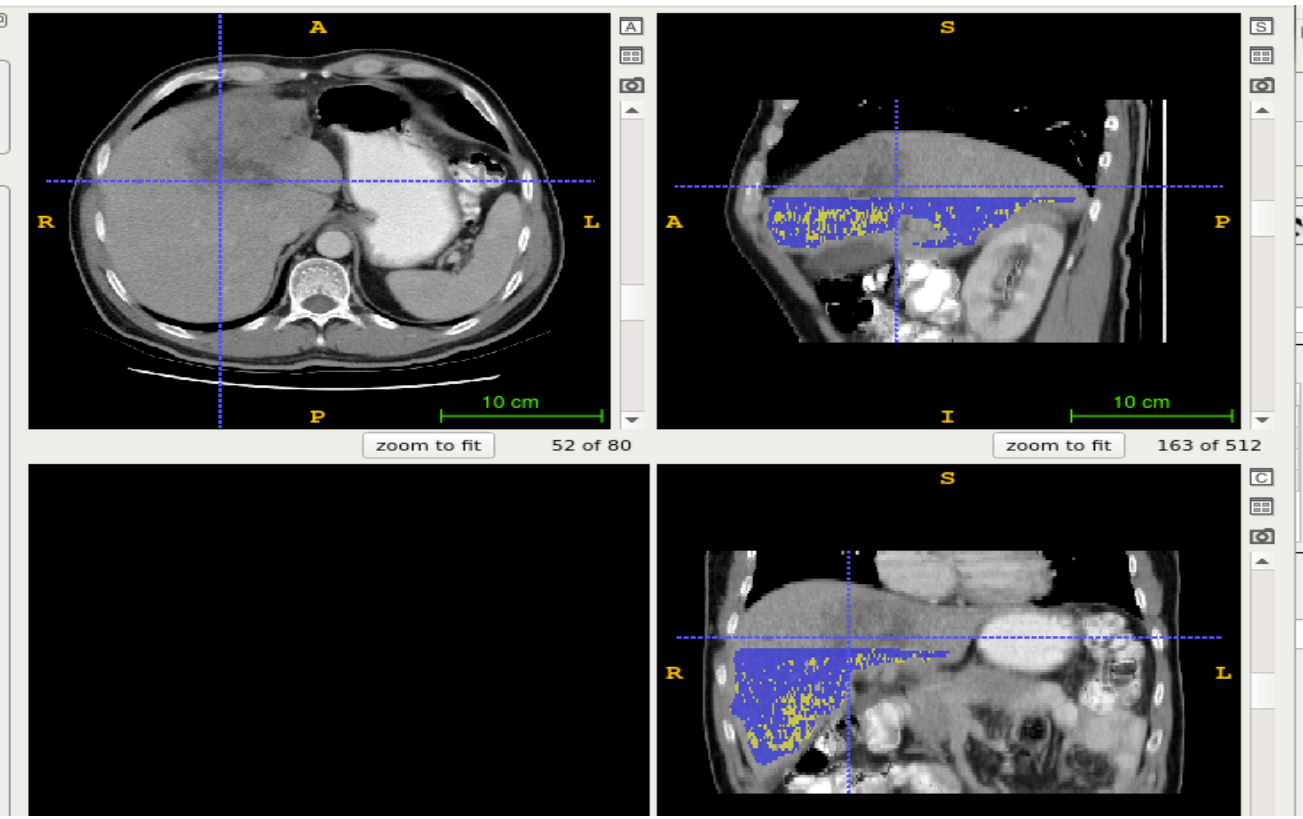
Holger R. Roth*[1], Le Lu[1], Amal Farag[1], Hoo-Chang Shin[1], Jiamin Liu[1], Evrim Turkbey[1], and Ronald M. Summers[1]

[1]Imaging Biomarkers and Computer-Aided Diagnosis Laboratory,, Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bethesda, MD 20892-1182, USA.

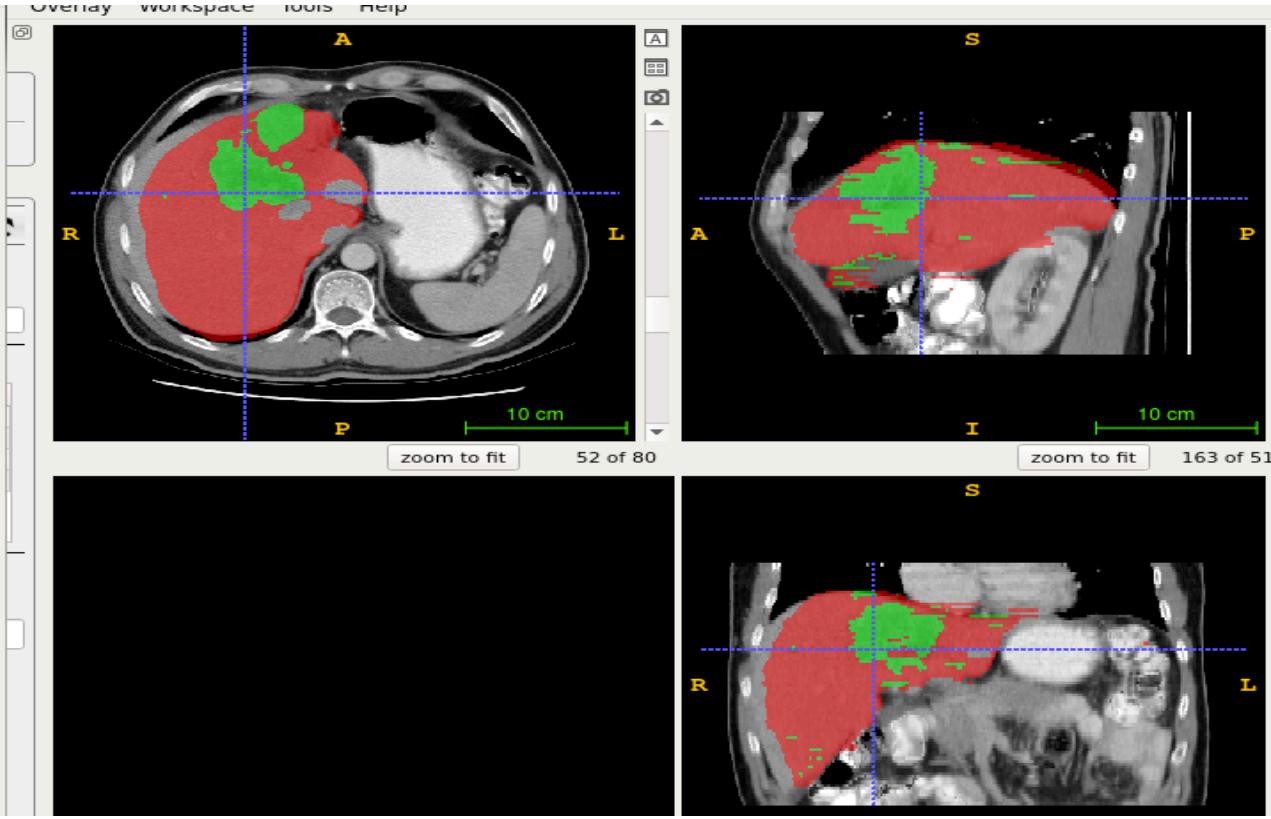
Benefits of Deep Learning

Comparison between manually handcrafted image features (RF) and deep NN

- RF usually works well on the training/cross validation, but not generalizing…

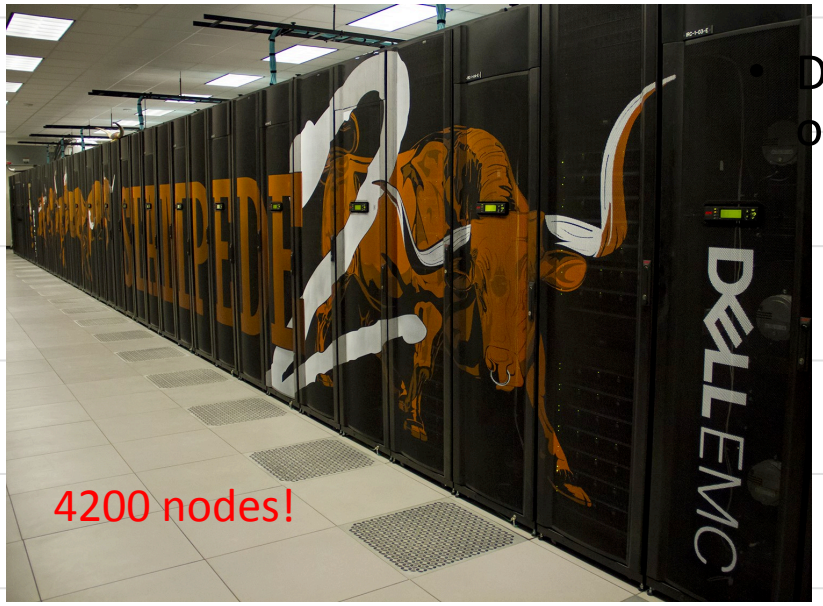


Hand Crafted Features



Learned Image Features

Distributed Training Speedup



- Data size, batch size, number of processors
  - Parallel work needs larger batch → faster epoch but longer to converge

arxiv.org/abs/1709.05011

Image net training on M40 GPU - 14days

2048 KNL nodes – 20 mins

horovod

Performance Normalized to 1 KNL Node

10
9
8
7
6
5
4
3
2
1
0

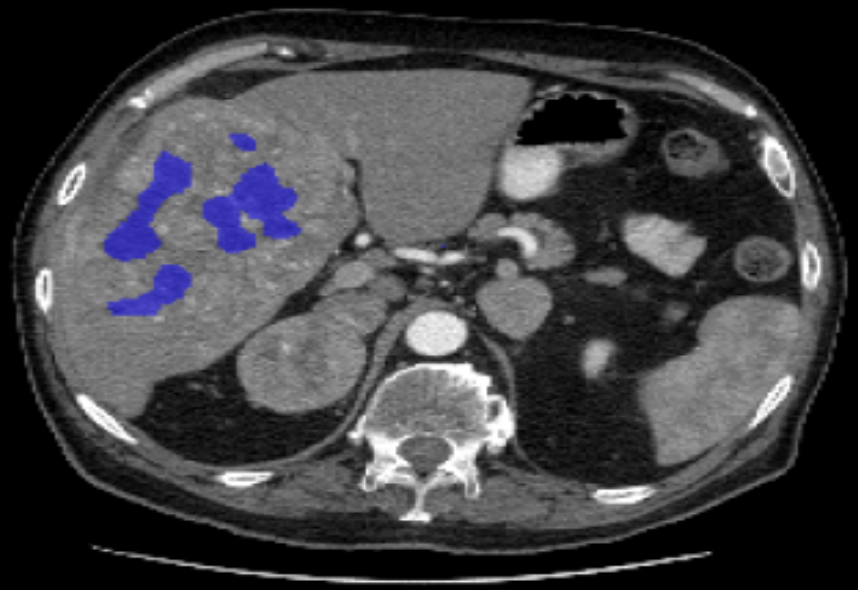
[VALUE]X
1.7X
3.2X
5.3X
8.6X
2.1X
5.4X

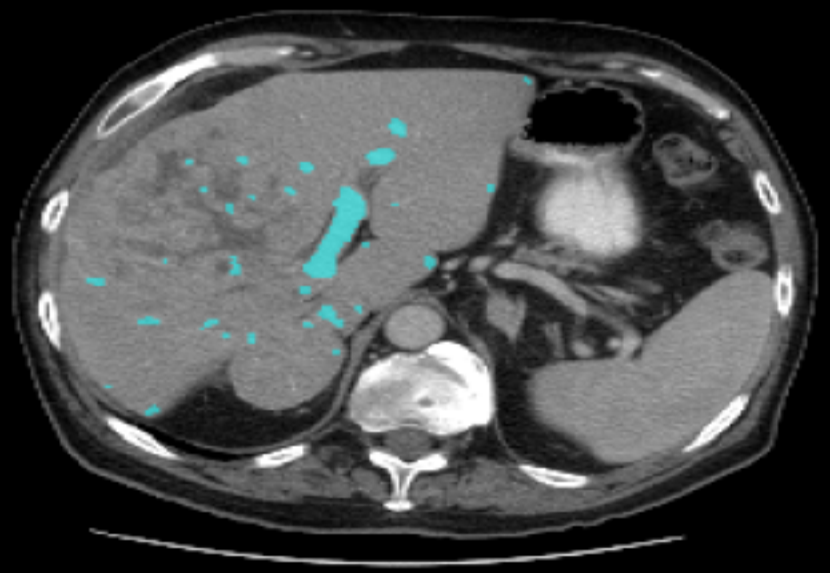
1 Node (batch=30)
2 Nodes (batch=30)
4 Nodes (batch=30)
8 Nodes (batch=30)
16 Nodes (batch=30)
Titan (batch=5)
P5000 (batch=10)

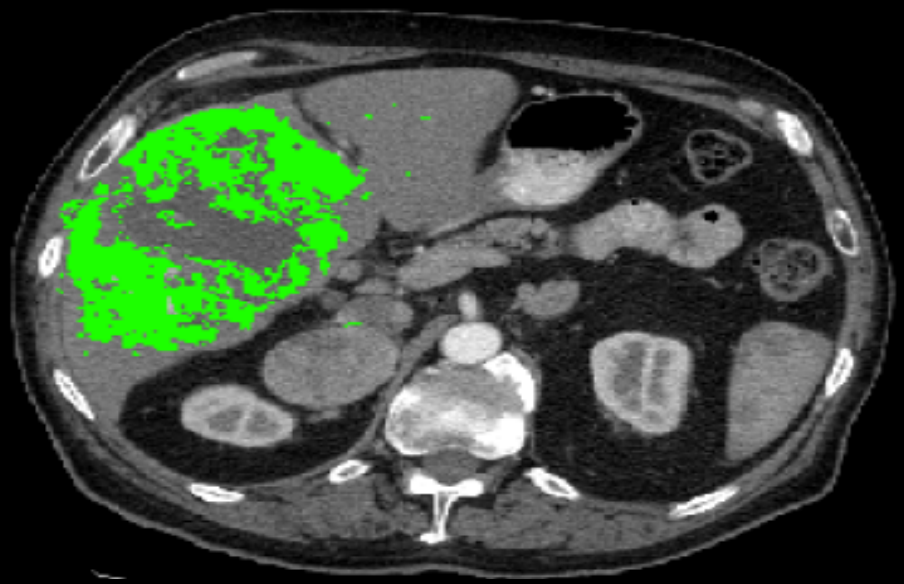
Device

Performance difference is mainly from memory latency issues

Pipeline – segmentation







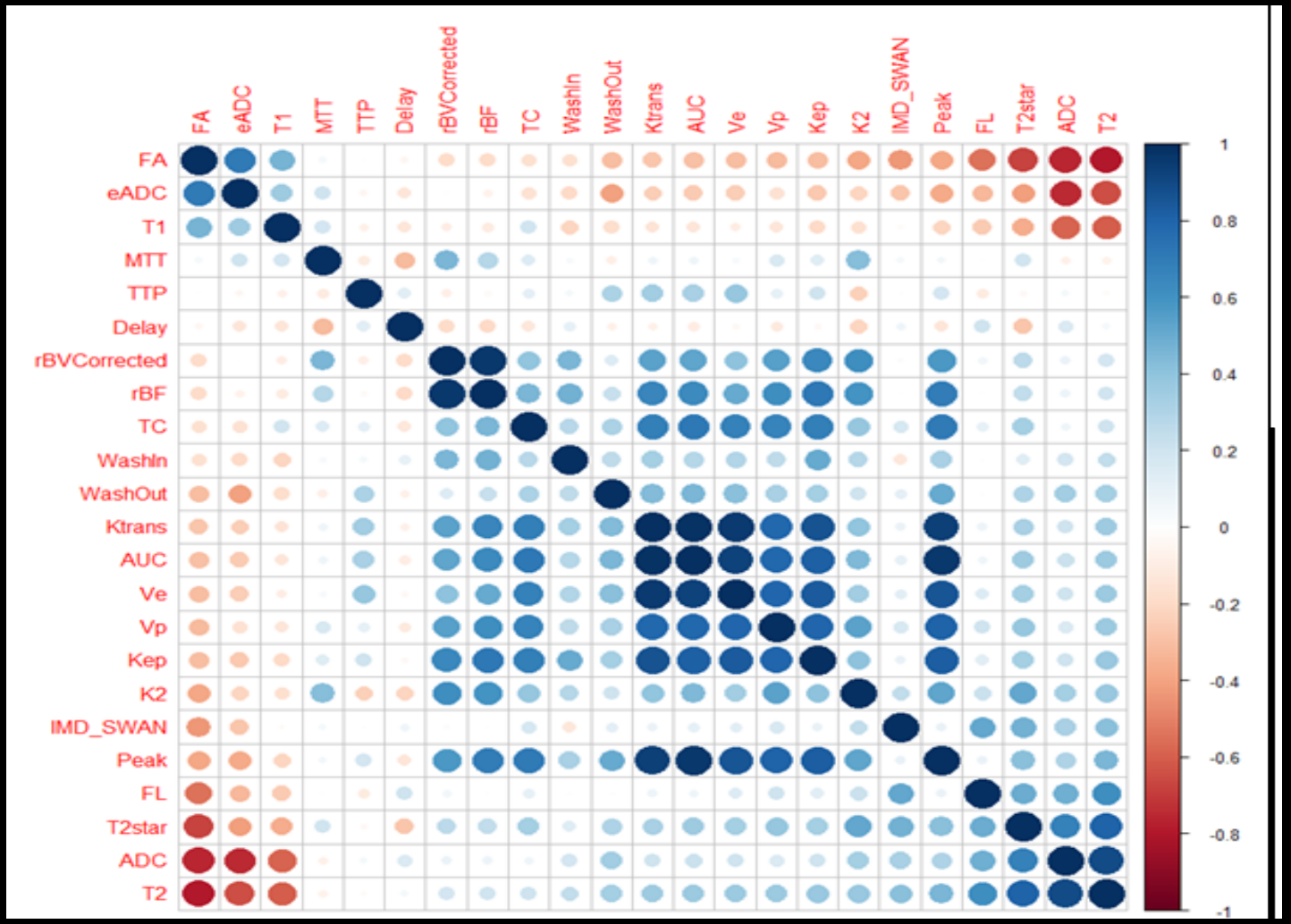
- Label Tumor
  - Enhancing Tumor
  - Necrosis
  - Background liver
  - Vessels
- Fully Automated!
  - 6mo → 2 days
- DSC accuracy for latest models was ~.6-.7 for RF model

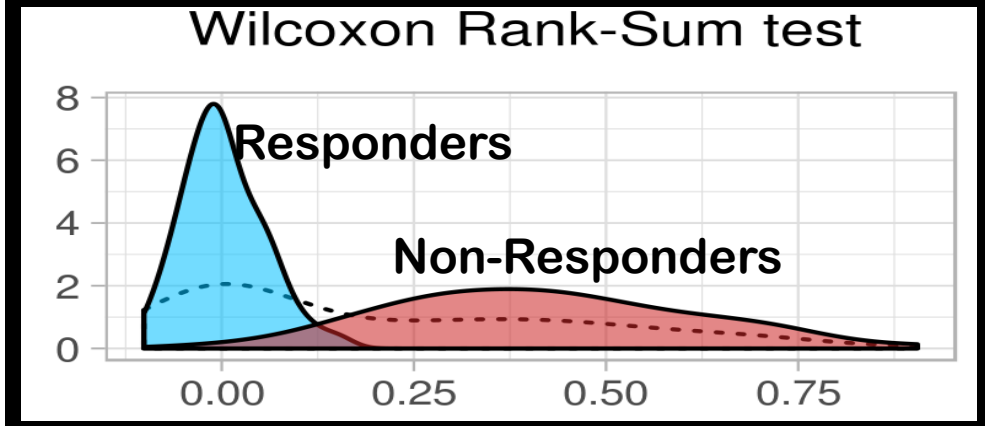
TACE	
viable	necrosis
0.670/0.651 (0.043)	0.766/0.724 (0.026)
0.601/0.551 (0.042)	0.935/0.927 (0.002)

Step 2 - Modeling

- Develop intuition for data – NO FREE LUNCH!
  - No correlation → not likely to predict

Input Variable Correlations

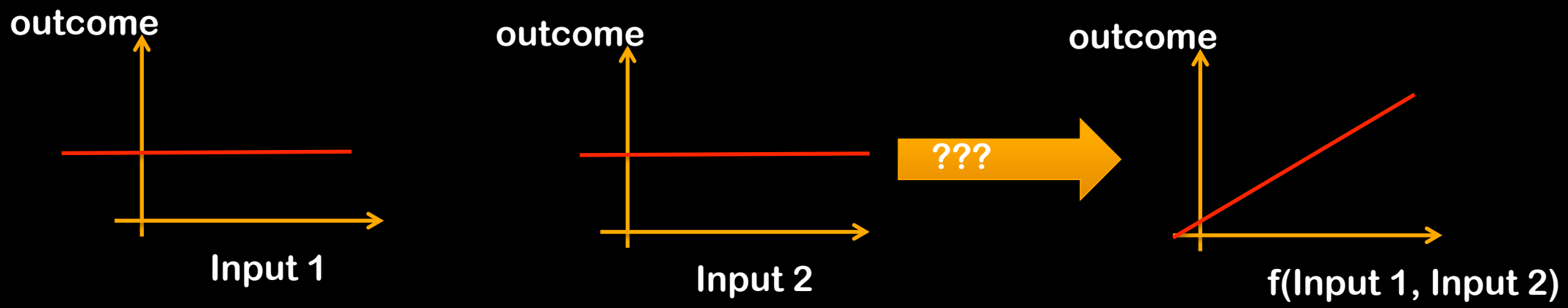




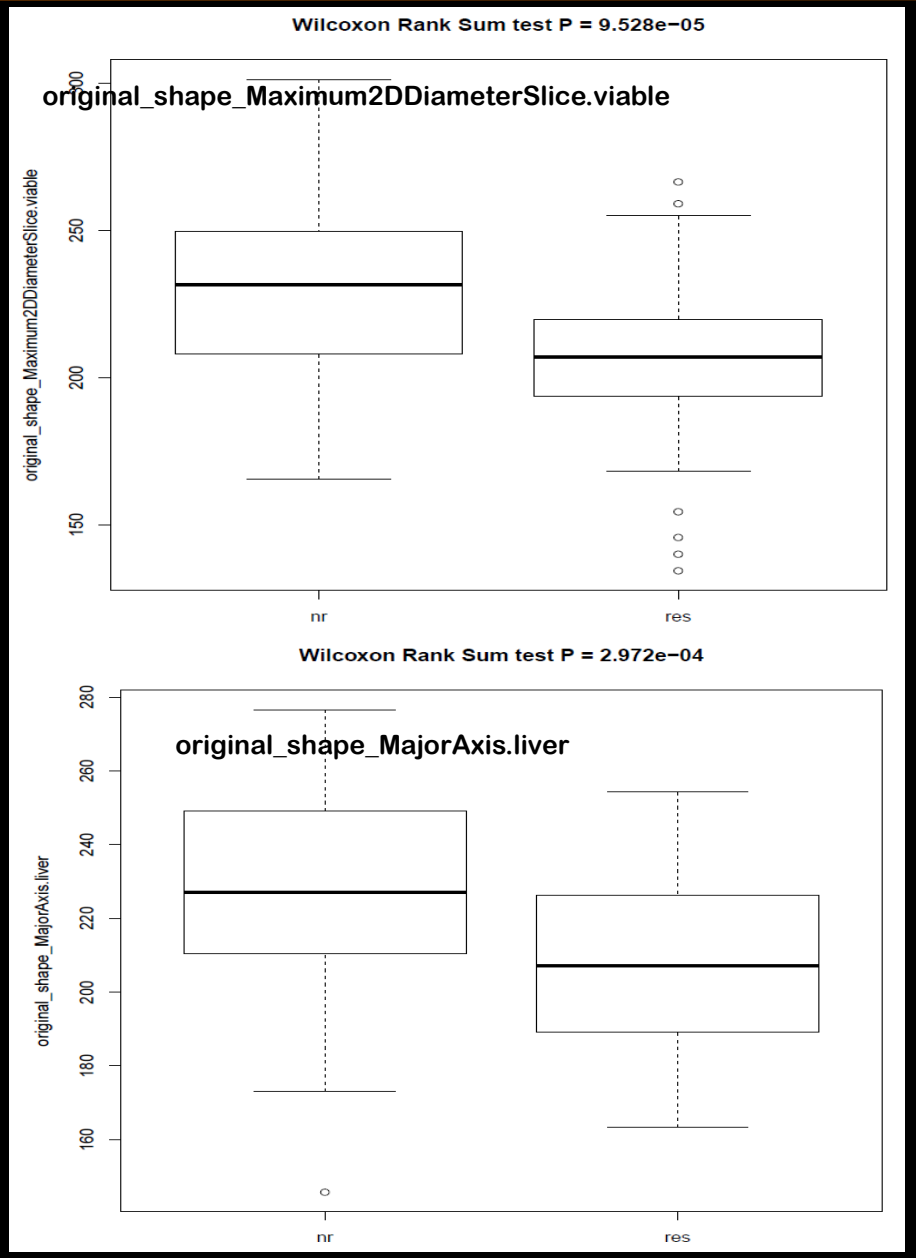
- Data Reduction
  - reduce model overfit
- Model Sensitivity Analysis
- Calibration
- Cross-Validation

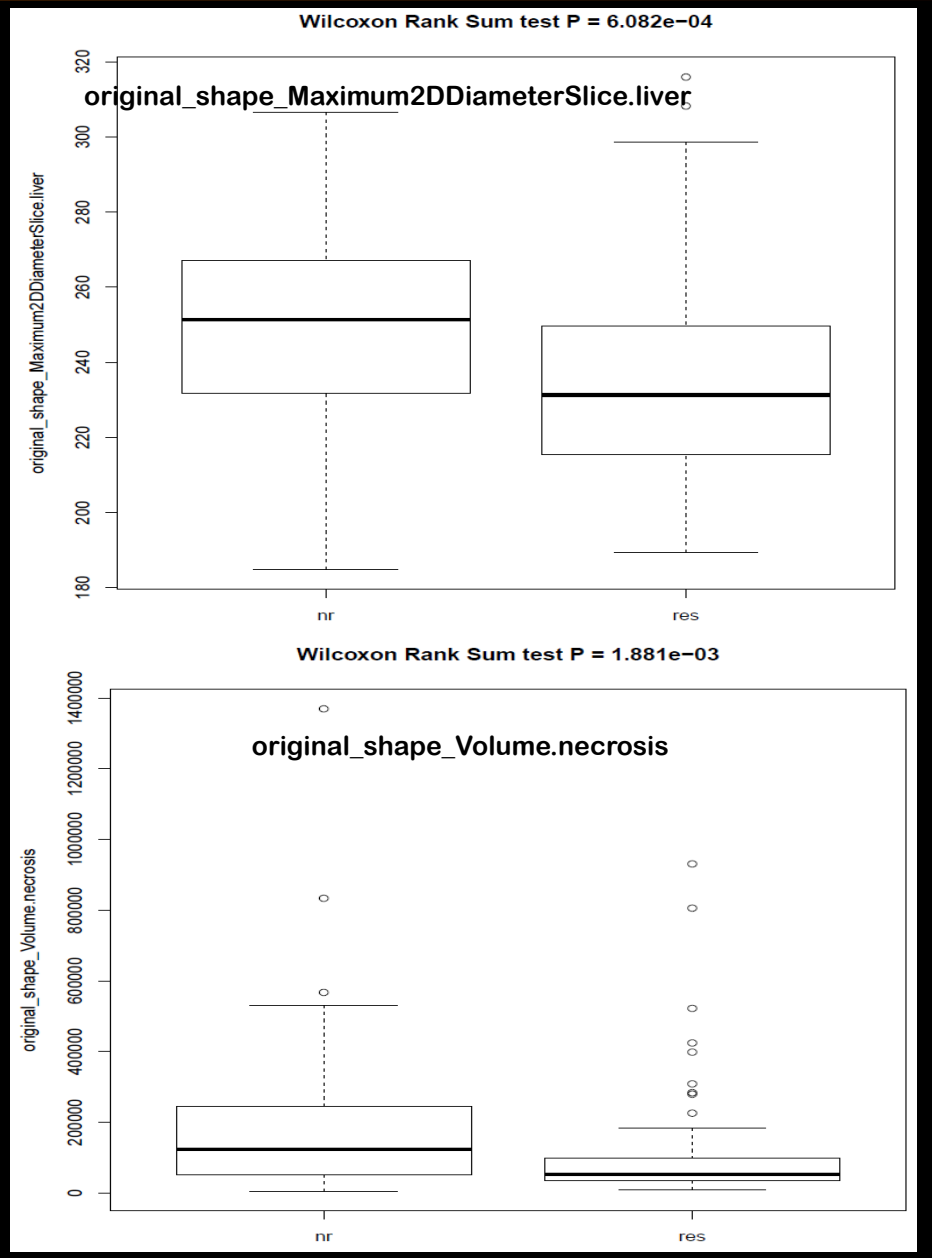
Modeling Principles

- What are the fundamental principles ?
  - Correlated inputs provide no new information
  - Correlates input with output are likely to build a predictive model
- Nonlinear combination of two uncorrelated inputs correlated with output ?... le magic

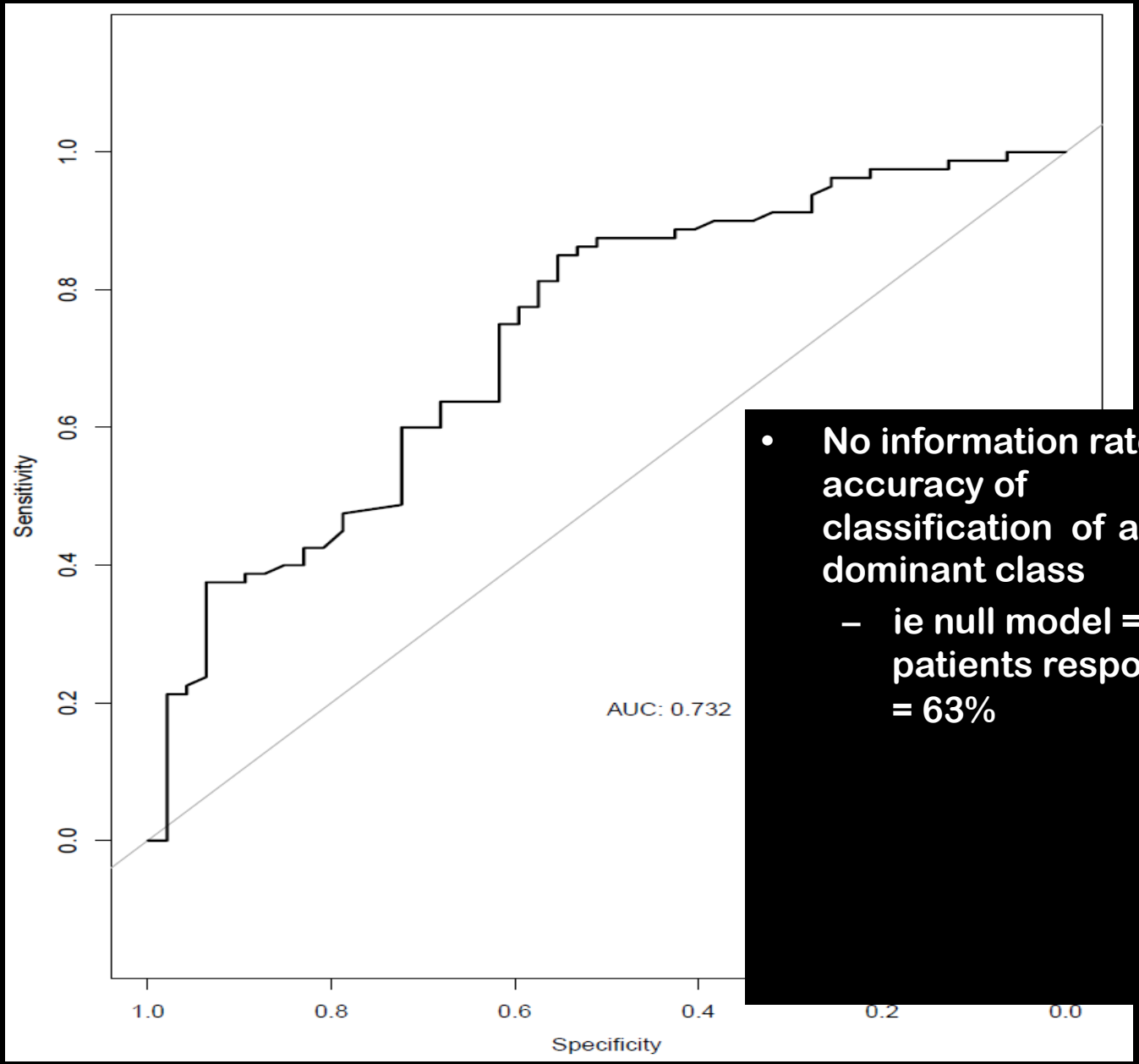


Pipeline – (Boruta method feature correlation)





Pipeline – prediction results



- No information rate = accuracy of classification of all the dominant class
  - ie null model = all patients responders = 63%

```
                Accuracy : 0.7323
                  95% CI : (0.6465, 0.8069)
    No Information Rate : 0.6299
    P-Value [Acc > NIR] : 0.009625

                   Kappa : 0.4049
 Mcnemar's Test P-Value : 0.229949

            Sensitivity : 0.8375
            Specificity : 0.5532
         Pos Pred Value : 0.7614
         Neg Pred Value : 0.6667
             Prevalence : 0.6299
         Detection Rate : 0.5276
   Detection Prevalence : 0.6929
      Balanced Accuracy : 0.6953

       'Positive' Class : res
```

true positive (TP)
eqv. with hit

true negative (TN)
eqv. with correct rejection

false positive (FP)
eqv. with false alarm, Type I error

false negative (FN)
eqv. with miss, Type II error

sensitivity, recall, hit rate, or true positive rate (TPR)

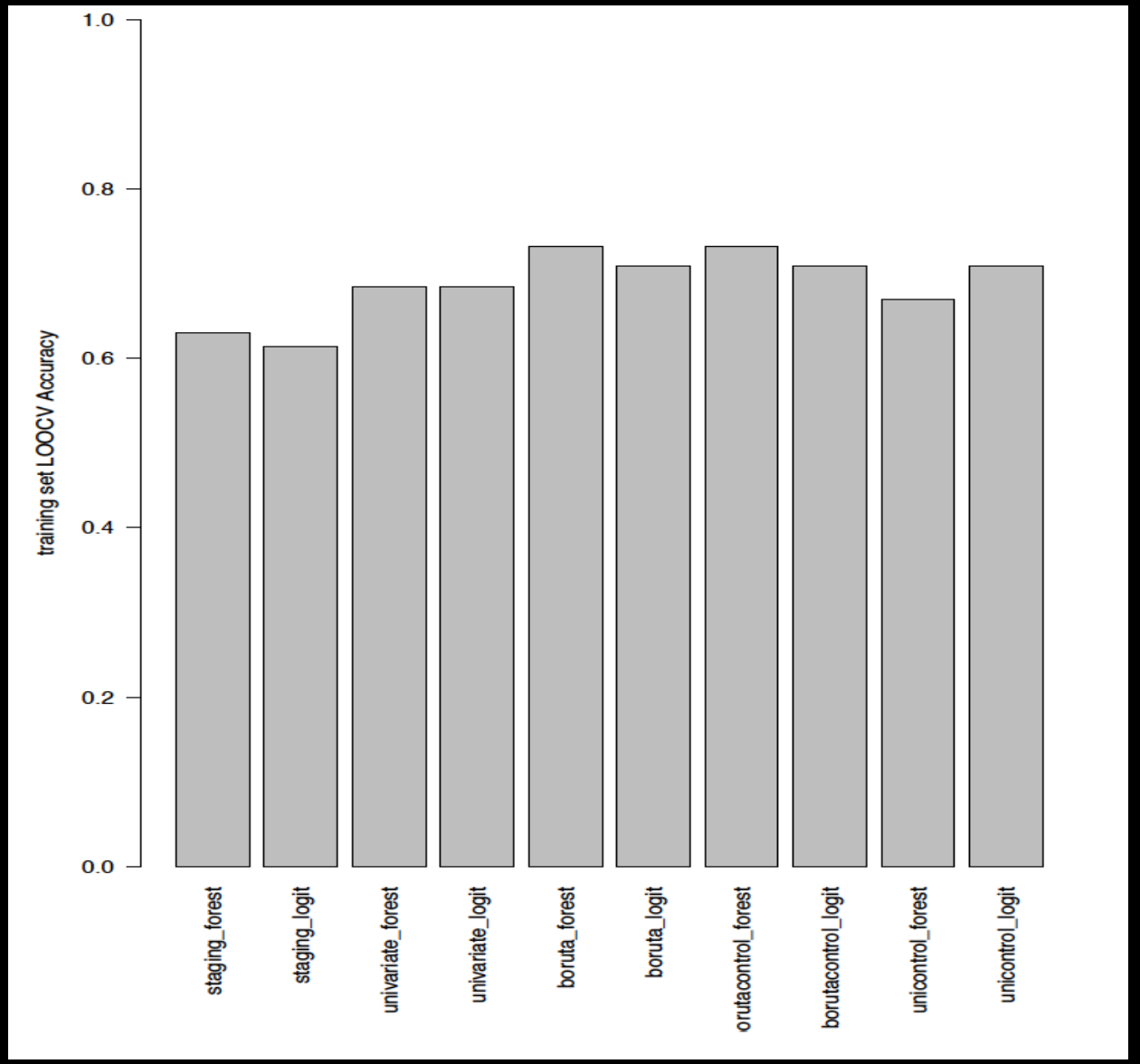
$$\mathrm{TPR} = \frac{\mathrm{TP}}{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} = 1 - \mathrm{FNR}$$

specificity, selectivity or true negative rate (TNR)

$$\mathrm{TNR} = \frac{\mathrm{TN}}{N} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} = 1 - \mathrm{FPR}$$

With and without BCLC

```
## $staging_forest
## [1] 0.6299213
##
## $staging_logit
## [1] 0.6141732
##
## $univariate_forest
## [1] 0.6850394
##
## $univariate_logit
## [1] 0.6850394
##
## $boruta_forest
## [1] 0.7322835
##
## $boruta_logit
## [1] 0.7086614
##
## $borutacontrol_forest
## [1] 0.7322835
##
## $borutacontrol_logit
## [1] 0.7086614
##
## $unicontrol_forest
## [1] 0.6692913
##
## $unicontrol_logit
## [1] 0.7086614
```



Clinical endpoint (TTP)

- TTP based on mRECIST
  - Difficult to improve prognosis for TTP < 5mo

Oncology

Liver Cancer – Clinical Study

Oncology 2017;93(suppl 1):120–126
DOI: 10.1159/000481242

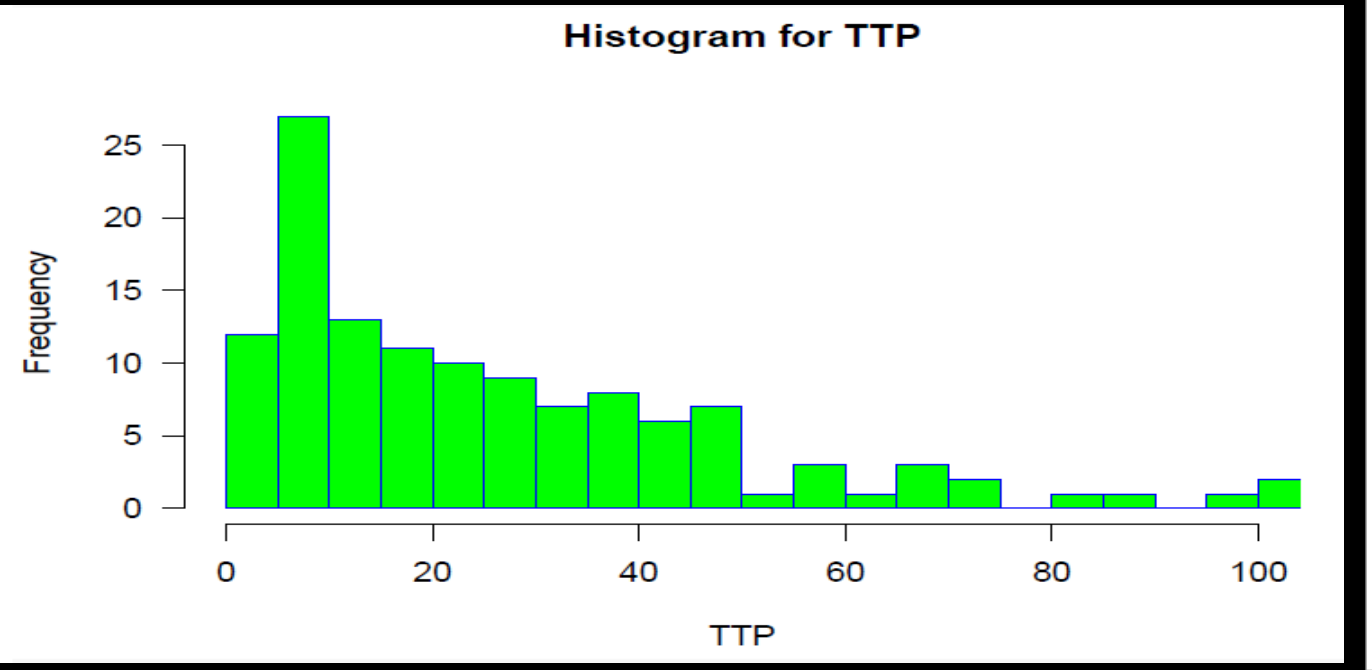
Published

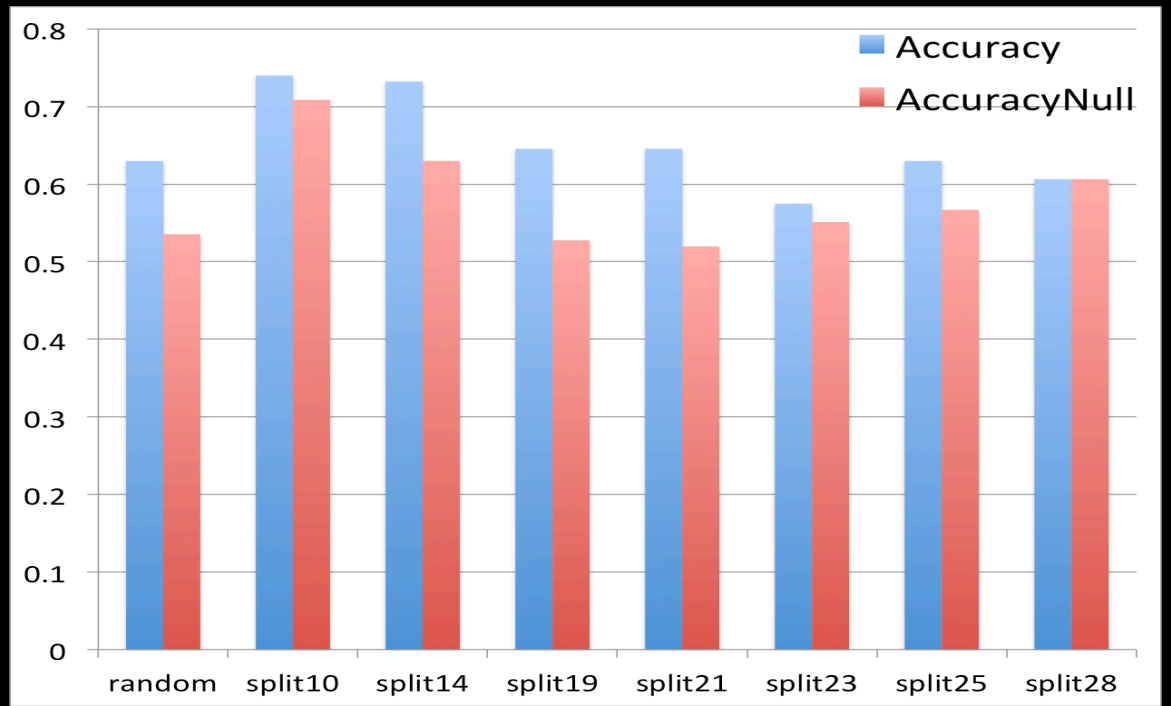
Validation of Newly Proposed Time to Transarterial Chemoembolization Progression in Intermediate-Stage Hepatocellular Carcinoma Cases

Hirofumi Izumoto[a] Atsushi Hiraoka[a] Yoshihiro Ishimaru[b] Tadashi Murakami[b]

A TTP cutoff of 14 weeks was used to stratify patients as follows

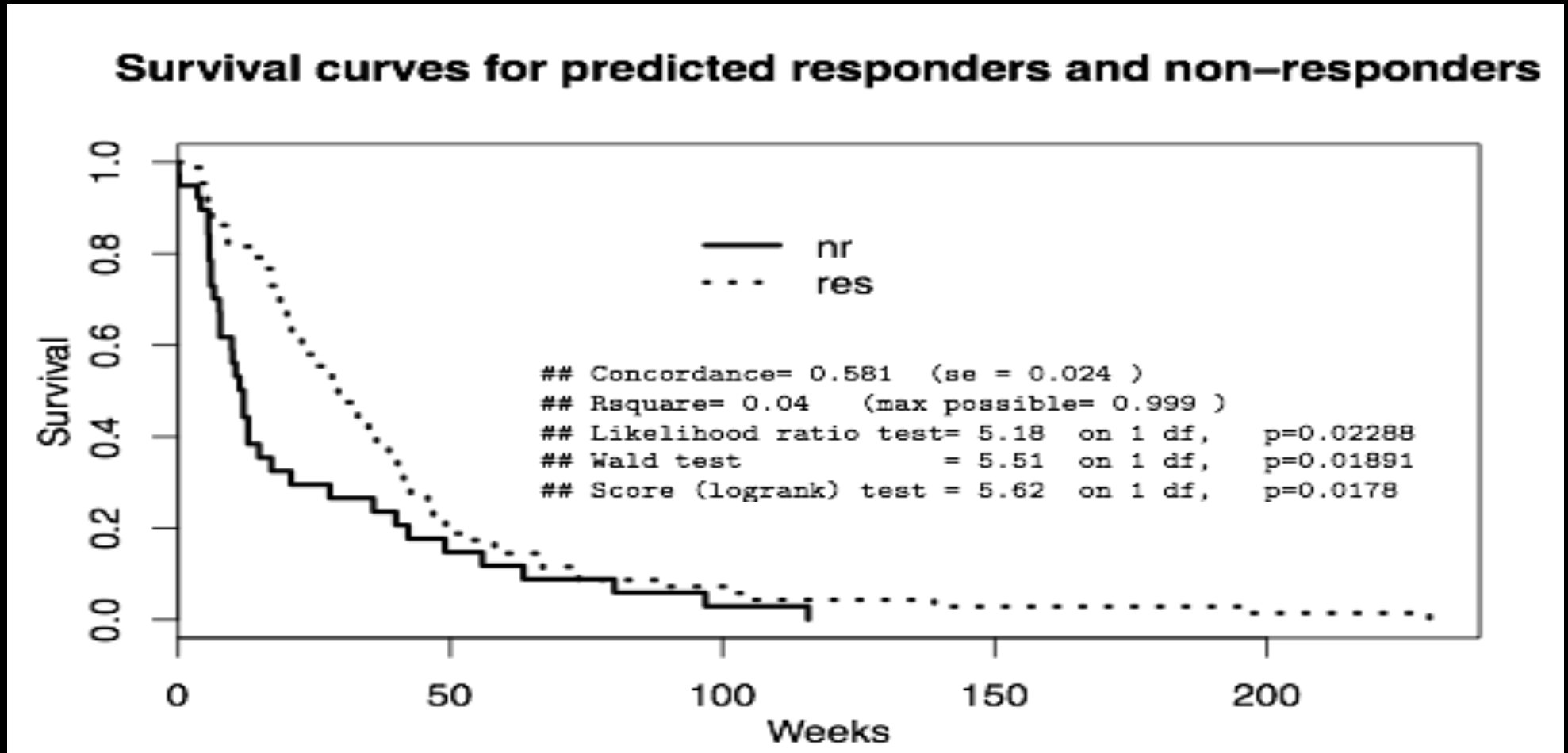
- **TTP ≥ 14 wks** were considered as **TACE response**
- **TTP < 14 wks** were considered as **TACE refractory**





Conclusion

- We found that prediction of HCC response to TACE using quantitative imaging and clinical measurements of pretreatment lesions is a potentially useful clinical tool that can assist in patient selection for TACE.



Future work

- Larger patient populations, prospective trial?
- Natural progression Binary → multiclass → TTP regression
- Has the max accuracy been achieved? Or can HPC find a model with improved prediction accuracy
- Training data update:
  - Need image intensity standardization beyond daily QC protocols. The trigger times of the bolus is different and causes problems.
  - Need landmark registration validation of the different phases
  - Need more training data to eliminate vessels, kidneys, and unwanted segmentation artifacts

MOST IMPORTANT!!! VIEW YOUR DATA!!!

- Tools to view the data are VITAL

imaginguids.csv - LibreOffice Calc

File Edit View Insert Format Tools Data Window Help

Arial 10

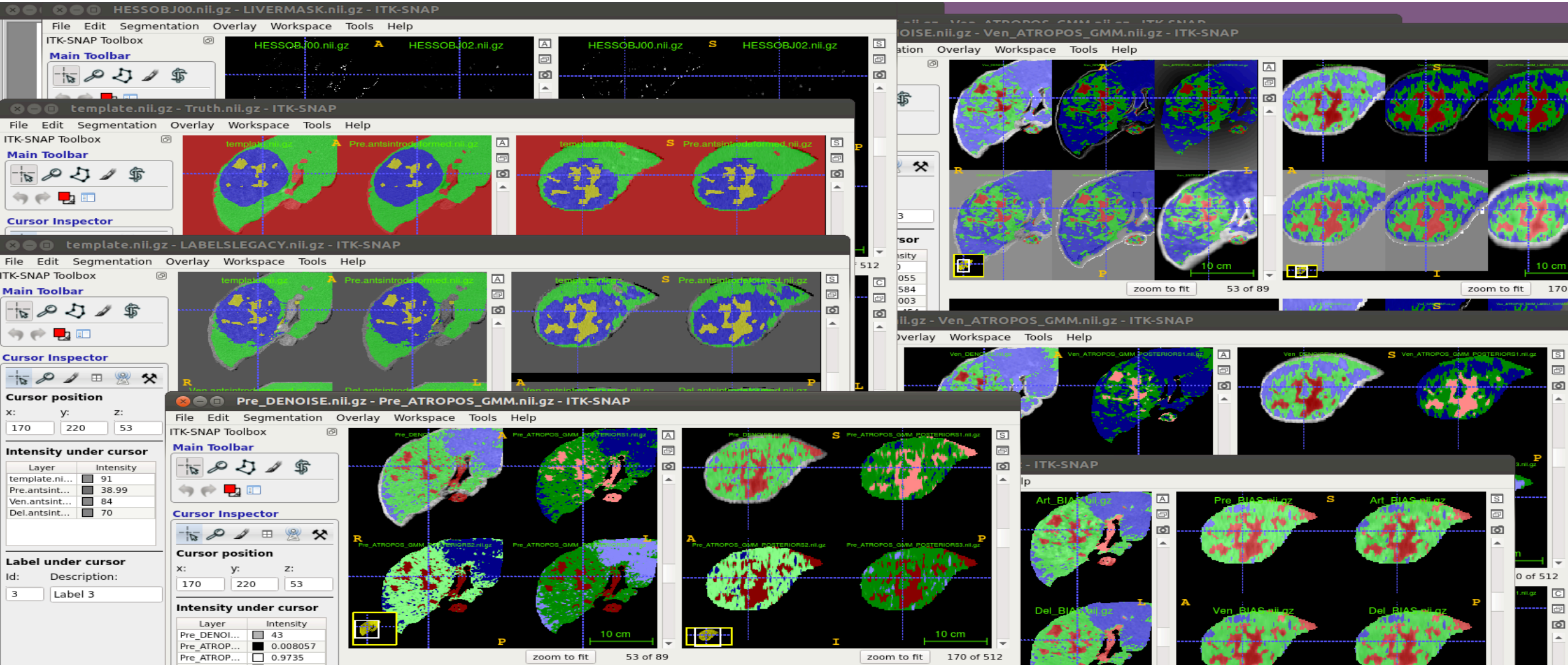
A1 id

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	LesionNumber	NeedDropDownQA	dataID	IHCCdataID	CRCdataID	truthID	responderID	StudyUIDPath	QATruth	LiverMask	RFModel	NNModel	QASolution ForceView=1	QADialogue	TimeID
2	1	4	0			0	0	ImageDatabas	#NAME?	True	True	True	Usable	#####SelectFromDialogueDropDown#####	baseline
3	2	4	3			3	2	ImageDatabas	#NAME?	True	True	False	DelRegistrationError	#####SelectFromDialogueDropDown#####	baseline
4	3	4	0			0	0	ImageDatabas	#NAME?	True	True	True	ArtRegistrationError	#####SelectFromDialogueDropDown#####	baseline
5	4	2	0			0	0	ImageDatabas	#NAME?	True	True	True	Usable	#####SelectFromDialogueDropDown#####	baseline
6	5	3	3			4	1	ImageDatabas	#NAME?	True	True	False	MaskError	#####SelectFromDialogueDropDown#####	baseline
7	6	1	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
8	7	4	0			0	0	ImageDatabas	#NAME?	True	True	True	MaskError	#####SelectFromDialogueDropDown#####	baseline
9	8	3	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
10	9	3	0			0	0	ImageDatabas	#NAME?	True	True	True	MaskError	#####SelectFromDialogueDropDown#####	baseline
11	10	3	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
12	11	2	0			0	0	ImageDatabas	#NAME?	True	True	True	MaskError	#####SelectFromDialogueDropDown#####	baseline
13	12	1	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
14	13	2	0			0	0	ImageDatabas	#NAME?	True	True	True	ArtRegistrationError	#####SelectFromDialogueDropDown#####	baseline
15	14	3	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
16	15	2	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
17	16	3	0			0	0	ImageDatabas	#NAME?	True	True	True	DelRegistrationError	#####SelectFromDialogueDropDown#####	baseline
18	17	3	3			3	2	ImageDatabas	#NAME?	True	True	False	MaskError	#####SelectFromDialogueDropDown#####	baseline
19	18	3	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
20	19	4	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
21	20	2	3			2	1	ImageDatabas	#NAME?	True	True	False	ProtocolErrorNeedNewArterial	#####SelectFromDialogueDropDown#####	baseline
22	21	3	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
23	22	3	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
24	23	3	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
25	24	3	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
26	25	4	0			0	0	ImageDatabas	#NAME?	True	True	True	ArtRegistrationError	#####SelectFromDialogueDropDown#####	baseline
27	26	4	0			0	0	ImageDatabas	#NAME?	True	True	True	MaskError	#####SelectFromDialogueDropDown#####	baseline
28	27	4	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
29	28	3	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
30	29	3	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
31	30	3	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
32	31	2	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
33	32	4	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
34	33	2	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
35	34	2	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
36	35	4	0			0	0	ImageDatabas	#NAME?	True	True	True	DelRegistrationError	#####SelectFromDialogueDropDown#####	baseline
37	36	4	0			0	0	ImageDatabas	#NAME?	True	True	True	MaskError	#####SelectFromDialogueDropDown#####	baseline
38	37	2	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline
39	38	2	0			0	0	ImageDatabas	#NAME?	True	True	True	DelRegistrationError	#####SelectFromDialogueDropDown#####	baseline
40	39	2	0			0	0	ImageDatabas	#NAME?	True	True	True	ArtRegistrationError	#####SelectFromDialogueDropDown#####	baseline
41	40	2	0			0	0	ImageDatabas	#NAME?	True	True	True	VenRegistrationError	#####SelectFromDialogueDropDown#####	baseline
42	41	2	0			0	0	ImageDatabas	#NAME?	True	True	True	ArtRegistrationError	#####SelectFromDialogueDropDown#####	baseline
43	42	3	0			0	0	ImageDatabas	#NAME?	True	True	True	VenRegistrationError	#####SelectFromDialogueDropDown#####	baseline
44	43	4	0			0	0	ImageDatabas	#NAME?	True	True	True	VenRegistrationError	#####SelectFromDialogueDropDown#####	baseline
45	44	4	0			0	0	ImageDatabas	#NAME?	True	True	True	DelRegistrationError	#####SelectFromDialogueDropDown#####	baseline
46	45	2	0			0	0	ImageDatabas	#NAME?	True	True	True	TumorLabelNotFound	#####SelectFromDialogueDropDown#####	baseline

Sheet1

Sheet 1 / 1 Default STD Sum=0

Most Important!! View your data!!



NOT a relational database

Challenges - Database

STILL USING TCIA/ MICCAI/ CROWDSOURCING DATA

Landscape of various database highlight opportunities for data science collaborative efforts for the various database to communicate

Redundancy in manual data effort = $$

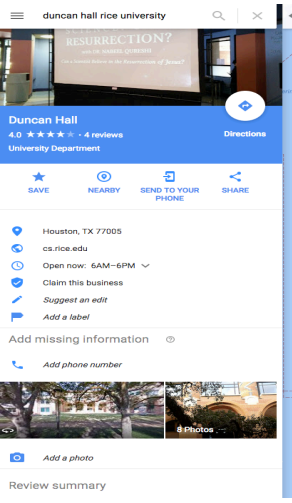
Bottleneck: Data curation

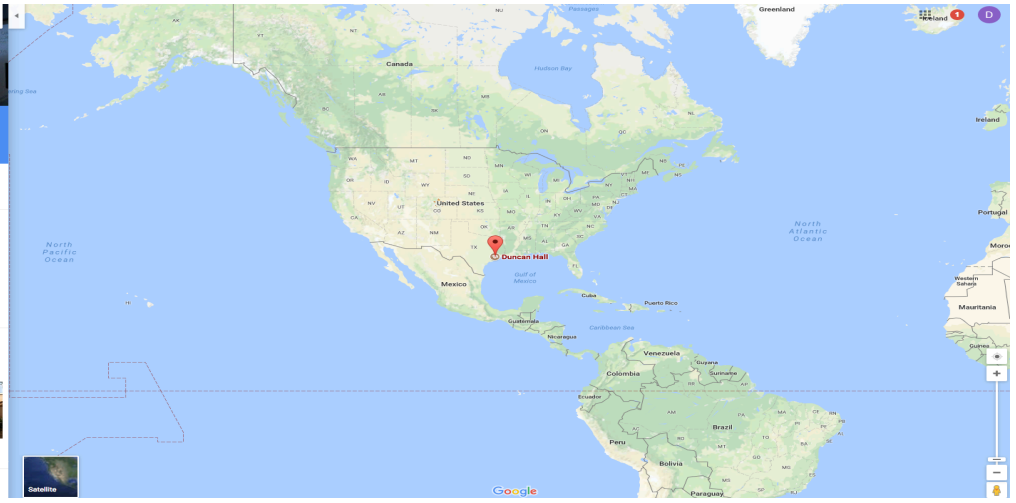
All researchers work together to develop useful infrastructure and not interfere with clinical operations or jeopardize PHI

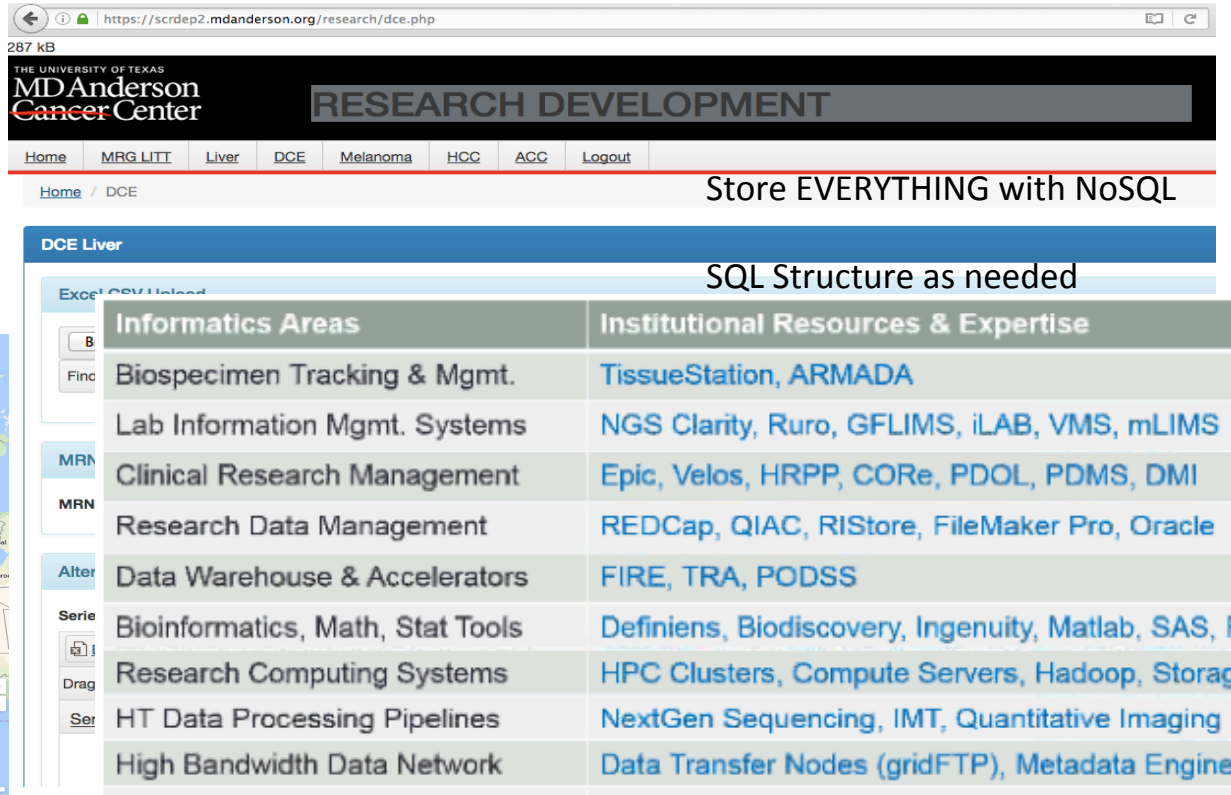
Leadership

Curation > 90% of time/effort

Database visualization



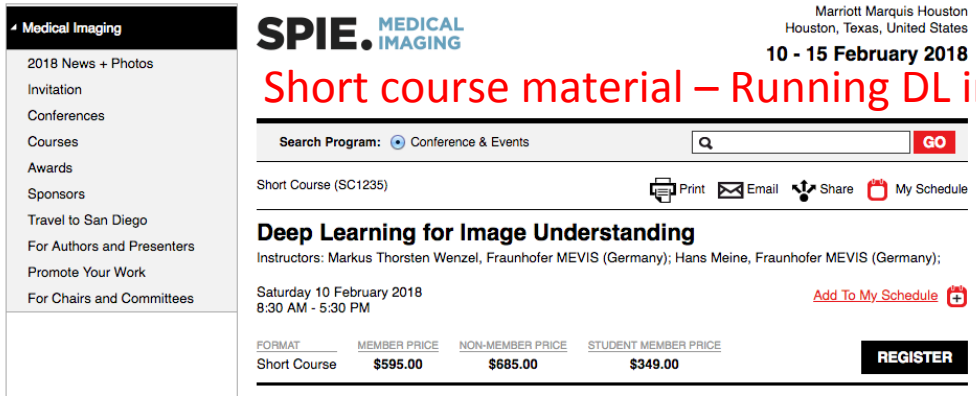




Store EVERYTHING with NoSQL

SQL Structure as needed

Informatics Areas	Institutional Resources & Expertise
Biospecimen Tracking & Mgmt.	TissueStation, ARMADA
Lab Information Mgmt. Systems	NGS Clarity, Ruro, GFLIMS, iLAB, VMS, mLIMS
Clinical Research Management	Epic, Velos, HRPP, CORe, PDOL, PDMS, DMI
Research Data Management	REDCap, QIAC, RIStore, FileMaker Pro, Oracle
Data Warehouse & Accelerators	FIRE, TRA, PODSS
Bioinformatics, Math, Stat Tools	Definiens, Biodiscovery, Ingenuity, Matlab, SAS, R
Research Computing Systems	HPC Clusters, Compute Servers, Hadoop, Storage
HT Data Processing Pipelines	NextGen Sequencing, IMT, Quantitative Imaging
High Bandwidth Data Network	Data Transfer Nodes (gridFTP), Metadata Engine
Computational/Data Sci Expertise	BCB & Biostatistics Depts., IS Div. (IAI & RISTS)



Short course material – Running DL in a Day

DeepEverything

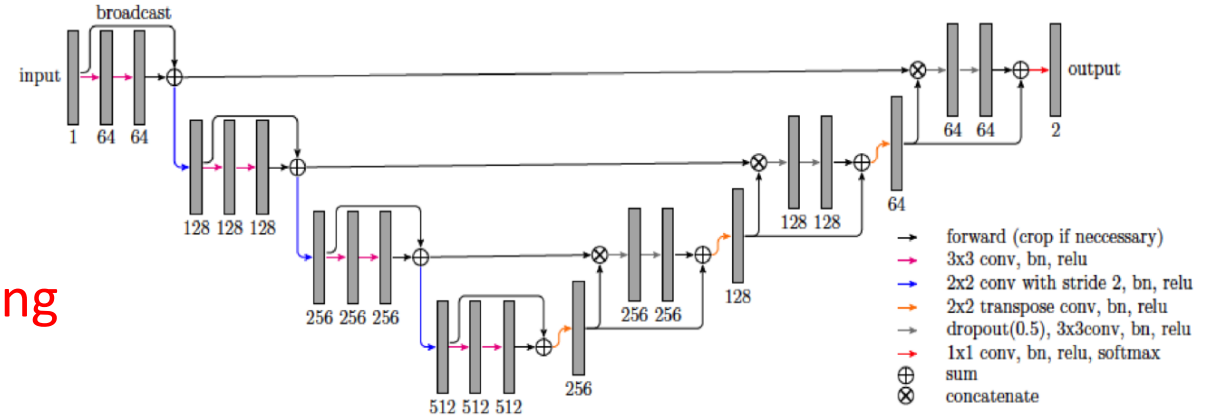
- Deep learning opportunities for theorists!
  - Works! No one understands why it works!
  - Goes against traditional thoughts of model building
    - What is a good model?

- Improve DSC accuracy to ~1.0 while simultaneously reducing variance ?
  - Cannot have both

- Have to try to get a bad result…
  - Data portal → upload new dataset → get dice >.9
    https://github.com/fuentesdt/livermask

Next steps - Physics based ML

- Spotting the matrix algebra is central to understanding the problem



- Convolution, Pooling, Upscaling are Linear Operators

$$O(m,n) = \max\{I(2m-1,2n-1), I(2m,2n-1), I(2m-1,2n), I(2m,2n)\}$$

$$= \underbrace{\begin{bmatrix} 0 & 1 & 0 & \dots \\ 1 & 0 & 0 & \dots \\ 1 & 0 & 0 & \dots \end{bmatrix}}_{P} I(:) \qquad P \in \mathbb{R}^{256^2 \times 128^2}$$

$$\hat{O}(m,n) = I\left(\left\lfloor \frac{m+1}{2} \right\rfloor, \left\lfloor \frac{n+1}{2} \right\rfloor\right) \qquad \hat{O} \in \mathbb{R}^{256 \times 256}$$

$$= \underbrace{\begin{bmatrix} 1 & 0 & 0 & \dots \\ 1 & 0 & 0 & \dots \\ 1 & 0 & 0 & \dots \\ 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ & & & \vdots \end{bmatrix}}_{U} \hat{I}(:) \qquad U \in \mathbb{R}^{128^2 \times 256^2}$$

- Product business model: Ease of use
  - Input group of images → normalization → registration → segmentation → output result is automated

- Best of both words - Mechanistic understanding of each layer with data driven accuracy

- Effectively PDE constrained optimization

$$y = W_N h_{N-1}(\dots W_3 h_2(W_2 h_1(W_1 a_0))$$

First principle physics based predictions

Non-Mechanistic Approach

Input/Output relationship learned from the data

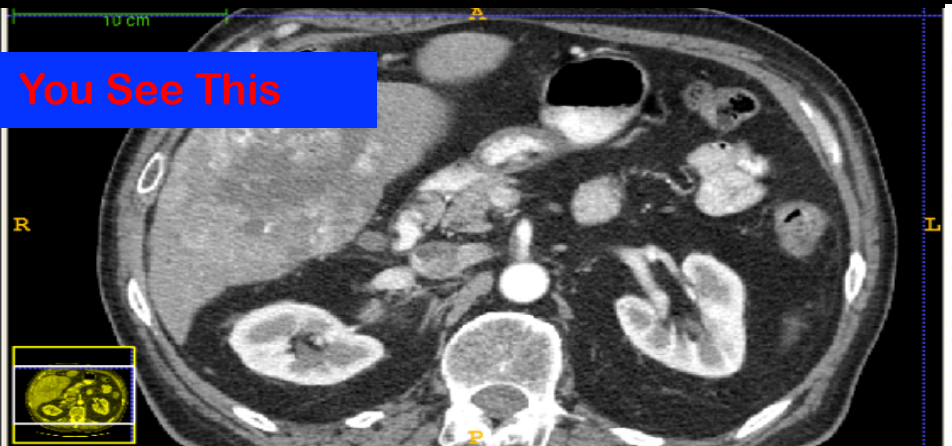
Statistical Science
2001, Vol. 16, No. 3, 199–231

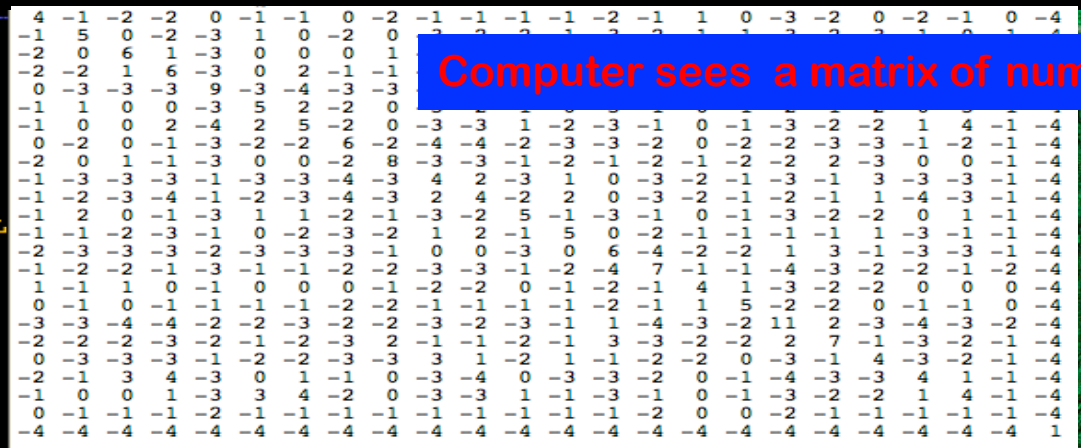
Statistical Modeling: The Two Cultures

Leo Breiman

- Conventional paradigm
  - Computer must be programmed to do something new
    - Meticulous detail
      - start from 1st principle physics, line by line instruction
- Machine Learning
  - Program something you don't know how to do yourself
  - How do you tell the difference between the liver, kidney, heart, etc
    - the liver is in the top left, has intensity threshold within a given range, simply connected, higher intensity values inside the liver are ok…. Robust?
    - Very difficult to write an analytical expression for this however ML provides a mechanism
      - train an algorithm to do a complex task by assembling a group of relatively trivial tasks
      - better than writing a single monolithic complex algorithm

- Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed
- Learning Theory: How many datasets are needed to achieve a certain prediction accuracy ?



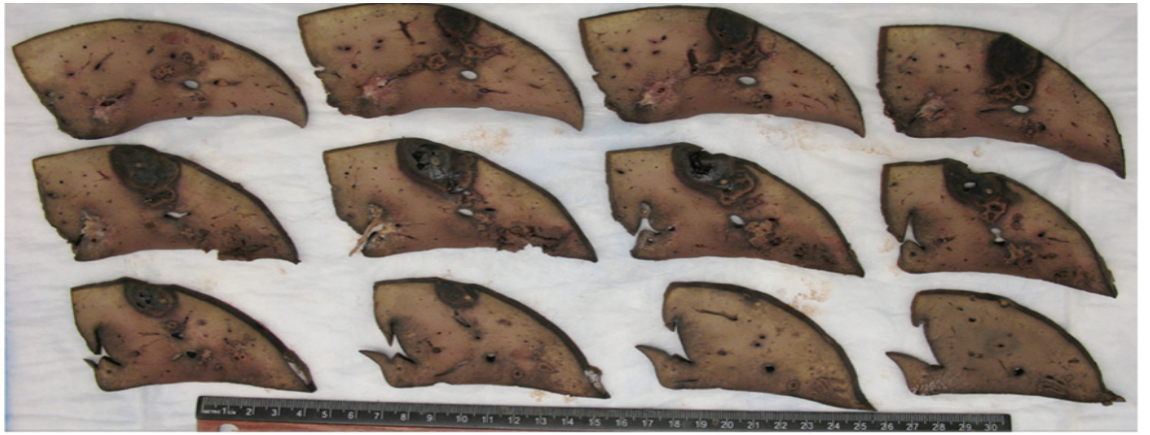


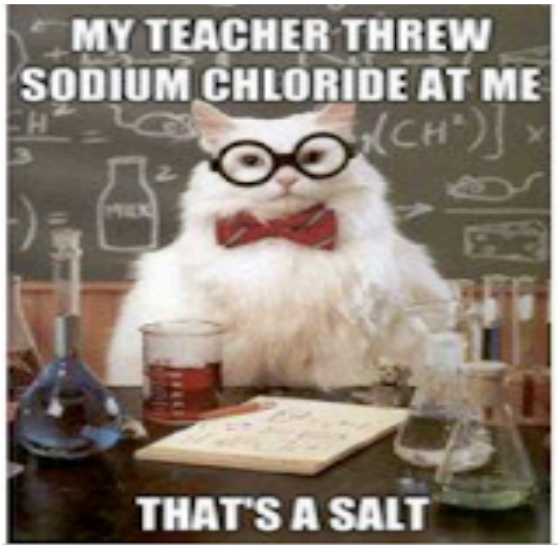
Outline

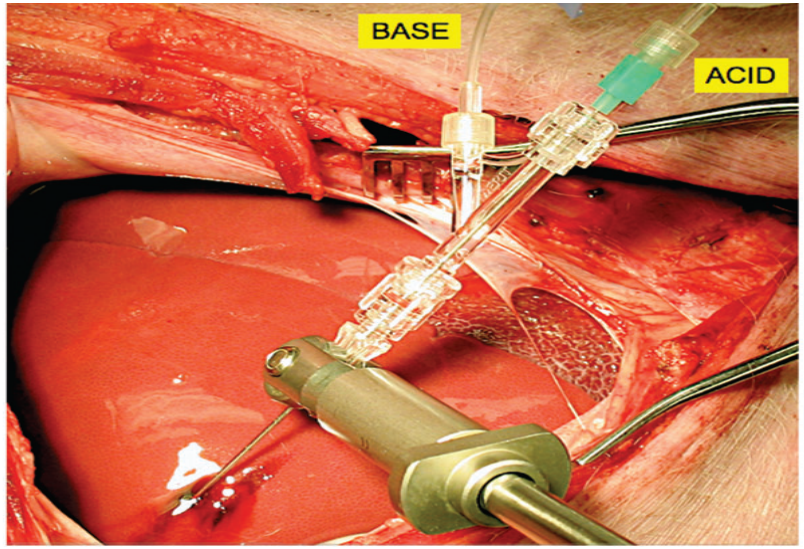
- Utilization of HPC
  - Hepatocellular Carcinoma (HCC)
    - AI
    - FEM
    - MD

Thermochemical Ablation (TCA)

- Cressman et al demonstrate efficacy of TCA for cell kill
  - coagulate a 18.9mL volume of blood perfused tissue in vivo animal models
  - sodium hydroxide + hydrochloric acid → salt + water + heat
- Motivation: Mathematical models to further study factors that affect the extent of ablation borders







$$NaOH(aq) + HCl(aq) \rightarrow NaCl(aq) + H_2O(l)$$

Governing Equations

- Heat transfer coupled to miscible flow
  - Low order FEM, stabilization, finite difference time stepping

$$S_{\text{vapor}} = 0 \qquad S_{\text{tissue}} = .8 \qquad \sum_{\alpha} S_{\alpha} = 1 \qquad v = -\frac{\kappa}{\mu}\nabla p \qquad -\nabla \cdot \left(\left((1 - S_t)\frac{\kappa}{\mu} \right) \nabla p \right) = 0$$

$$\tau = \text{rate constant} \left[\frac{1}{s} \right] \qquad \frac{\partial S_{\text{salt}}}{\partial t} + \nabla \cdot (S_{\text{salt}} v) = \tau \min \left(\frac{\rho_{\text{reactant}} S_{\text{reactant}}}{W_{\text{reactant}}}, \frac{\rho_{\text{buffer}} S_{\text{buffer}}}{W_{\text{buffer}}} \right) \frac{W_{\text{salt}}}{\rho_{\text{salt}}}$$

$$\frac{\partial S_{\text{blood}}}{\partial t} + \nabla \cdot (S_{\text{blood}} v) = \tau \min \left(\frac{\rho_{\text{reactant}} S_{\text{reactant}}}{W_{\text{reactant}}}, \frac{\rho_{\text{buffer}} S_{\text{buffer}}}{W_{\text{buffer}}} \right) \frac{W_{\text{blood}}}{\rho_{\text{blood}}}$$

$$\frac{\partial S_{\text{reactant}}}{\partial t} + \nabla \cdot (S_{\text{reactant}} v) = -\tau \min \left(\frac{\rho_{\text{reactant}} S_{\text{reactant}}}{W_{\text{reactant}}}, \frac{\rho_{\text{buffer}} S_{\text{buffer}}}{W_{\text{buffer}}} \right) \frac{W_{\text{reactant}}}{\rho_{\text{reactant}}}$$

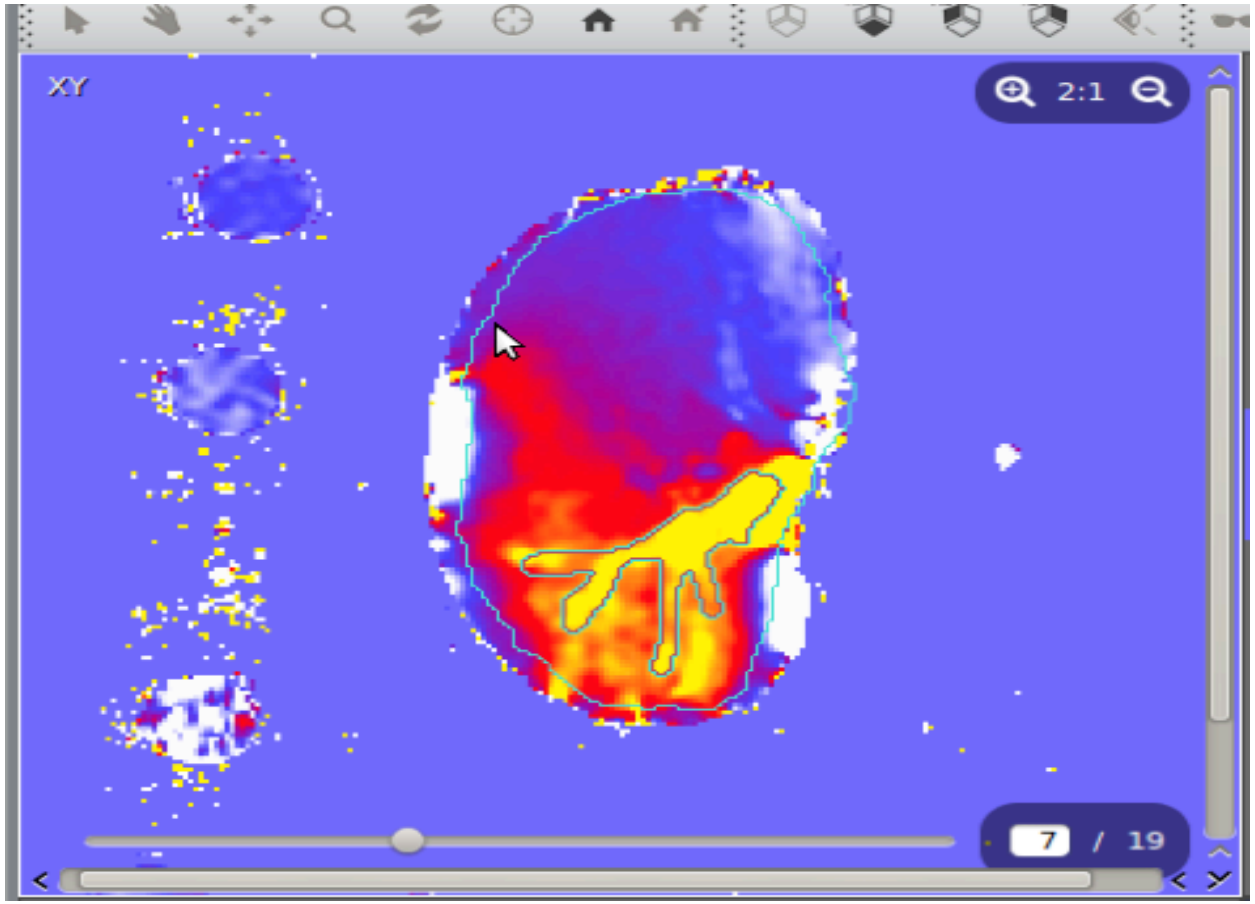
$$\frac{\partial S_{\text{buffer}}}{\partial t} + \nabla \cdot (S_{\text{buffer}} v) = -\tau \min \left(\frac{\rho_{\text{reactant}} S_{\text{reactant}}}{W_{\text{reactant}}}, \frac{\rho_{\text{buffer}} S_{\text{buffer}}}{W_{\text{buffer}}} \right) \frac{W_{\text{buffer}}}{\rho_{\text{buffer}}}$$

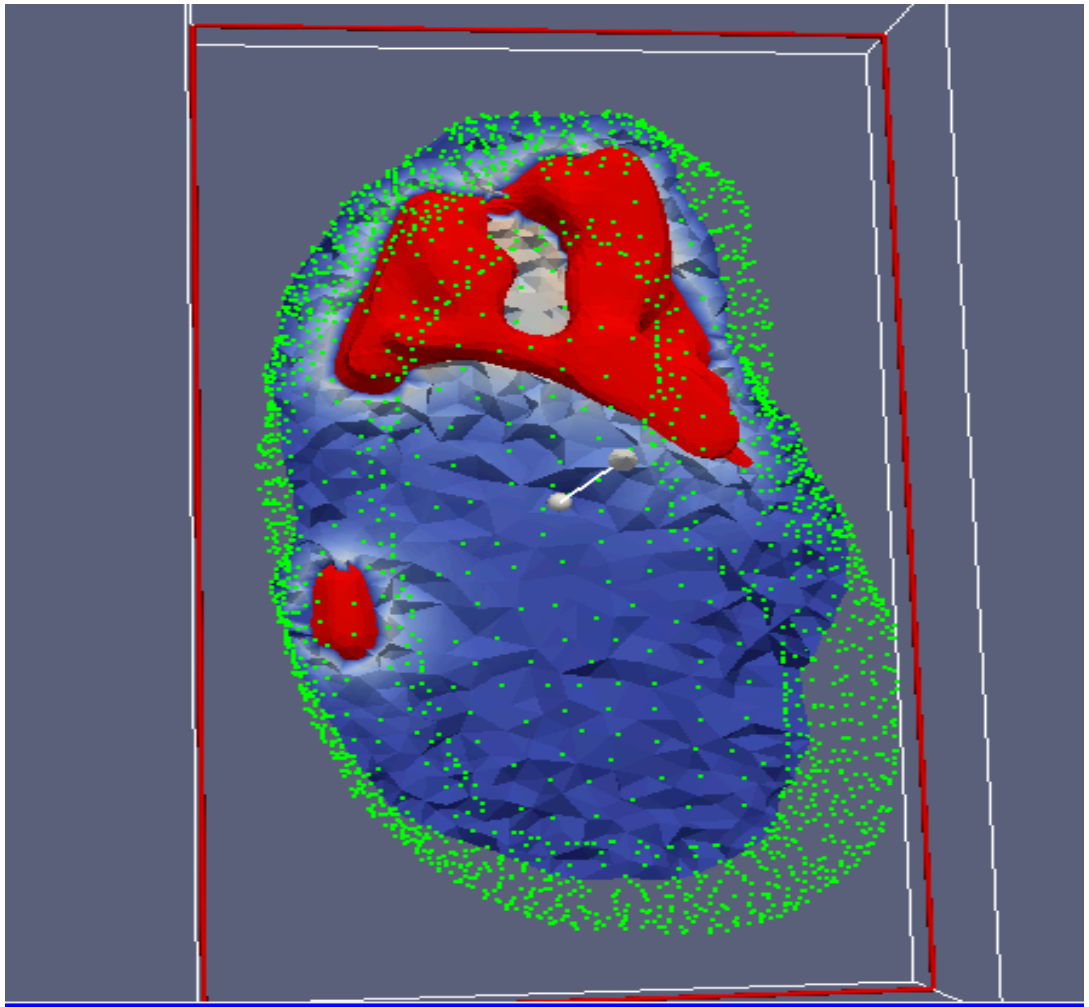
$$\left(\sum_{\alpha} (\rho c)_{\alpha} \right) \frac{\partial u}{\partial t} + S_p c_p \omega (u - u_a) + \sum_{\alpha = s,b,r} S_{\alpha} (\rho c)_{\alpha} v \nabla u = \nabla \cdot \left(\left(\sum_{\alpha} k_{\alpha} \right) \nabla u \right) + h_{\text{salt}} q_{\text{salt}}$$

$$\text{B.C./I.C.} \qquad S_s(x,t) = 1 \qquad v(x,t) = v_{\text{inject}} \qquad u(x,t) = u_{\text{inject}} \qquad x \in \partial\Omega_D$$

$$S_{\text{blood}}(x,0) = .3 \quad S_{\text{reactant}}(x,0) = 0 \quad S_{\text{buffer}}(x,0) = .1 \quad S_{\text{salt}}(x,0) = 0 \quad u(x,0) = u_0 \quad x \in \Omega$$

- Segment vessels on imaging
- Vessels provide boundary conditions for the simulation
- Tissue properties obtained from imaging



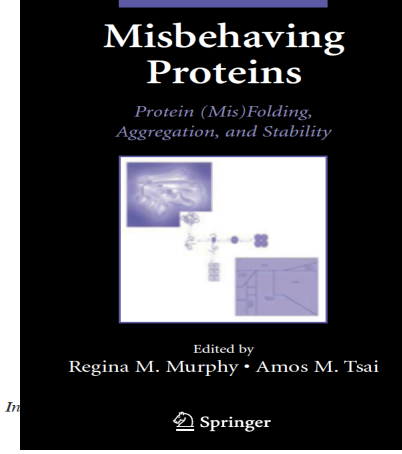


Outline

- Utilization of HPC
  - Hepatocellular Carcinoma (HCC)
    - AI
    - FEM
    - Molecular dynamics model for osmotic/thermal stress induced structural changes of proteins at the cellular scale
      - Alternative approach than usual FEM models for understanding ablation
      - Insight for characterizing fundamental thermal and osmotic damage mechanisms
        - Analogy: in vitro models invaluable system for studying complex in vivo behavior of cancer under controlled conditions
          - MD models allow detailed systematic investigation of correlations between temperature or osmolarity stress and structural changes potentially leading to tumor cell death.
      - Correlations with cell viability experiments representative of TCA environment

Insight from Molecular Models

- Literature: salt induced strengthening of hydrophobic interactions
  - With respect to the Hofmeister series, the salts used for TCA function as kosmotropes. These act to preferentially hydrate proteins and are preferentially excluded from the protein-solvent interface
  - Water concentration near interface increases burial forces for hydrophobic residues
    - MD quantifies “burial force increase” in terms of free energy
- Salt addition stabilizes the protein
  - Free energy is lowered as a function of salt concentration
  - Increased stability of nonfunctional protein conformations induced during heating may have a role in TCA induced cell stress
  - Less of an affect on unfolded states
- Molecular models allow us to further study the effect of salt and temperature as well as ionic and Van der Waals forces in ligand-receptor interactions

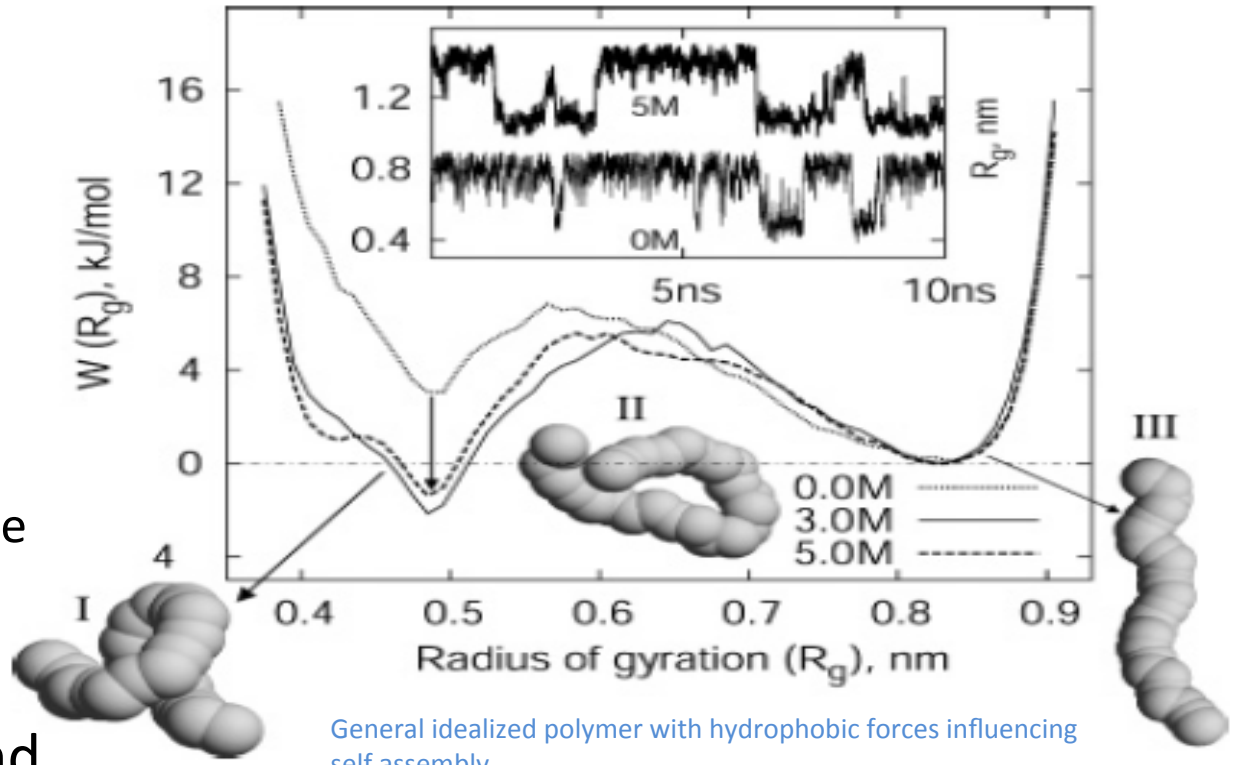


642 J. Phys. Chem. B 2005, 109, 642−651

On the Salt-Induced Stabilization of Pair and Many-body Hydrophobic Interactions

Tuhin Ghosh, Amrit Kalra,† and Shekhar Garde*

The Howard P. Isermann Department of Chemical & Biological Engineering, Rensselaer Polytechnic In
Troy, New York 12180



General idealized polymer with hydrophobic forces influencing self assembly

Extracellular Model

- Molecular models of the extracellular environment as a first step
  - Intracellular models have additional confounding factors for experimental comparison
- Idealized fibronectin/integrin molecular models of extracellular environment to further study molecular scale effects
  - Guided by availability of PDB models: 4MMX, 1FNA
- Cell attachment to ECM, including fibronectin, correlated with viability
  - arginylglycylaspartic acid (RGD) domain important for fibronectin binding to integrins on the cell surface included in these models
- Hypothesis: extracellular proteins influence survival → infer mechanisms of TCA reduced viability from observed structural changes
  - Matching experimental/simulation thermal/osmotic conditions

[CANCER RESEARCH 44, 3022–3028, July 1984]

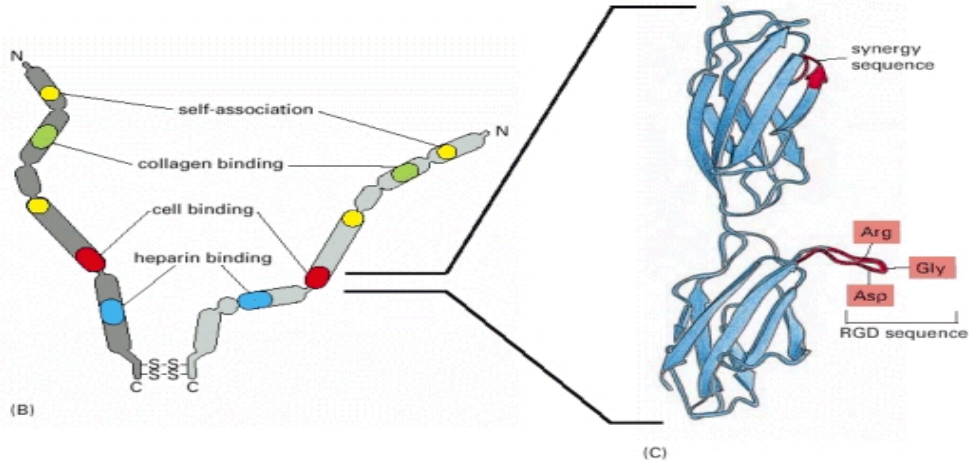
Fibronectin Synthesized by a Human Hepatoma Cell Line[1]

James E. Glasgow[2] and Robert W. Colman

Thrombosis Research Center [J. E. G., R. W. C.], and Hematology-Oncology Section of the Department of Medicine [R. W. C.], Temple University School of Medicine, Philadelphia, Pennsylvania 19140

Proteins combination on PHBV microsphere scaffold to regulate Hep3B cells activity and functionality: A model of liver tissue engineering system

Xin Hao Zhu,[1] Seng Keat Gan,[2] Chi-Hwa Wang,[1] Yen Wah Tong[1,2]
[1]*Department of Chemical and Biomolecular Engineering, National University of Singapore, 21 Lower Kent Ridge Road, Singapore 119077*
[2]*Division of Bioengineering, National University of Singapore, 21 Lower Kent Ridge Road, Singapore 119077*



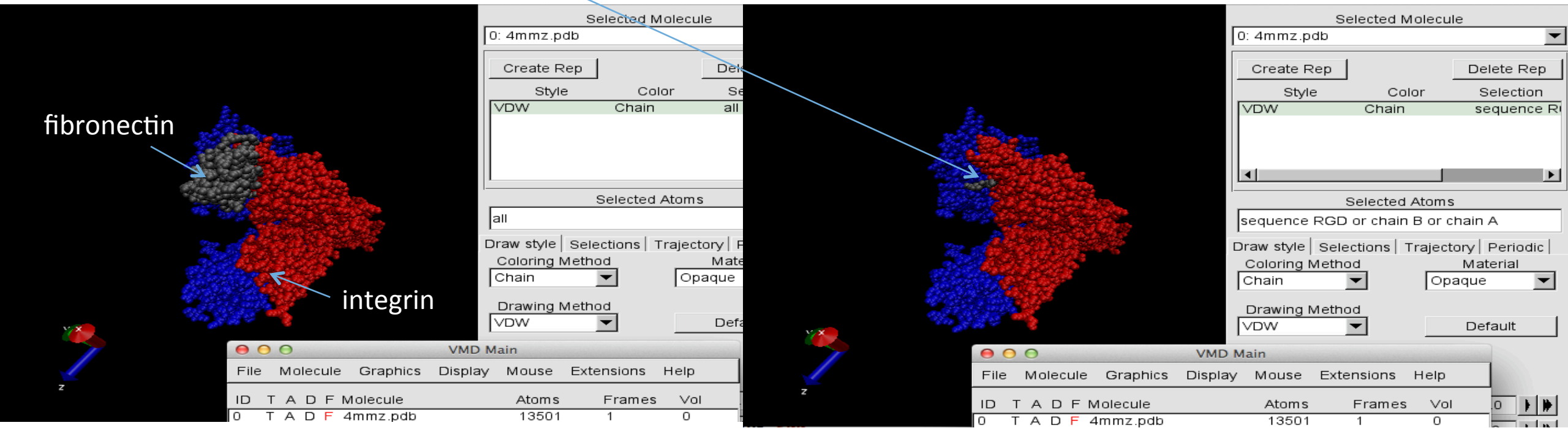
Model System

nature structural & molecular biology

Structural basis for pure antagonism of integrin $\alpha_V\beta_3$ by a high-affinity form of fibronectin

Johannes F Van Agthoven[1,4], Jian-Ping Xiong[1,4], José Luis Alonso[2], Xianliang Rui[2], Brian D Adair[1], Simon L Goodman[3] & M Amin Arnaout[1,2]

- Fibronectin bound to integrin
  - Binding domain: RGD motif



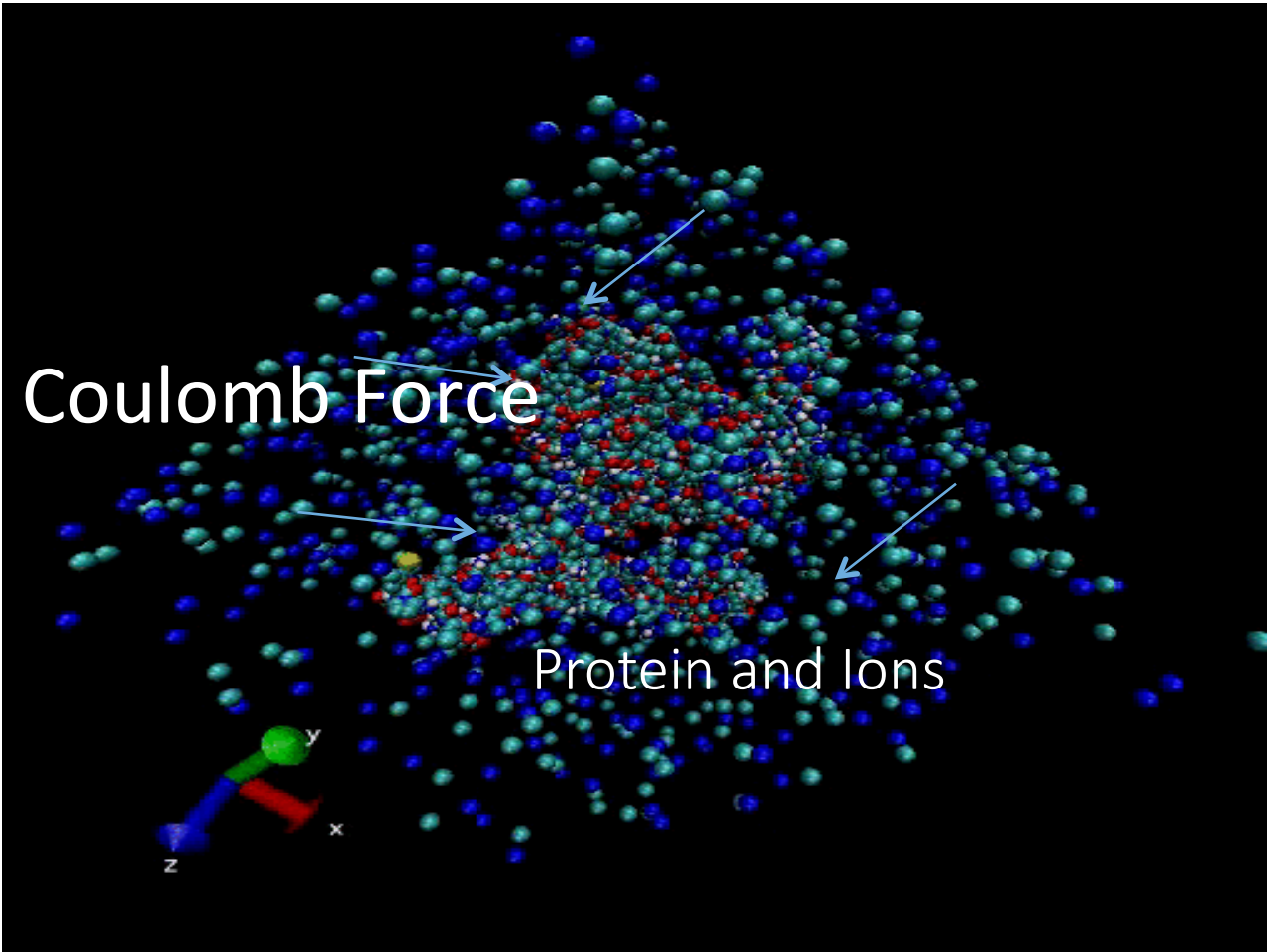
The force is given by the negative gradient of the potential energy.

$$\mathbf{F} = -\nabla V(\mathbf{r})$$

Knowing the force allows us to accelerate the atoms in the direction of the force.

$$\mathbf{F} = m\mathbf{a}$$

Simulation box ~ 10nm



General Algorithm

Need starting position, velocity, and all pairwise interactions for all 1e6, 1e9 particles in a simulation "box"

1. Input initial conditions

Potential interaction V as a function of atom positions
Positions r of all atoms in the system
Velocities v of all atoms in the system

⇓

repeat 2,3,4 for the required number of steps:

2. Compute forces

The force on any atom

$$\boldsymbol{F}_i = -\frac{\partial V}{\partial \boldsymbol{r}_i}$$

is computed by calculating the force between non-bonded atom pairs:

$$\boldsymbol{F}_i = \sum_j \boldsymbol{F}_{ij}$$

plus the forces due to bonded interactions (which may depend on 1, 2, 3, or 4 atoms), plus restraining and/or external forces. The potential and kinetic energies and the pressure tensor may be computed.

⇓

3. Update configuration

The movement of the atoms is simulated by numerically solving Newton's equations of motion

$$\frac{\mathrm{d}^2\boldsymbol{r}_i}{\mathrm{d}t^2} = \frac{\boldsymbol{F}_i}{m_i}$$

or

$$\frac{\mathrm{d}\boldsymbol{r}_i}{\mathrm{d}t} = \boldsymbol{v}_i; \quad \frac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t} = \frac{\boldsymbol{F}_i}{m_i}$$

⇓

4. if required: **Output step**

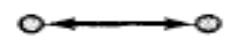
write positions, velocities, energies, temperature, pressure, etc.

Bond Potentials

Born-Oppenheimer approximation: motion of atomic nuclei and electrons decoupled

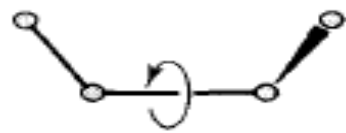
$$\mathcal{V}(\mathbf{r}^N) = \sum_{\text{bonds}} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{\text{angles}} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{\text{torsions}} \frac{V_n}{2}(1 + \cos(n\omega - \gamma))$$
$$+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right)$$

Parameters calibrated to thermodynamic properties in idealized experimental scenarios and using quantum mechanical simulations



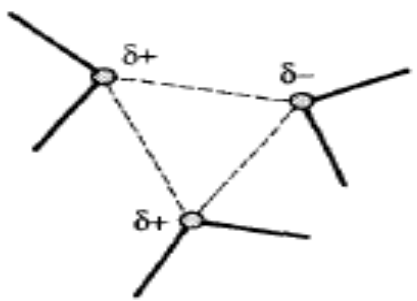
Bond stretching

Angle bending

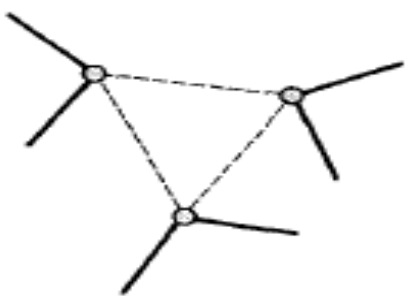


Bond rotation (torsion)

Bond	l_0 (Å)	k (kcal mol^{-1} Å^{-2})
Csp^3-Csp^3	1.523	317
Csp^3-Csp^2	1.497	317
$Csp^2=Csp^2$	1.337	690
$Csp^2=O$	1.208	777
Csp^3-Nsp^3	1.438	367
C–N (amide)	1.345	719



Non-bonded interactions (electrostatic)



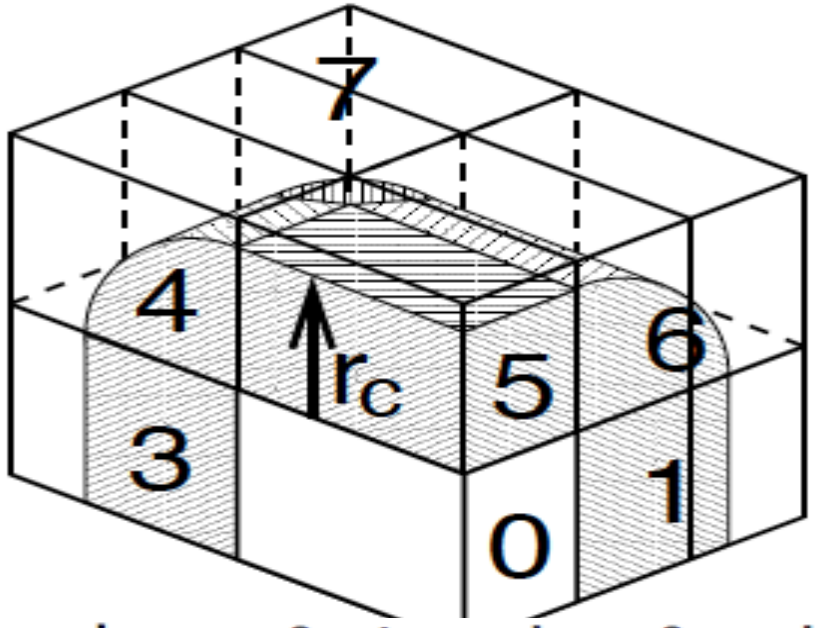
Non-bonded interactions (van der Waals)

Domain Decomposition for Parallelism

```
500000000 steps, 1000000.0 ps.
step 900, will finish Thu Aug 16 09:52:22 2018imb F 11% pme/F 0.56
NOTE: Turning on dynamic load balancing

step 1400, will finish Mon Aug 13 02:02:26 2018vol 0.83  imb F 12% pme/F 0.72
```

Local Machine Linux Workstation ~ 9months

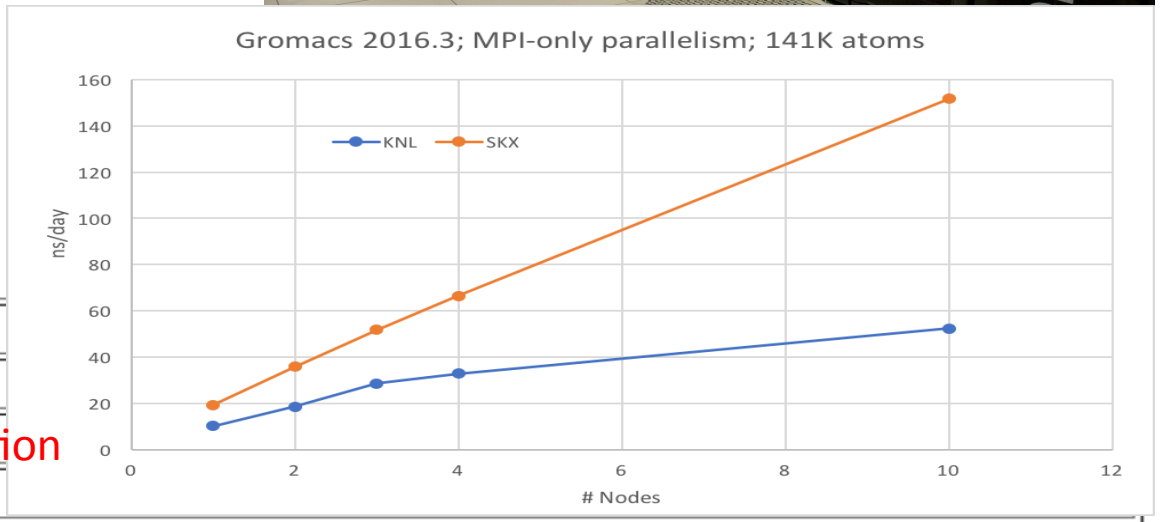



```
Tasks: 468 total,   2 running, 466 sleeping,   0 stopped,   0 zombie
%Cpu(s): 75.5 us, 24.5 sy,  0.0 ni,  0.0 id,  0.0 wa,  0.0 hi,  0.0 si,  0.0 st
KiB Mem : 19808771+total, 12982587+free,  6989848 used, 61271984 buff/cache
KiB Swap: 97654784 total, 97654784 free,        0 used. 18828867+avail Mem
  scroll coordinates: y = 1/468 (tasks), x = 1/12 (fields)
  PID USER      PR  NI    VIRT    RES    SHR S  %CPU %MEM     TIME+ COMMAND
23603 fuentes   20   0 1776356 118612  12052 R  2370  0.1   7:36.03 gmx
40915 fuentes   20   0 1395232 369232  59264 S   8.3  0.2   2488:13 nautilus
```

Flop Time << memory transfer << message passing (MPI)

Stampede2 > 200000 cores

- Simulation time of 1 microsecond
  - months (xeon, 12 cores, 2.4 GHz, local) → weeks (knights landing, 256 cores, 1.2 GHz) → 4 days (skylake, 96 cores, 2.1 GHz, vectorized instruction set)



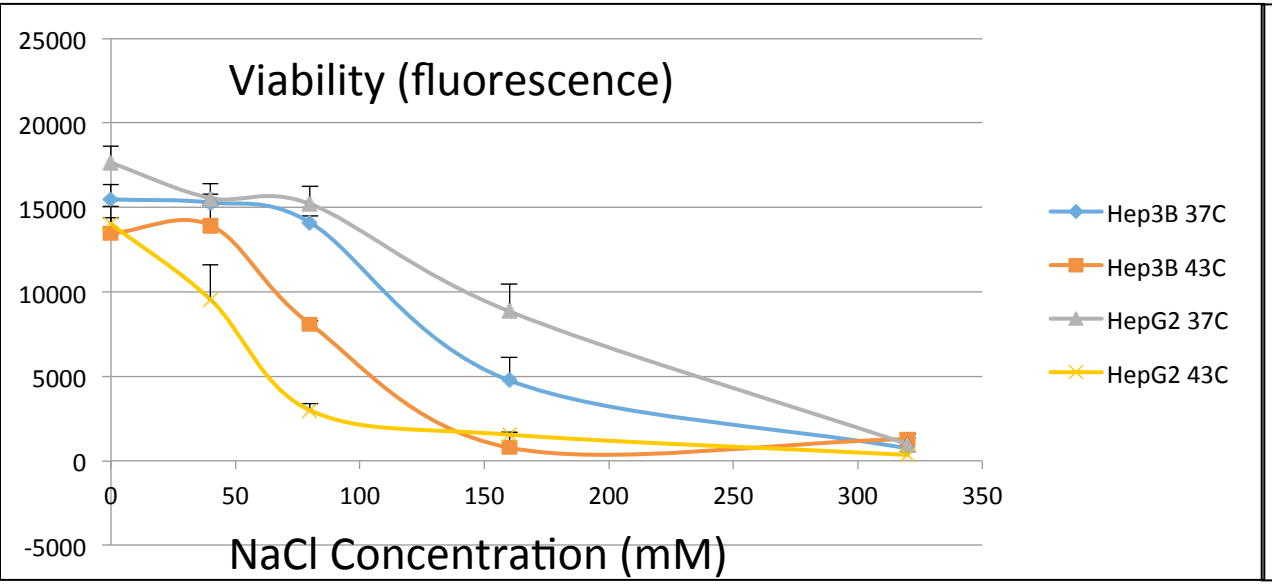
Model:	Intel Xeon Platinum 8160 ("Skylake")
Total cores per SKX node:	48 cores on two sockets (24 cores/socket)
Hardware threads per core:	2
Hardware threads per node:	48 x 2 = 96
Clock rate:	2.1GHz nominal (1.4-3.7GHz depending on instruction set and number of active cores)
RAM:	192GB (2.67GHz)
Cache:	32KB L1 data cache per core; 1MB L2 per core; 33MB L3 per socket. Each socket can cache up to 57MB (sum of L2 and L3 capacity).
Local storage:	144GB `/tmp` partition on a 200GB SSD. Size of `/tmp` partition as of 14 Nov 2017.

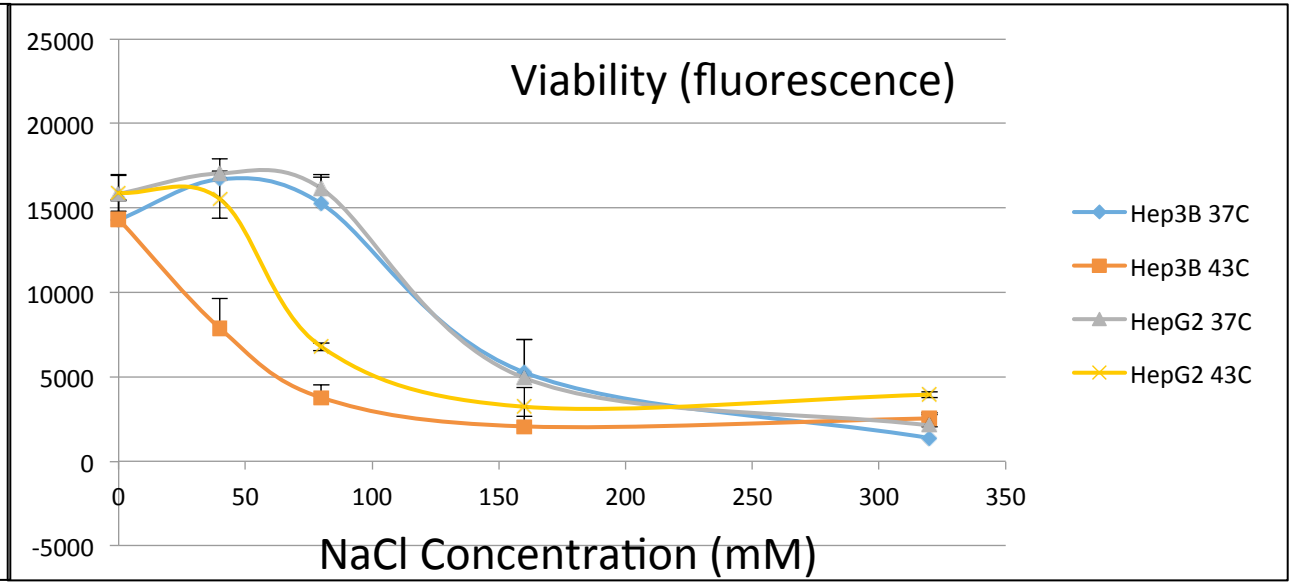
More sockets → less communication bottle neck

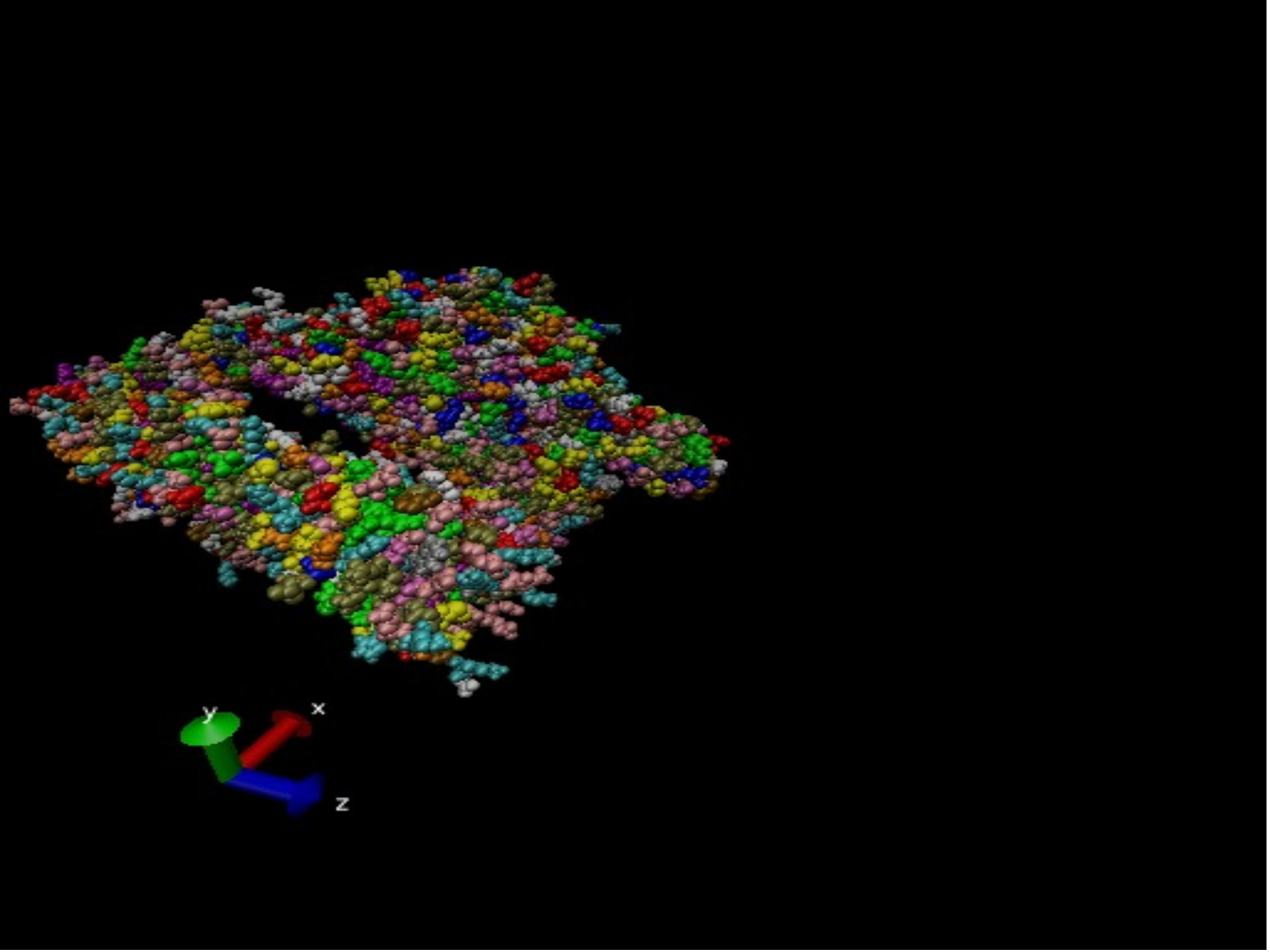
-DCMAKE_C_FLAGS="-std=gnu99 -O3 -xCORE-AVX2 -axMIC-AVX512,CORE-AVX512 -mkl=sequential -g ”
-DCMAKE_CXX_FLAGS="-std=c++11 -O3 -xCORE-AVX2 -axMIC-AVX512,CORE-AVX512 -mkl=sequential -g ”

Experimental Setup

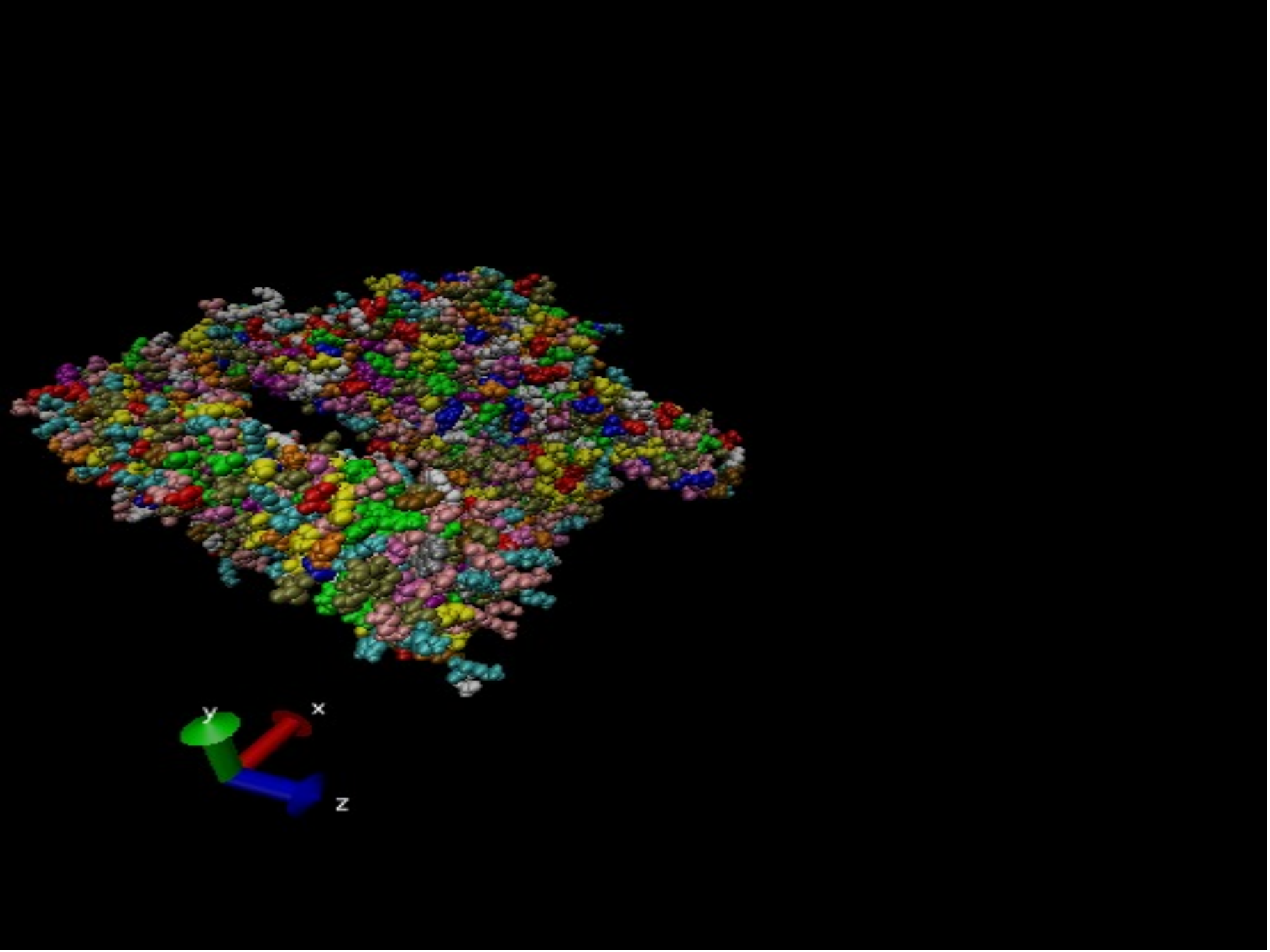
- Human HCC lines HepG2 and Hep3B were subject to combined hyperthermal stress (43C for 1hr) and hyperosmotic stress (24h) with the Sodium salts in the Hofmeister series below, at concentrations of 0, 40, 80, 160 and 320mM:
- After 1h at 43C the cells were returned to 37C.
- After 24h of treatment initiation, the salts were removed from the cultures, and cell viability was measured using AlamarBlue® 3, 24 and 48h after removal of salts. *PLOTS CORRESPOND TO 48h TIMEPOINT.*



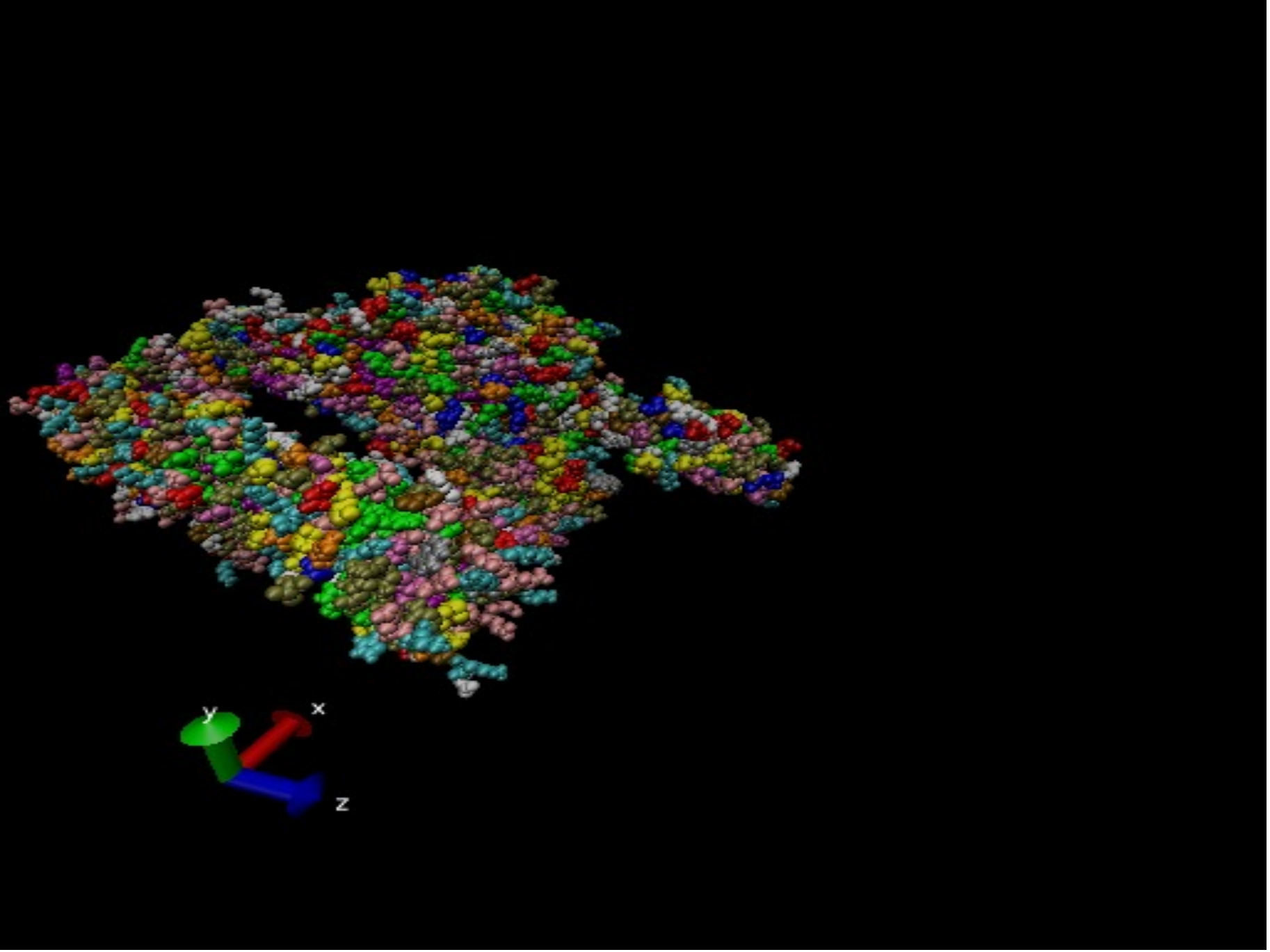




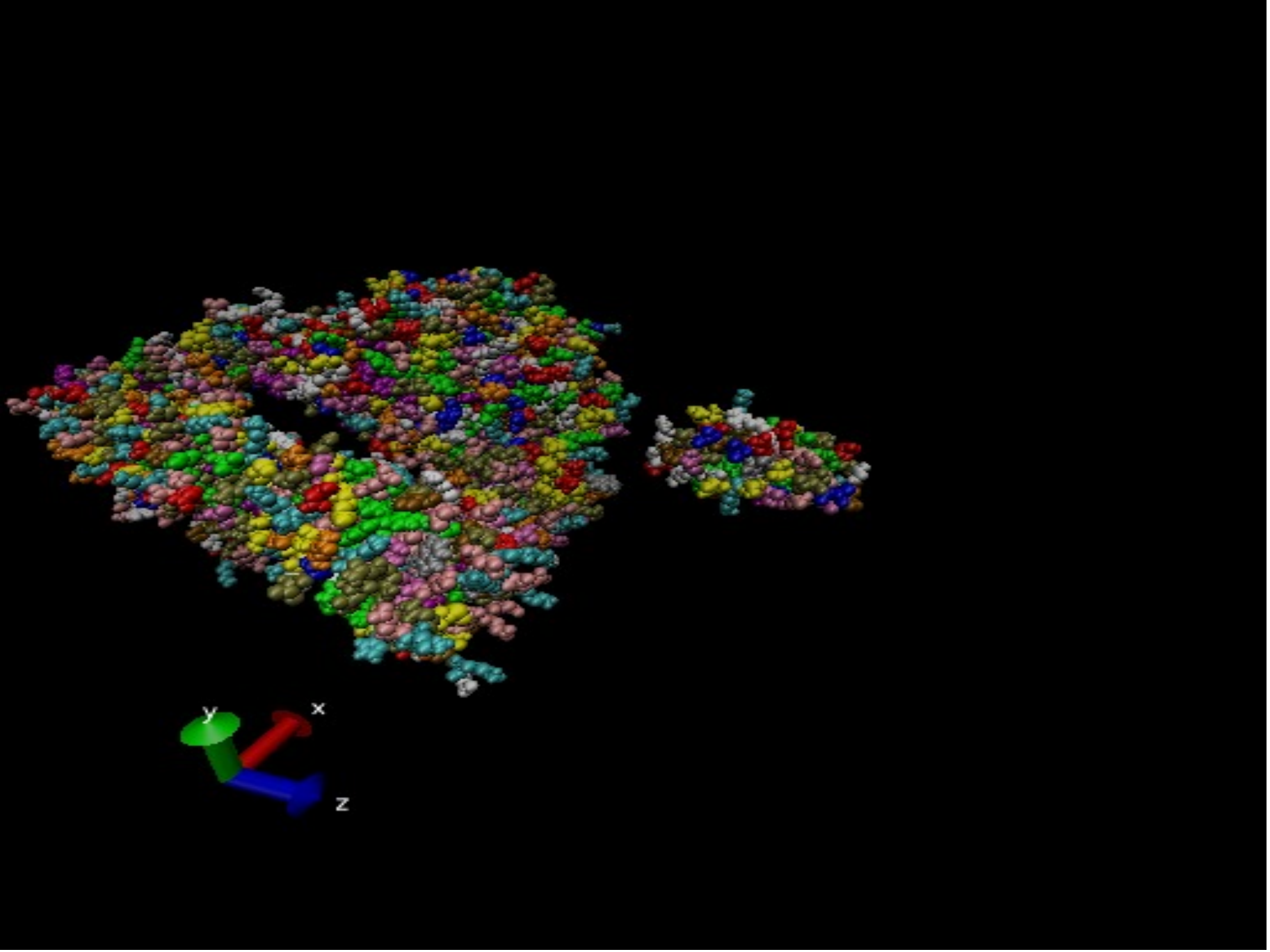
Linear Pull



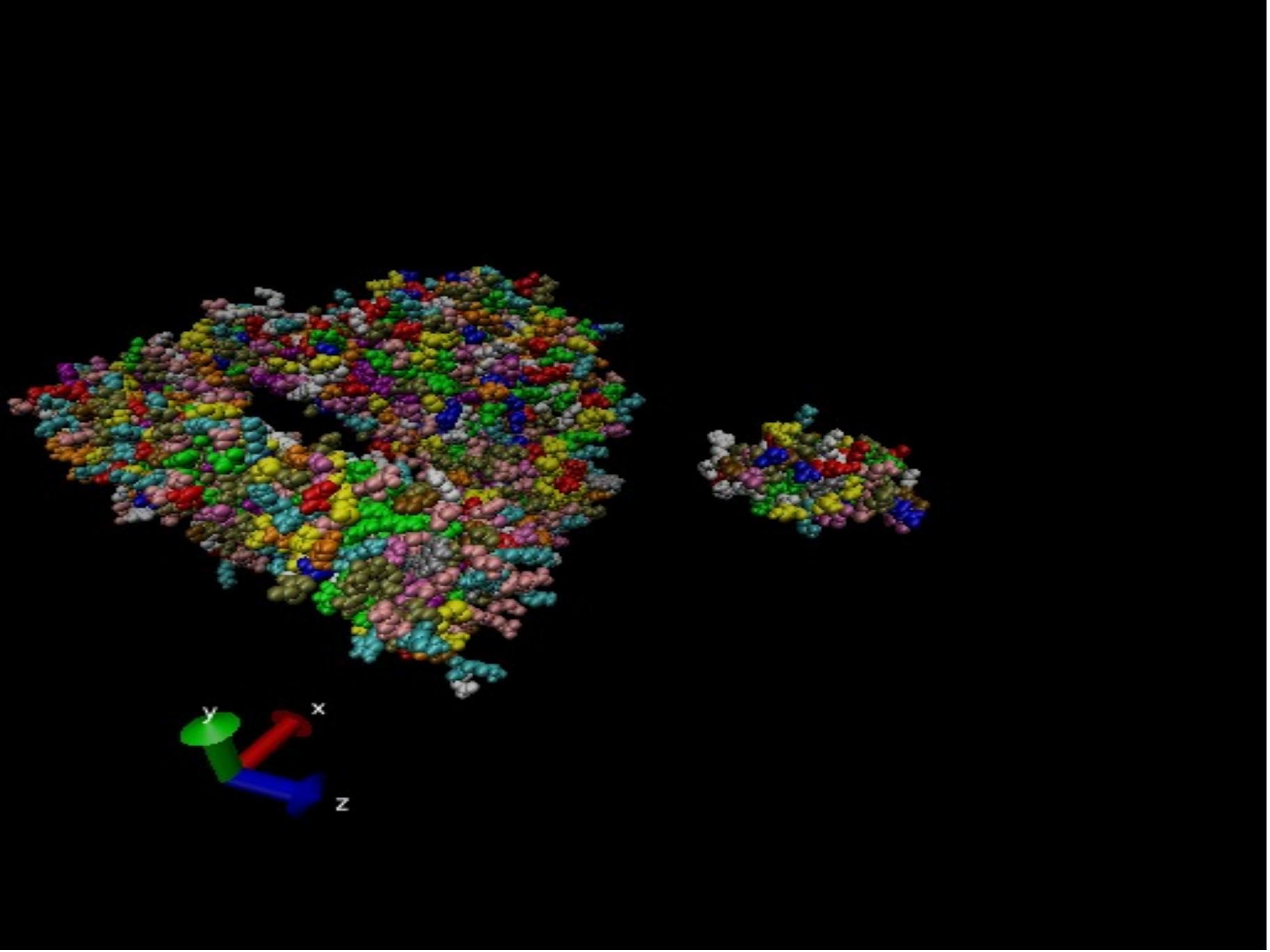
Linear Pull



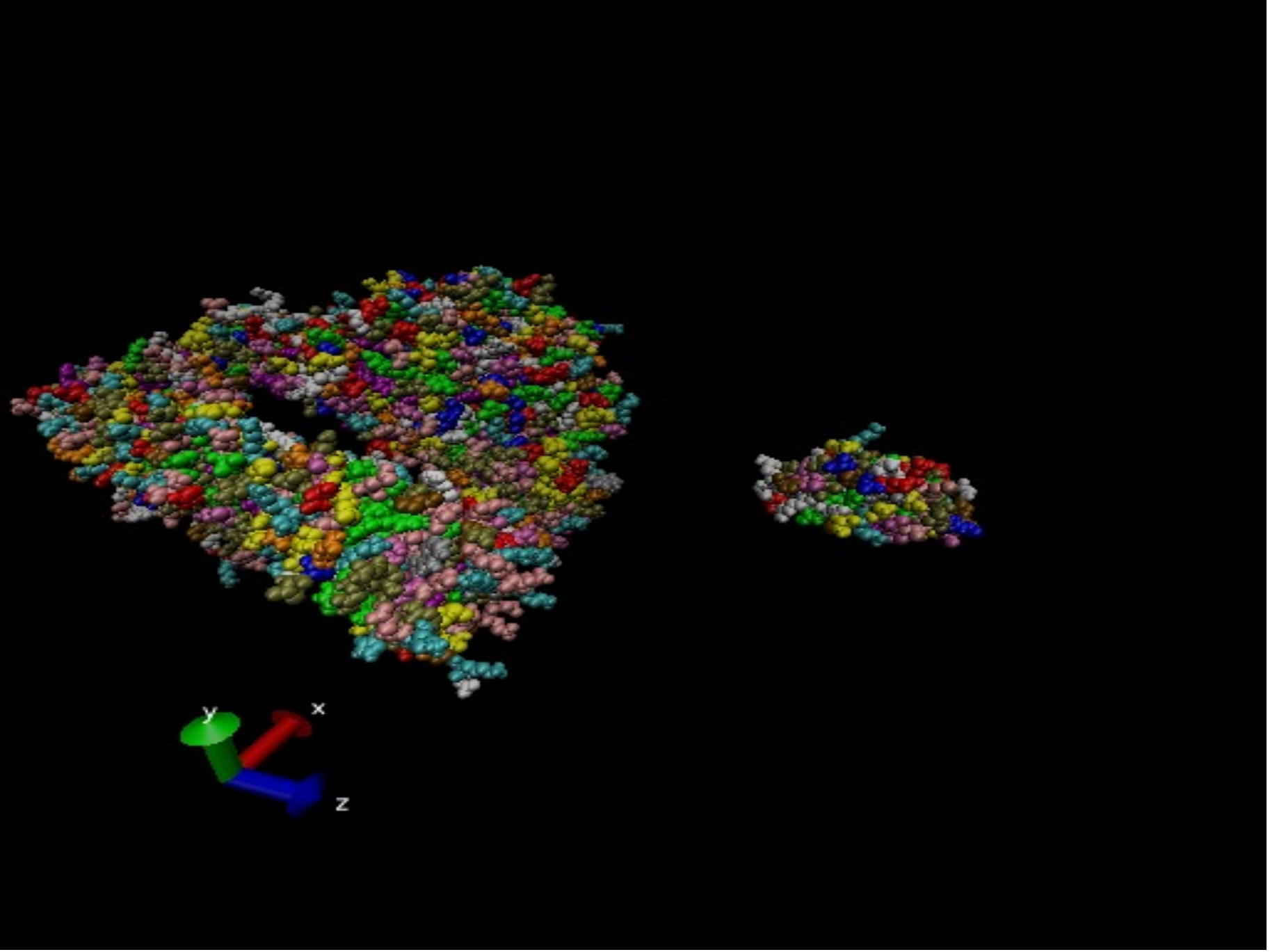
Linear Pull



Linear Pull

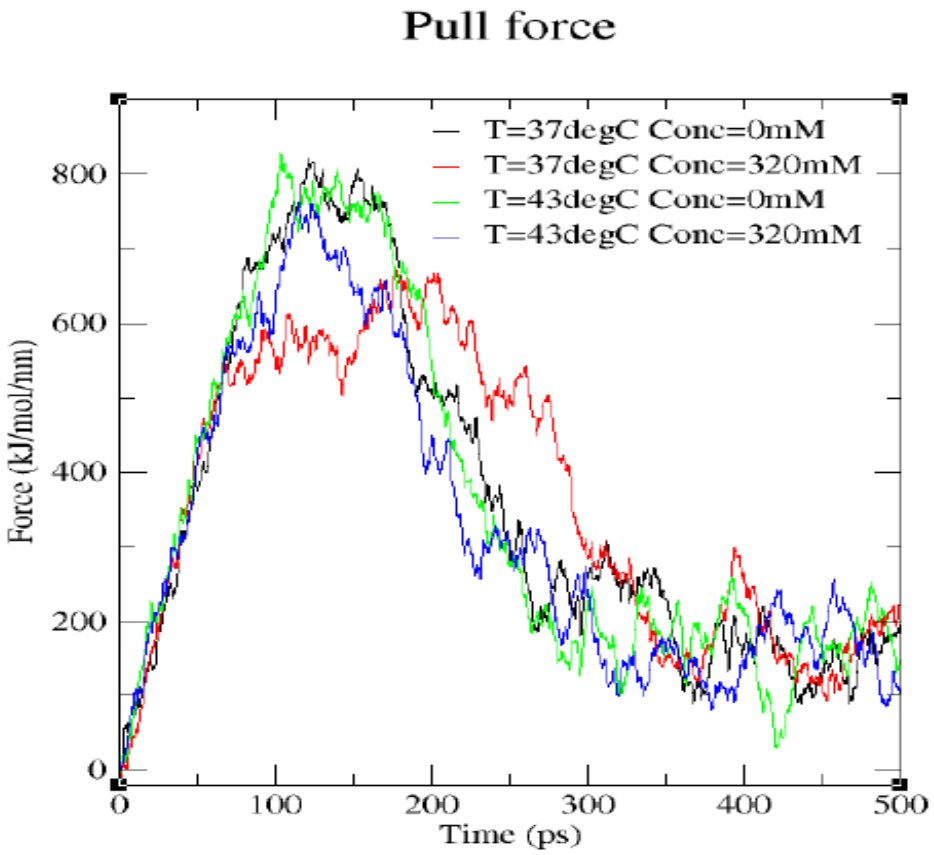


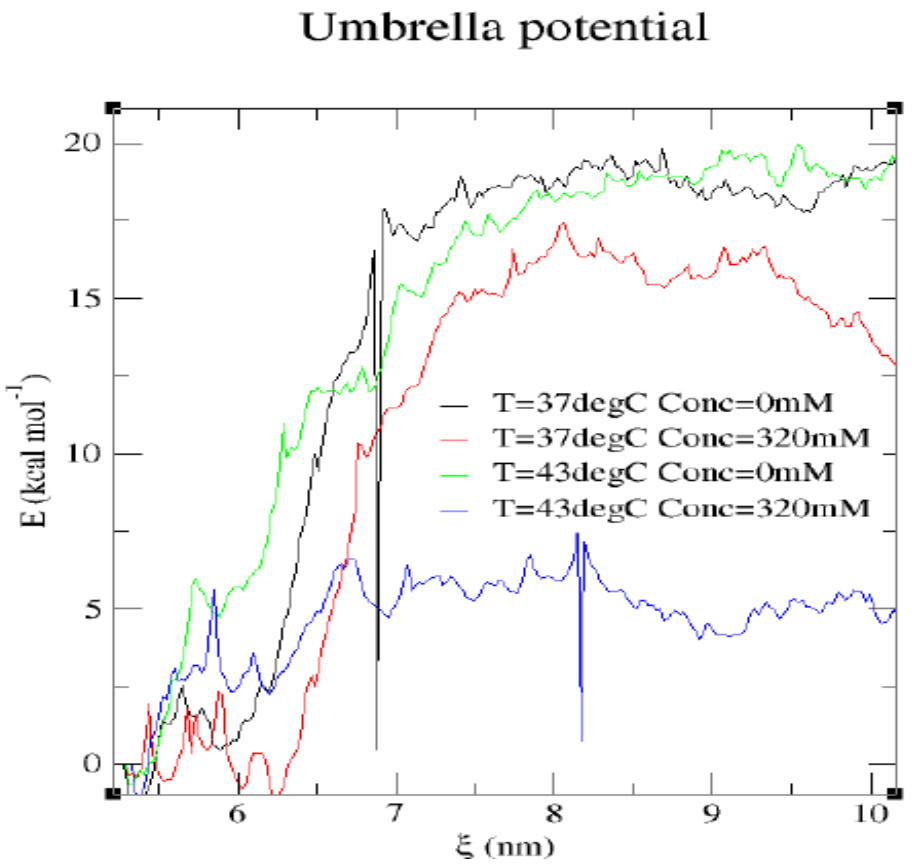
Linear Pull



Linear Pull

Synergy of Salt and Temperature





At the largest hyperthermal and hyperosmotic stress, fibronectin binding to integrin was energetically less favorable. Correlations with viability again suggest that less favorable cell attachment conditions contribute to reduce cellular viability at the TCA ablation boundaries.

Intuitively forces opposing the system tendency to drift to lower free energies are related

Summary

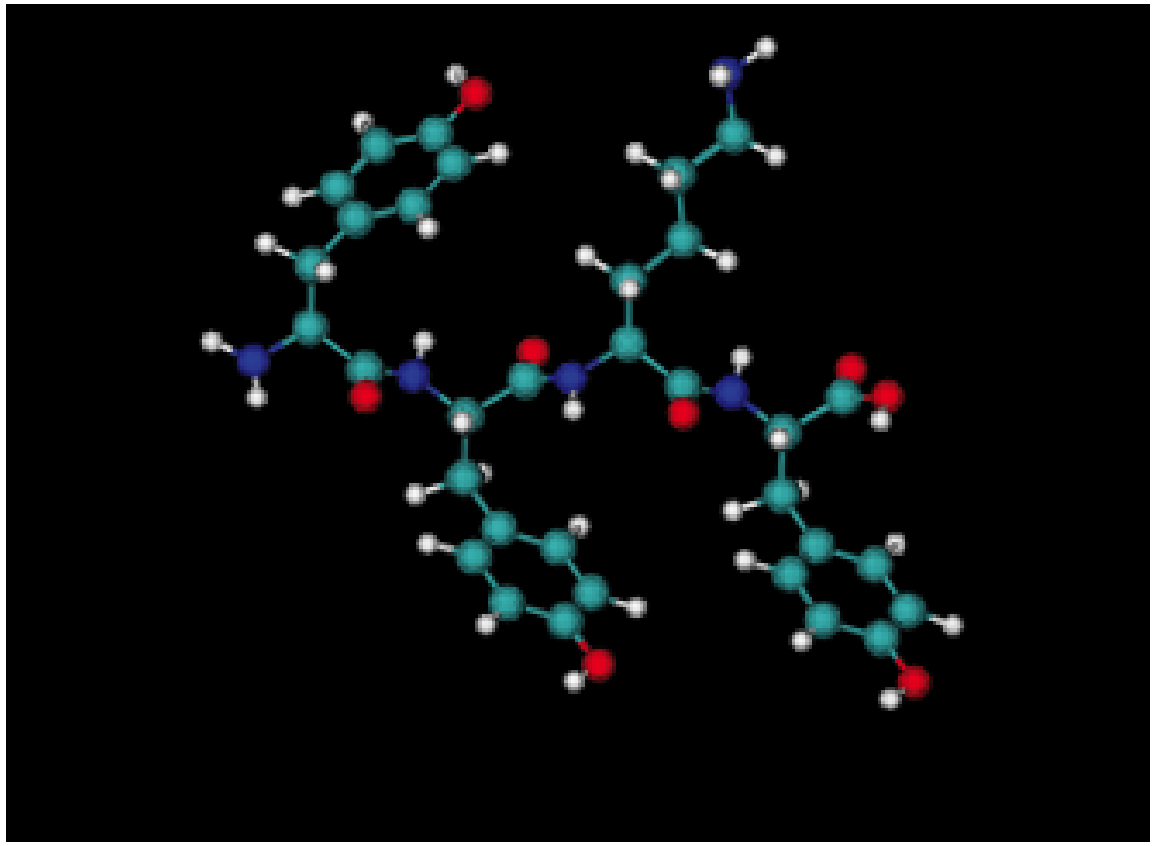
- Insight for further experimental design

- MD simulations mainly correlated with salt concentration → salt concentration correlated with survival != MD simulations correlated with survival
  - Need to repeat viability experiments to confirm influence of TCA blocking of cell attachment on viability

- Limited by PDB models. Fibronectin/Integrin chosen because PDB file available
  - Future work, upstream/downstream intracellular HSP targets ?

- Would expect any model to show increase `hydrophobic hiding' in salt environment. As well as salt concentration to decrease binding energy from Van der Waals Forces
  - Integrin inhibitor drug trials failed 10yrs ago… need more modern receptor-ligand system

Molecular dynamics investigation of self-assembly of peptide-containing nanostructures



Tetra peptide – TYR-TYR-LYS-TYR

dtfuentes@mdanderson.org

https://github.com/ImageGuidedTherapyLab