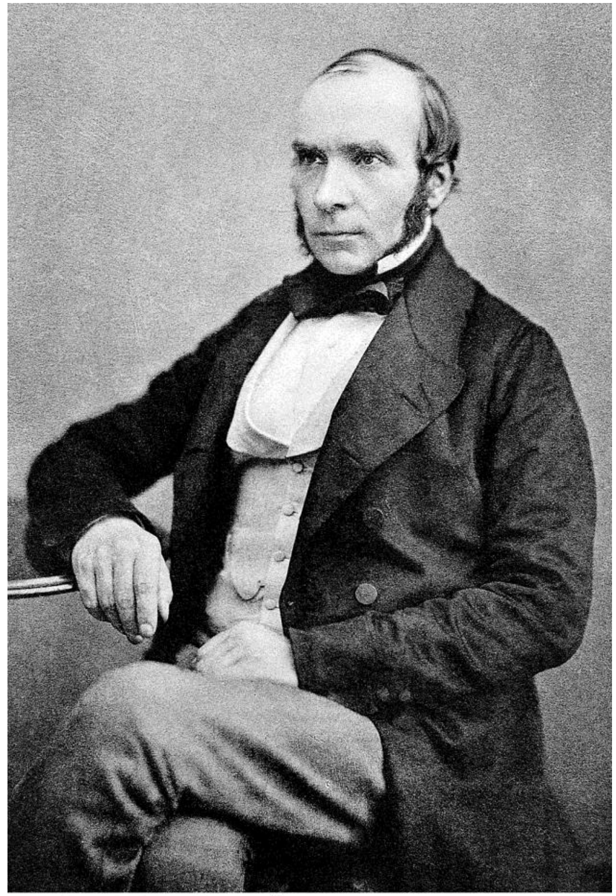
Towards Big Data in Radiation Oncology

Dr. rer. nat. Klaus H. Maier-Hein (né Fritzsche)
Head of junior research group Medical Image Computing

GERMAN
CANCER RESEARCH CENTER
IN THE HELMHOLTZ ASSOCIATION

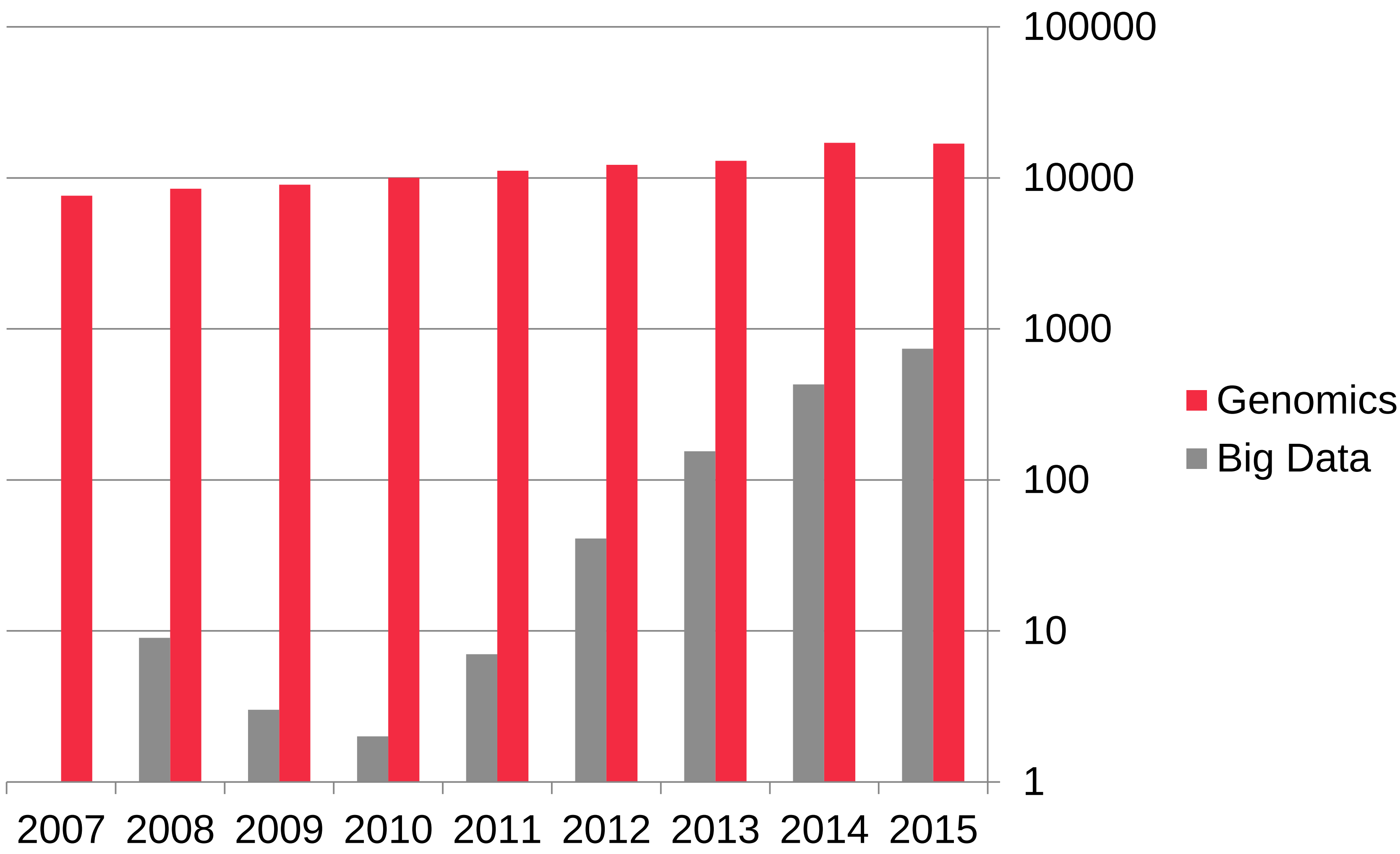


John Snow

The rise of big clinical databases

- Medicare Provider Analysis (MedPAR)
  - Diagnosic/procedure information, limited patient characteristics
- Swedish Colon Cancer Registry
  - Clinical factors, outcomes, co-morbidities
- The Cancer Genome Atlas (TCGA) and The Cancer Imaging Archive (TCIA)
  - Genomic and imaging data on a variety of cancers
  - Official repository for Nature Publishing Group

„Big data“ in Pubmed



Big data

- **Volume:** the quantity of generated and stored data
- **Variety:** the type and nature of the data
- **Velocity:** the speed at which the data is generated and processed
- **Variability:** the inconsistency in the data set can hamper processing
- **Veracity:** the varying quality of captured data

Hilbert, M. (2015). Digital Technology and Social Change [Open Online Course at the University of California] https://canvas.instructure.com/courses/949415

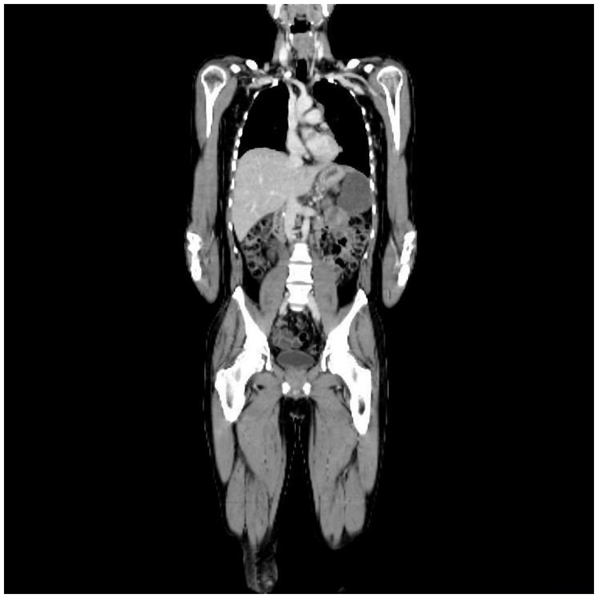
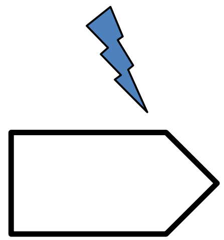


Image data
(many gigabytes)

Big Data
useless!



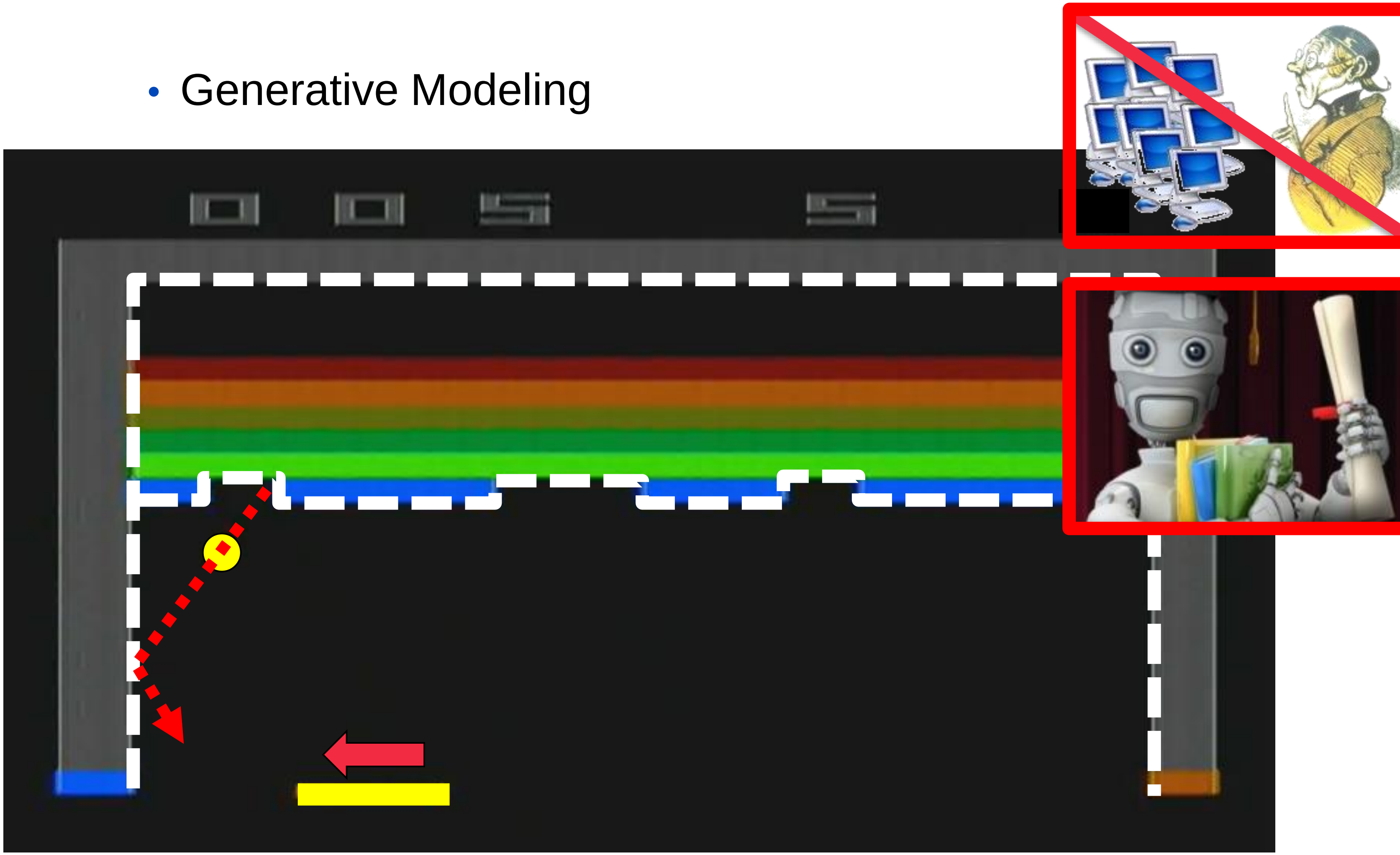
MRT Institut
am Friedrichshain

Medical report
(few kilobytes)

Atari 2600 Game „Breakout“

Previous Approach

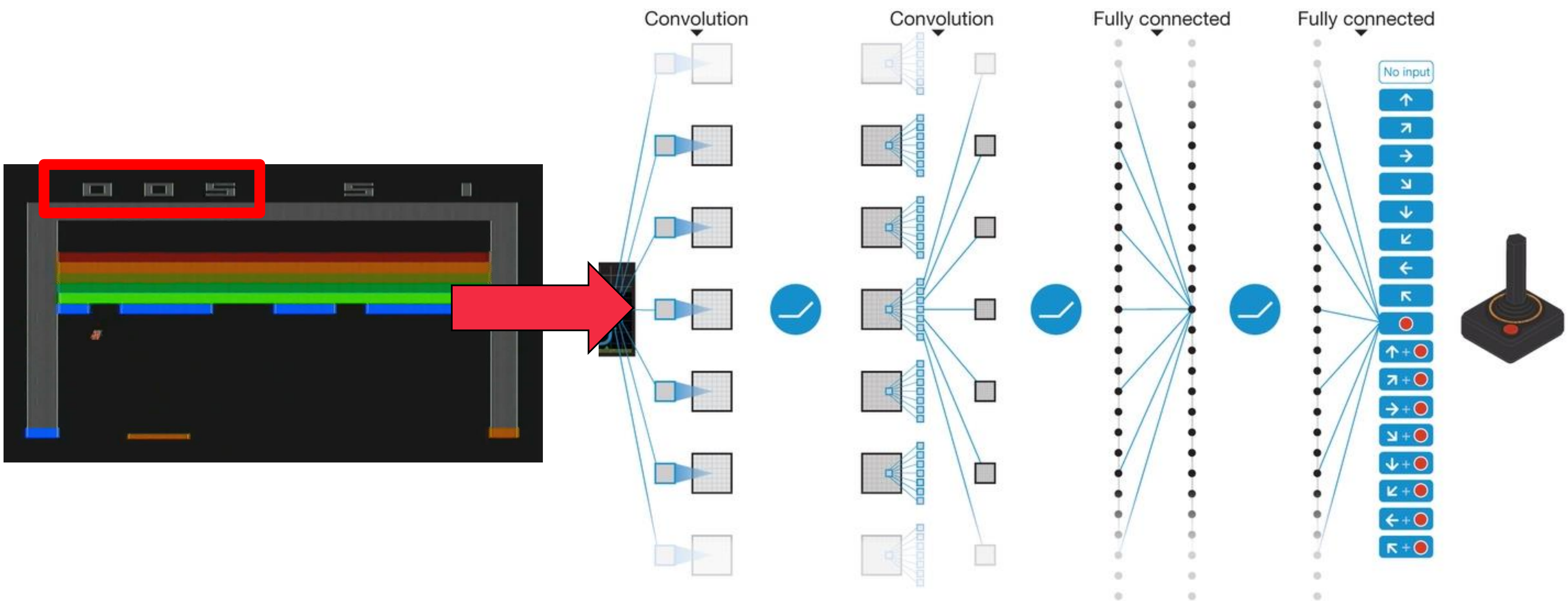
- Generative Modeling



Data-driven approach (discriminative)

- Screen
(40.000 pixels)
- Score

Action



Mnih *et al.*, Human-level control through deep reinforcement learning. Nature 518, 529–533 (2015)

Results

10 min training

Beginner level

120 min training

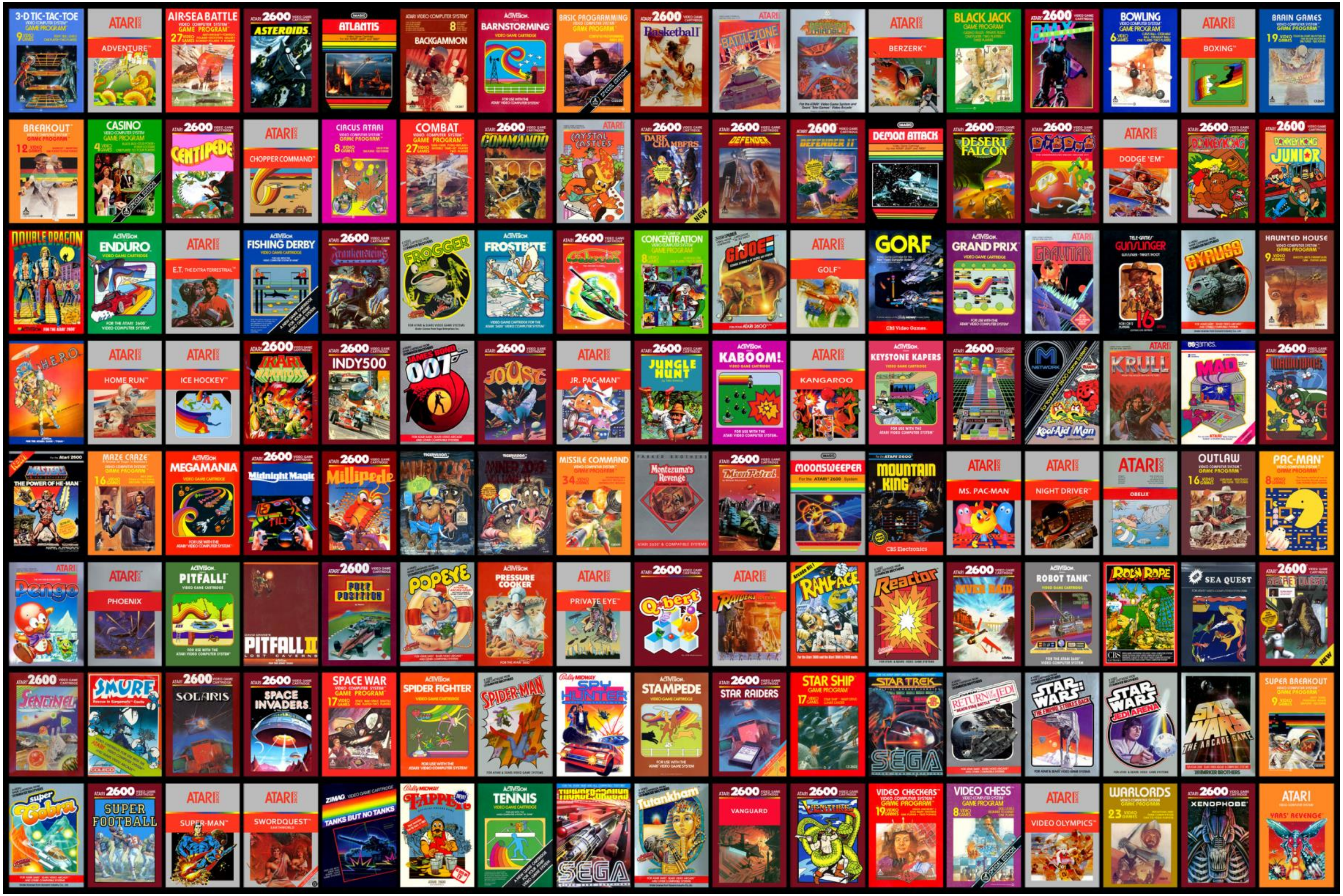
Expert level

240 min training

13x better than human

Mnih *et al.*, Human-level control through deep reinforcement learning. Nature 518, 529–533 (2015)

Results

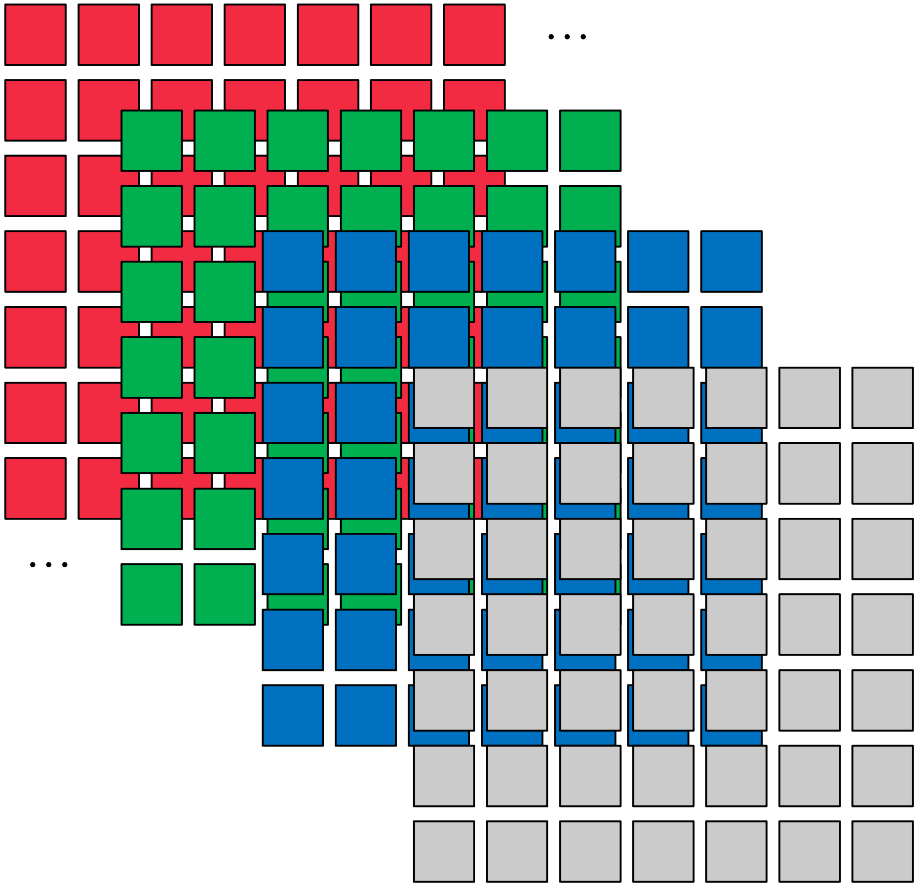

3-D TIC-TAC-TOE
ADVENTURE
AIR-SEA BATTLE
ASTEROIDS
ATLANTIS
BACKGAMMON
BARNSTORMING
BASIC PROGRAMMING
Basketball
BERZERK
BLACK JACK
BOWLING
BOXING
BRAIN GAMES
BREAKOUT
CASINO
CHOPPER COMMAND
CIRCUS ATARI
COMBAT
DEMON ATTACK
DODGE 'EM
ENDURO
E.T. THE EXTRA-TERRESTRIAL
FISHING DERBY
FROSTBITE
CONCENTRATION
GOLF
GORF
GRAND PRIX
HAUNTED HOUSE
HOME RUN
ICE HOCKEY
INDY500
007
JR. PAC-MAN
KABOOM!
KANGAROO
KEYSTONE KAPERS
KRULL
THE POWER OF HE-MAN
MAZE CRAZE
MEGAMANIA
MISSILE COMMAND
Montezuma's Revenge
MOONSWEEPER
MOUNTAIN KING
MS. PAC-MAN
NIGHT DRIVER
OUTLAW
PAC-MAN
PHOENIX
PITFALL!
PITFALL II
PRESSURE COOKER
PRIVATE EYE
ROBOT TANK
SEA QUEST
SOLARIS
SPACE INVADERS
SPACE WAR
SPIDER FIGHTER
STAMPEDE
STAR RAIDERS
STAR SHIP
SEGA
SUPER BREAKOUT
SUPER FOOTBALL
SUPER-MAN
SWORDQUEST
TANKS BUT NO TANKS
TENNIS
Tutankham
VANGUARD
VIDEO CHECKERS
VIDEO CHESS
VIDEO OLYMPICS
WARLORDS
XENOPHOBE
YARS' REVENGE

What does all this mean for us?

What does all this mean for us?

- Computers are great in learning from images

I = 84*84 *4 = 28,224 numbers



What does all this mean for us?

- Computers are great in learning from images

28,224

Image 1

Image 2

Image 3

…

…

Data matrix

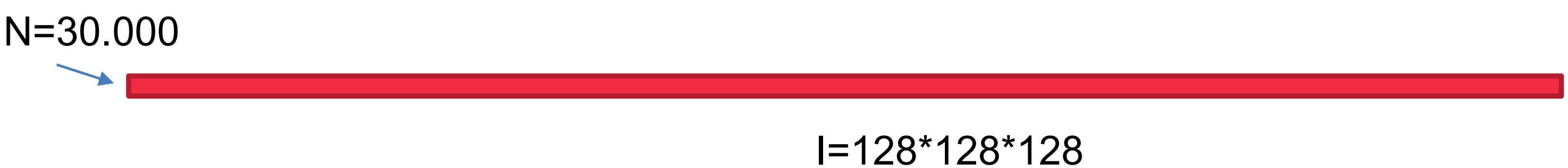
N=50.000.000
(1585 slides)

What about medical imaging?

- Computers are great in learning from images

BUT

- Data matrices are wiiiddeeeee ….



- Meaningful data annotation difficult to get

- Image data = dark data

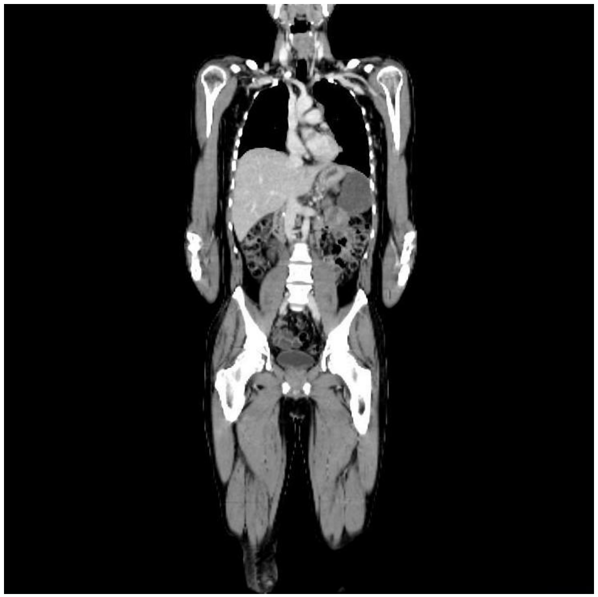
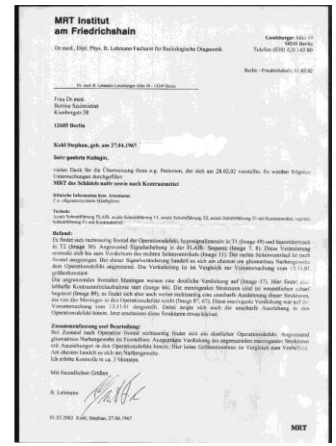


Image data
(many gigabytes)

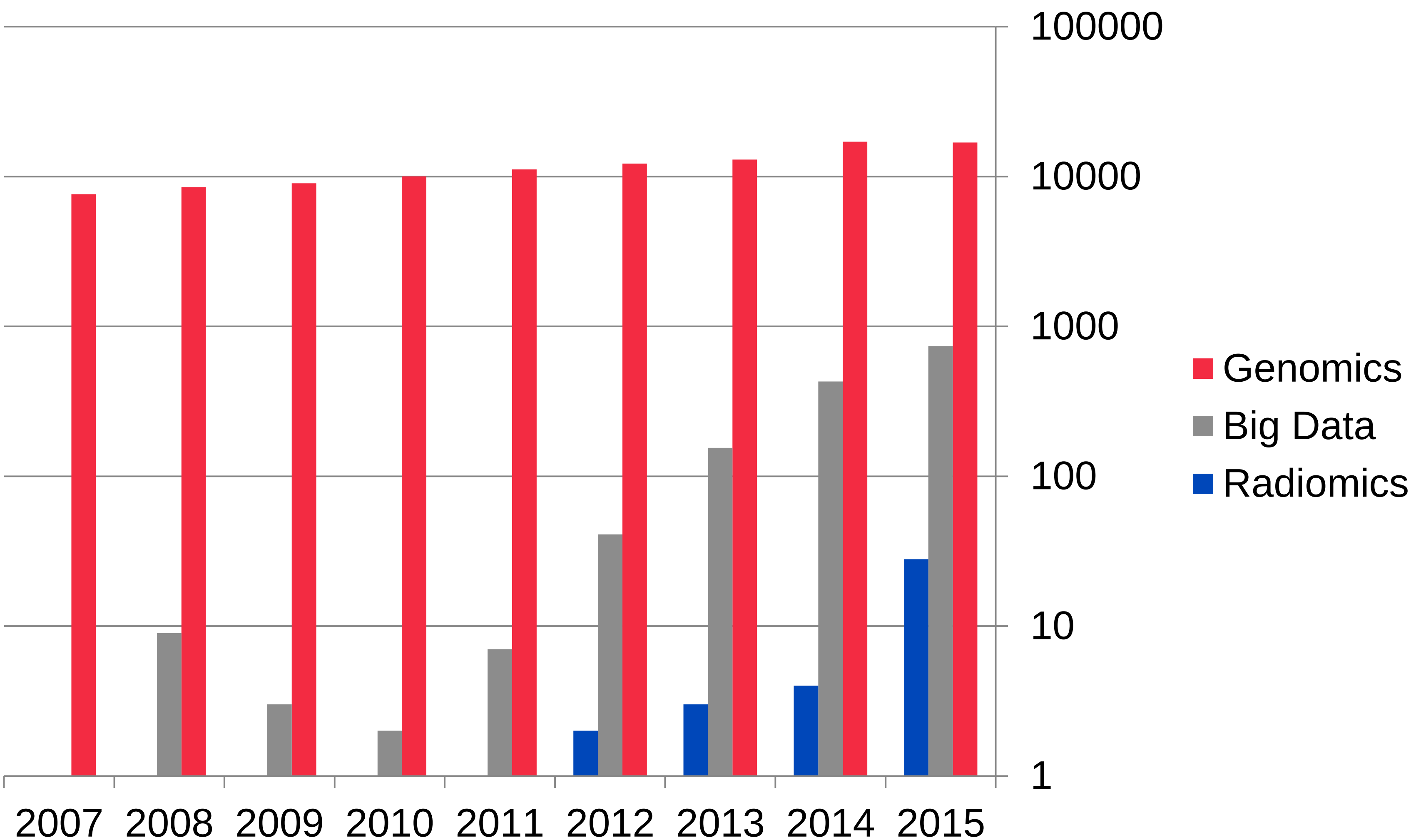
Big Data
useless!



MRT Institut
am Friedrichshain

Medical report
(few kilobytes)

„Radiomics“ in Pubmed



-omics – a field of study studying the *totality* of some sort

-ome – addresses the objects of study

Radiomics (idea)

- Clinical practice
  - Segmentation of tumor (RECIST, CTV, GTV)
  - Not predictive (overall/progression free survival)
  - Qualitative descriptions (“peripherally enhancing spiculated mass in left lobe”)

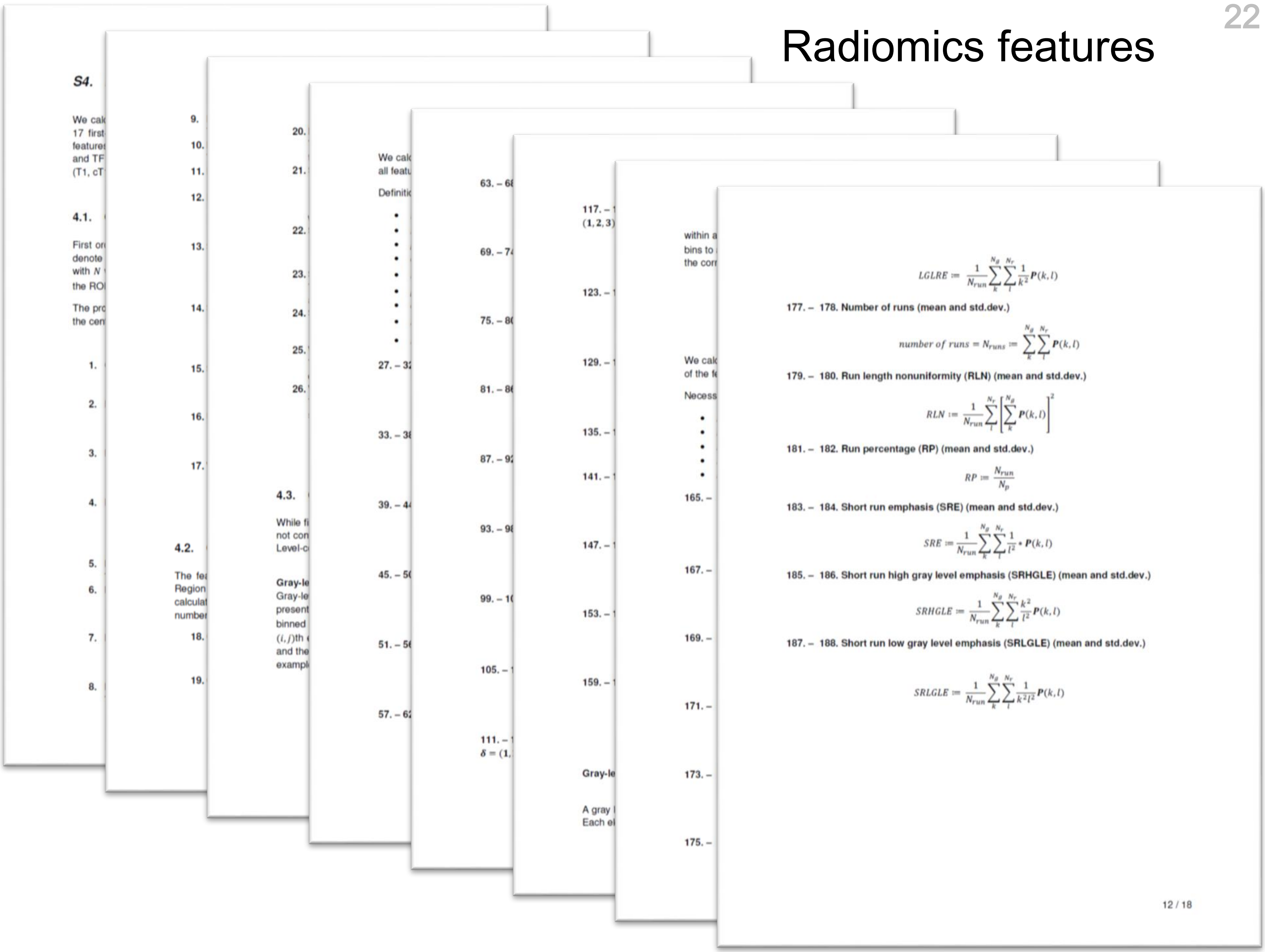
- Potential of images
  - Intensity, Texture, Shape
  - Growth and response predictions
  - …

Radiomics (definition)

- “extraction and analysis of large amounts of advanced **quantitative** imaging features with **high throughput**” [1]
- core hypothesis: resulting models provide diagnostic, prognostic or predictive information [2]
- Radio**ge**nomics

[1] Kumar V et al. Radiomics: the process and the challenges. Magn Reson Imag 30, 1234–1248 (2012).
[2] Lambin P et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur. J. Cancer 48, 441–446 (2012).

Radiomics features



$$LGLRE := \frac{1}{N_{run}} \sum_{k}^{N_g} \sum_{l}^{N_r} \frac{1}{k^2} \boldsymbol{P}(k,l)$$

177. – 178. Number of runs (mean and std.dev.)

$$number\ of\ runs = N_{runs} := \sum_{k}^{N_g} \sum_{l}^{N_r} \boldsymbol{P}(k,l)$$

179. – 180. Run length nonuniformity (RLN) (mean and std.dev.)

$$RLN := \frac{1}{N_{run}} \sum_{l}^{N_r} \left[\sum_{k}^{N_g} \boldsymbol{P}(k,l) \right]^2$$

181. – 182. Run percentage (RP) (mean and std.dev.)

$$RP := \frac{N_{run}}{N_p}$$

183. – 184. Short run emphasis (SRE) (mean and std.dev.)

$$SRE := \frac{1}{N_{run}} \sum_{k}^{N_g} \sum_{l}^{N_r} \frac{1}{l^2} * \boldsymbol{P}(k,l)$$

185. – 186. Short run high gray level emphasis (SRHGLE) (mean and std.dev.)

$$SRHGLE := \frac{1}{N_{run}} \sum_{k}^{N_g} \sum_{l}^{N_r} \frac{k^2}{l^2} \boldsymbol{P}(k,l)$$

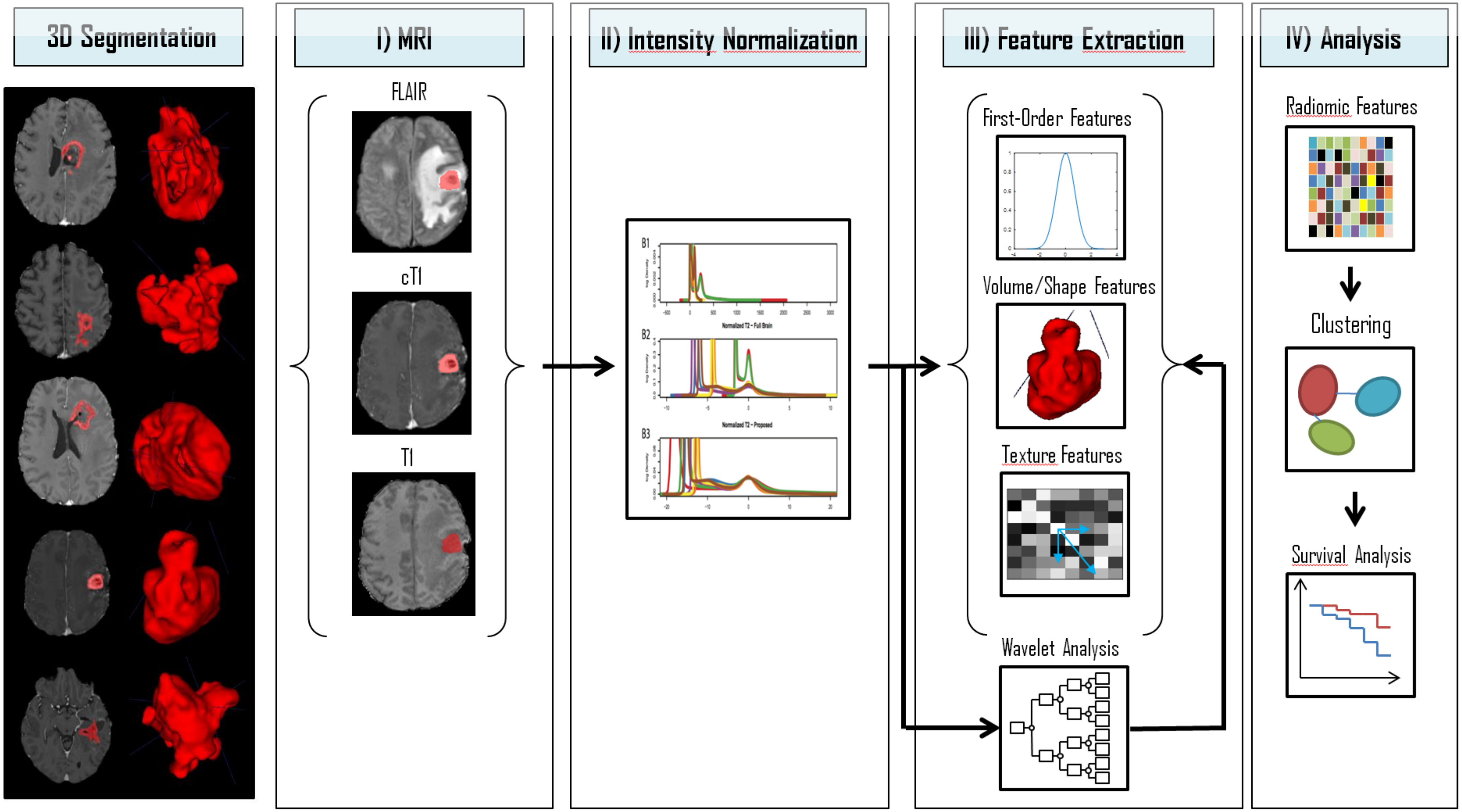
187. – 188. Short run low gray level emphasis (SRLGLE) (mean and std.dev.)

$$SRLGLE := \frac{1}{N_{run}} \sum_{k}^{N_g} \sum_{l}^{N_r} \frac{1}{k^2 l^2} \boldsymbol{P}(k,l)$$

12 / 18

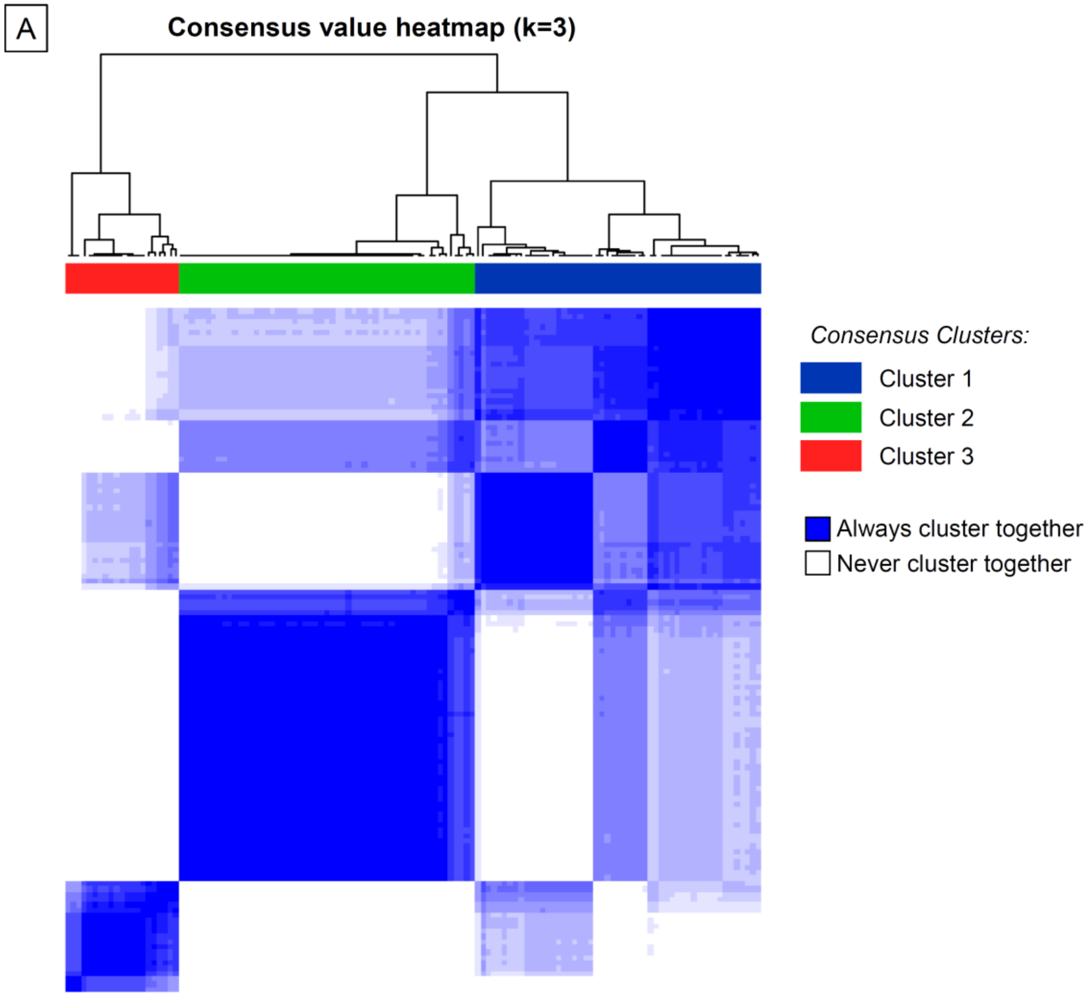
Predicting antiangiogenic treatment response in glioblastoma

- 4842 quantitative MRI features (T1, cT1, FLAIR), 129 patients



Kickingereder P et al. Large-scale radiomic profiling of glioblastoma identifies an imaging signature for predicting and stratifying antiangiogenic treatment response. (submitted)

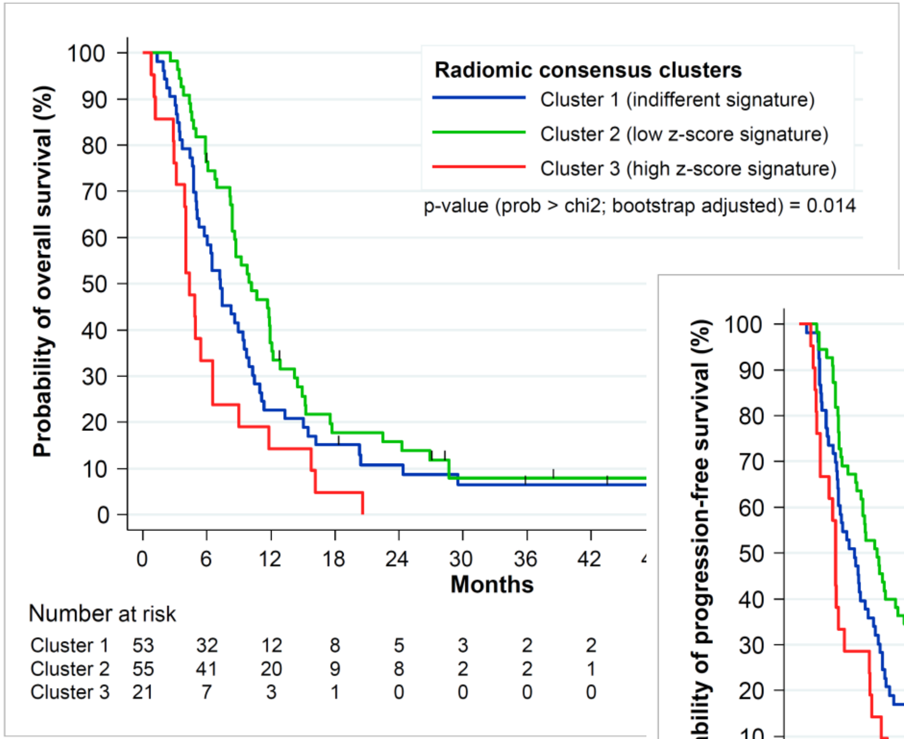
Predicting antiangiogenic treatment response in glioblastoma

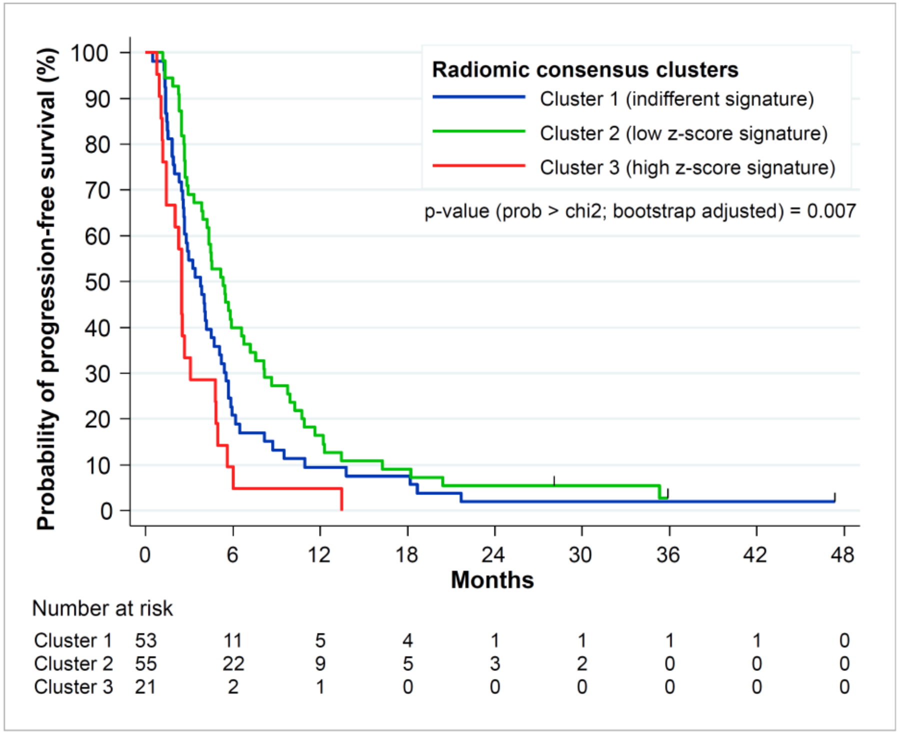


Kickingereder P et al. Large-scale radiomic profiling of glioblastoma identifies an imaging signature for predicting and stratifying antiangiogenic treatment response. (submitted)

Klaus H. Maier-Hein
Medical Imaging Computing Group

Predicting antiangiogenic treatment response in glioblastoma





Kickingereder P et al. Large-scale radiomic profiling of glioblastoma identifies an imaging signature for predicting and stratifying antiangiogenic treatment response. (submitted)

Contribution to personalized medicine

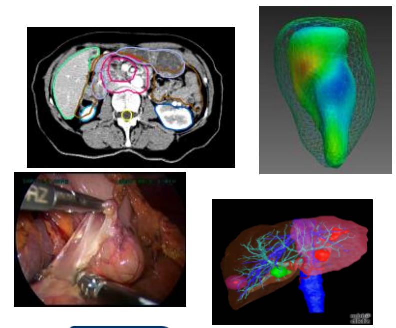
- Genomic/proteomic technologies **blind** to spatial + temporal tumor heterogeneity
  - Based on biopsy or invasive surgery
  - Analysis of small portions of the tumor
- Strengthen the role of medical imaging in „personalized medicine“
  - Comprehensive view of entire tumor
  - Monitor development and progression of disease/response to therapy
  - Noninvasive, existing data

Big data challenges

- Capture
- Manage
- Process
- Share
- Integrate
- Analyze
- Interpret

Big data challenges

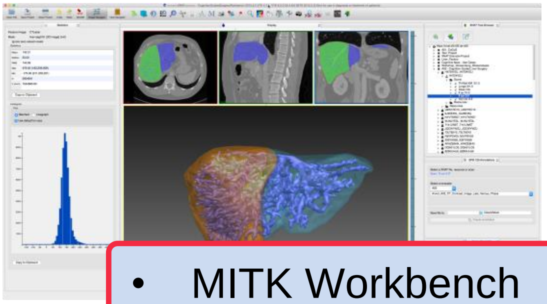
- Capture
- Manage
- Process
- Share
- Integrate
- Analyze
- Interpret



XNAT:
- Study collections
- Study objects
- Imaging sessions
- Annotations

Big data challenges

- Capture
- Manage
- Process
- Share
- Integrate
- Analyze
- Interpret



- MITK Workbench
- CTK Command Line
- Python Scripting

XNAT:
- Study collections
- Study objects
- Imaging sessions
- Annotations

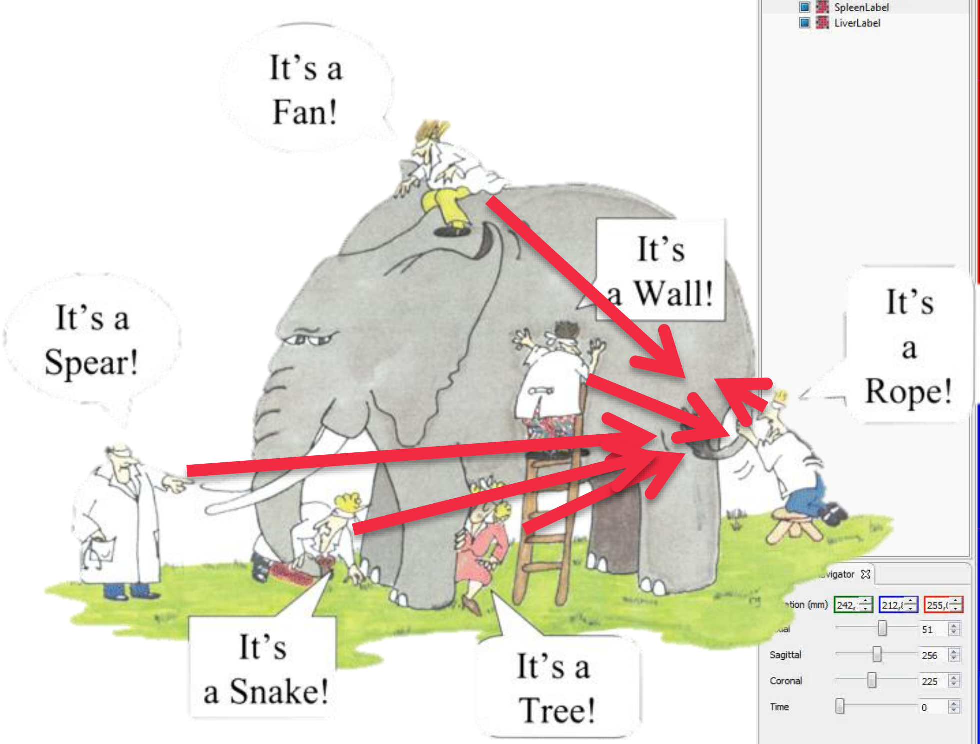
Klaus H. Maier-Hein

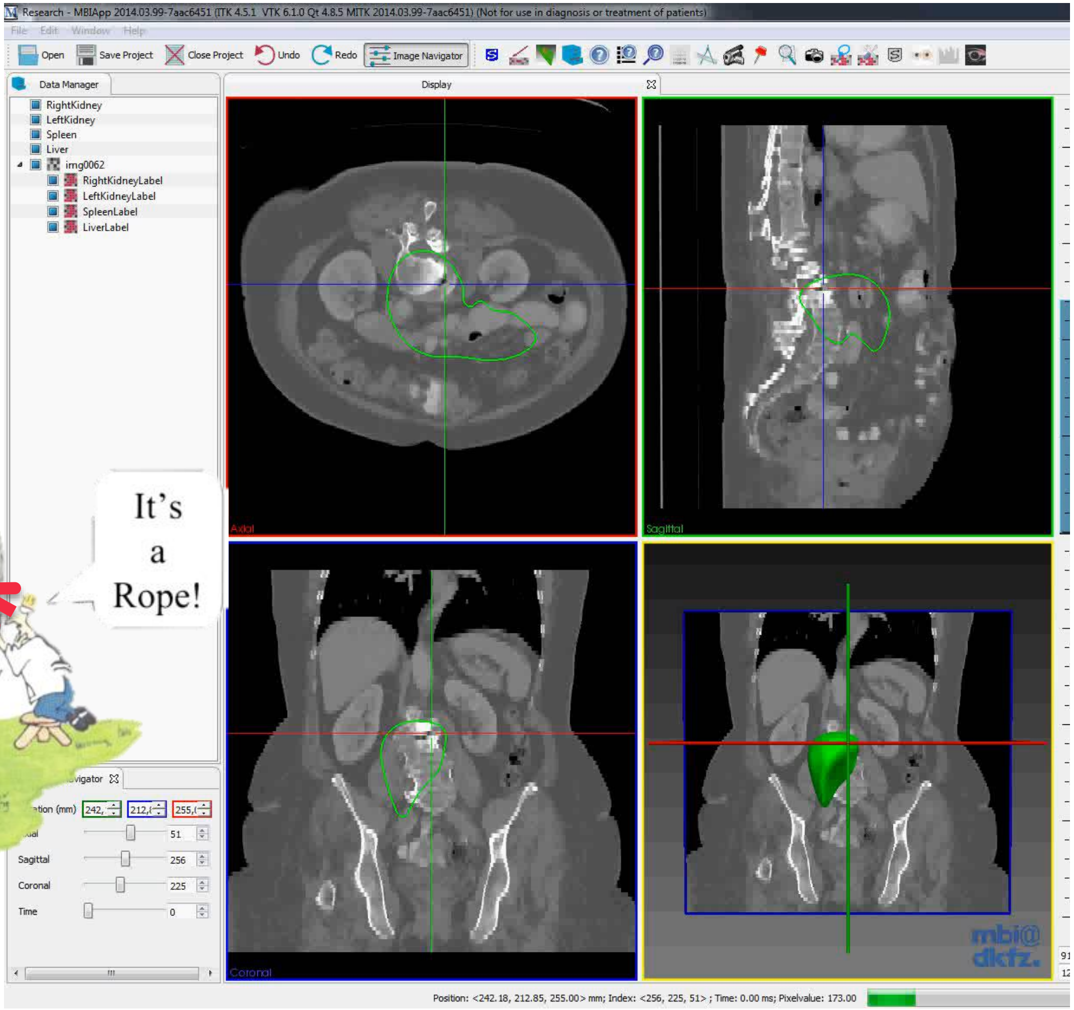
Medical Imaging Computing Group

Novel generation of machine-learning driven segmentation techniques

~~Question asked: IS THIS …?~~

Novel question: WHERE IS …?



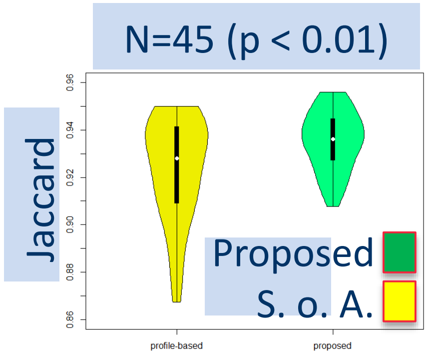


Norajitra T, et al. Improved 3D Statistical Shape Model Guidance using Regression-based Landmark Detection. *IEEE TMI*, 2015 (submitted)

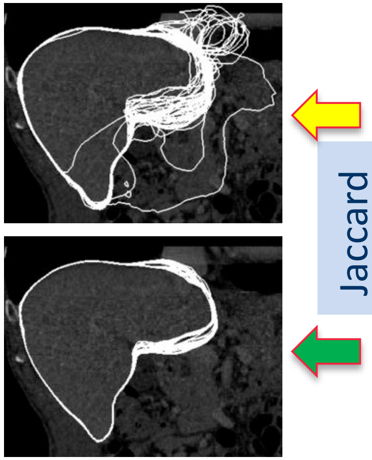
Norajitra T, et al. 3D Regression Voting on CT-Volumes fo the Human liver for SSM Surface Appearance Modeling. *Shape 2014*

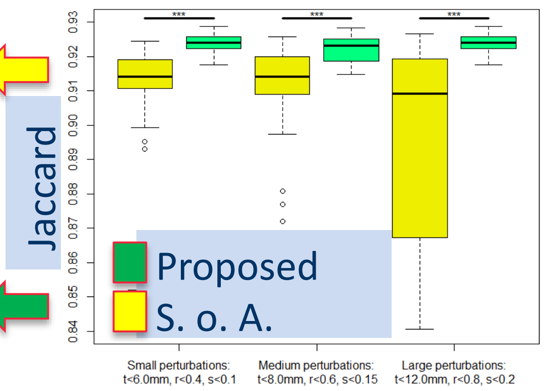
Norajitra T, et al. 3D SSM incorporating 3D Random Forest Regression Voting for Robust CT Liver Segmentation, *SPIE 2015*

Novel generation of machine-learning driven segmentation techniques

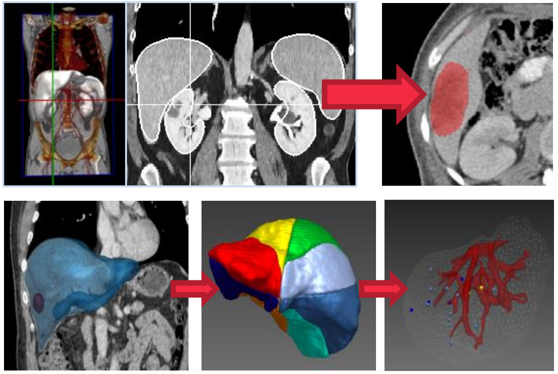


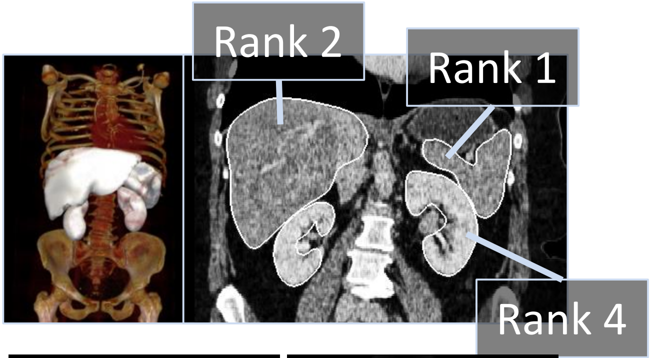
Improved Accuracy

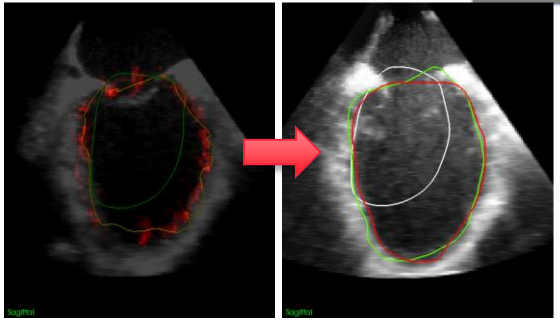




Independent of Initialization

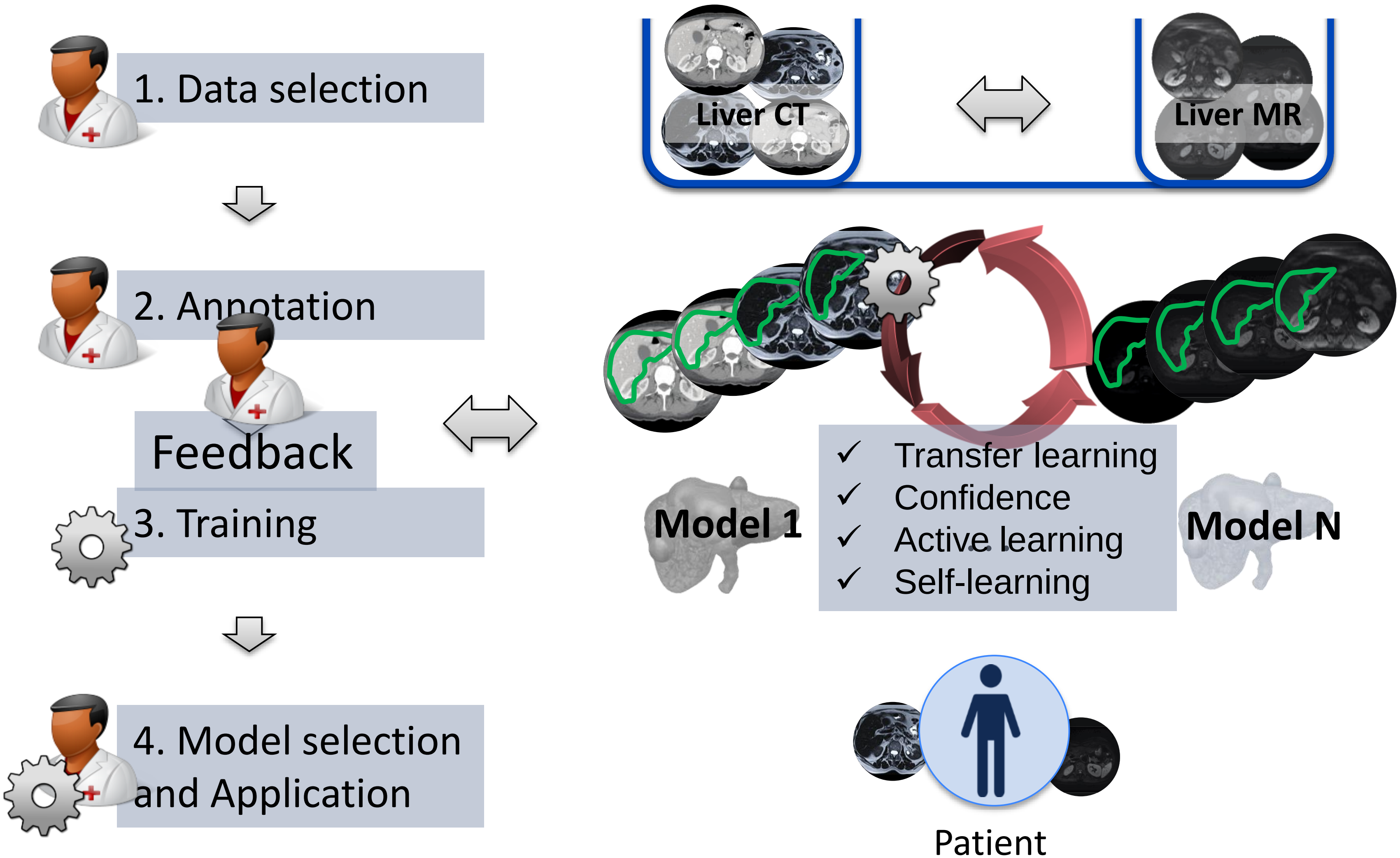






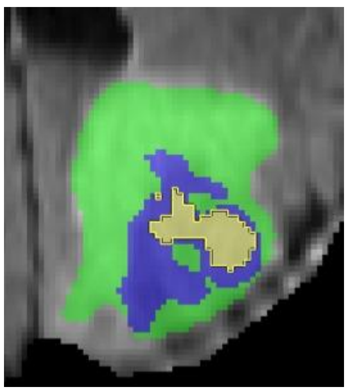
[1] Norajitra T, et al. **IEEE-TMI 2015.** 3D Statistical Shape Models incorporating Landmark-wise RRF. *(submitted)*
[2] Norajitra T, et al. **Shape 2014.** 3D Regression Voting on Human liver CTs for SSM Surface Appearance Modeling.
[3] Norajitra T, et al. **SPIE 2015.** 3D SSM incorporating 3D RRFVoting for Robust CT Liver Segmentation.

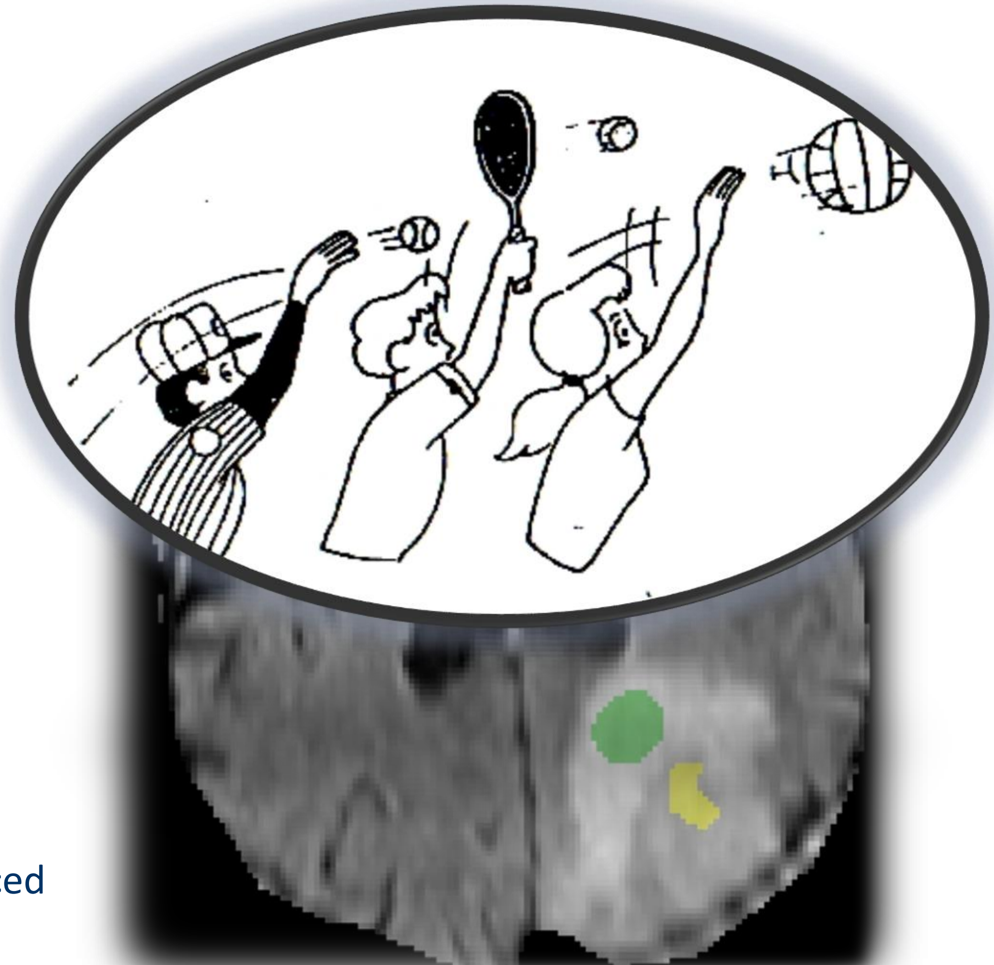
Self-learning models



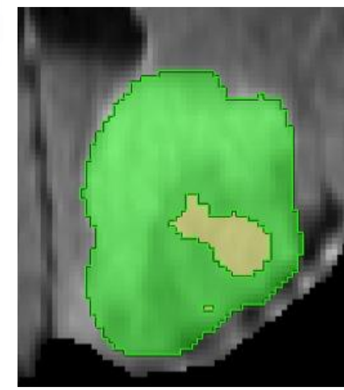
Transfer learning solves data annotation problem

Expert 1





Expert 2

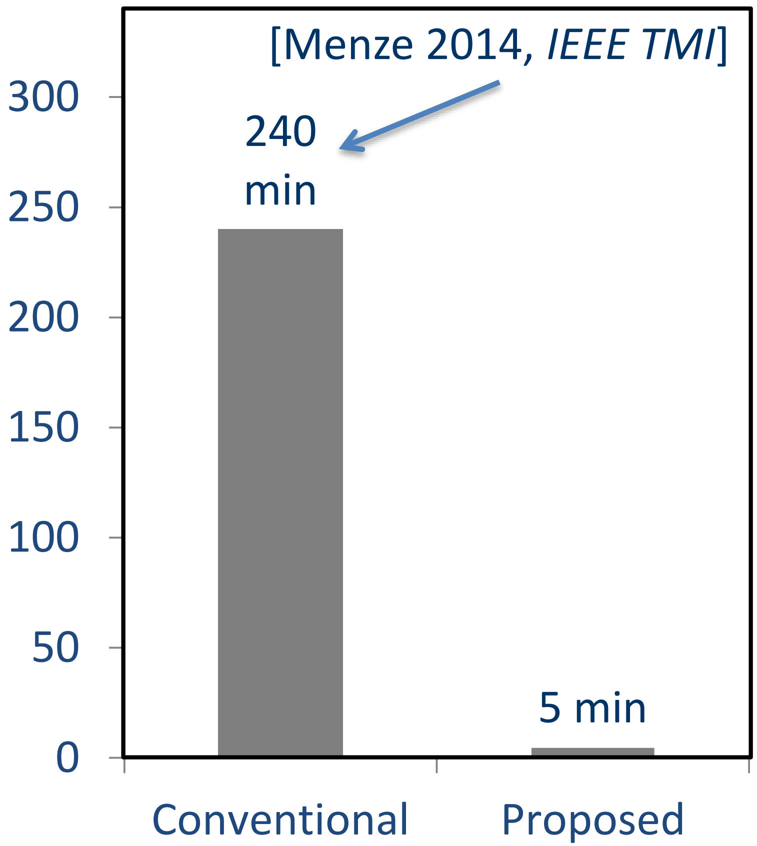


- Edema
- Active tumor
- Contrast-enhanced

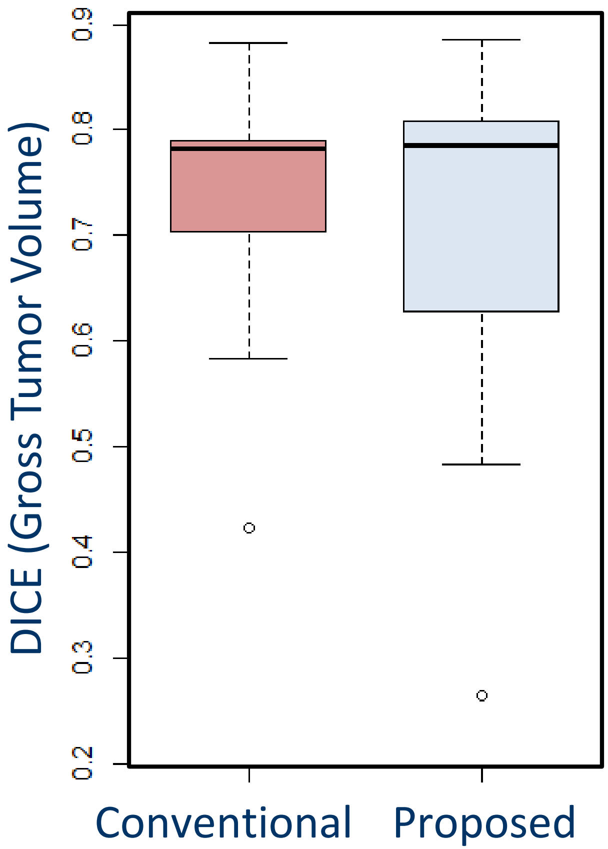
[1] **Götz M, et al.** DALSA: Domain Adaptation for Supervised Learning from Sparsely Annotated MR Images. *IEEE TMI* 2015
[2] **Götz M, et al.** Extremely randomized trees based brain tumor segmentation, *Proc. MICCAI 2014 BraTS Challenge*, 2014
[3] **Götz M, et al.** Input Data Adaptive Learning (IDAL) for sub-acute Ischemic Stroke Lesion Segmentation. *Springer LLNC,* 2015

Recuced annotation time, similar accuracy

Labeling Time



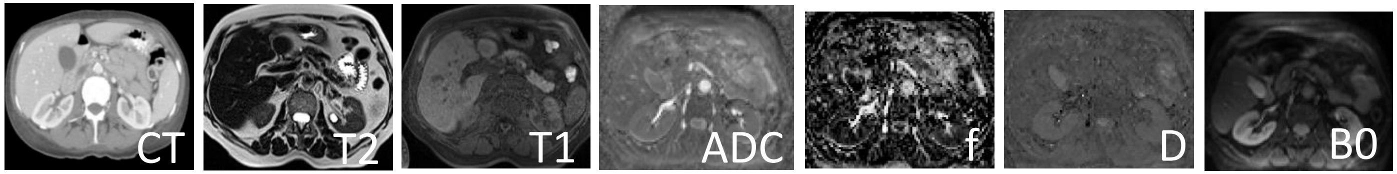
Segmentation Quality



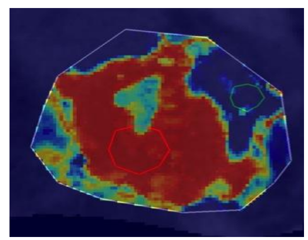
[1] **Götz M, et al.** DALSA: Domain Adaptation for Supervised Learning from Sparsely Annotated MR Images. *IEEE TMI* 2015
[2] **Götz M, et al.** Extremely randomized trees based brain tumor segmentation, *Proc. MICCAI 2014 BraTS Challenge*, 2014
[3] **Götz M, et al.** Input Data Adaptive Learning (IDAL) for sub-acute Ischemic Stroke Lesion Segmentation. *Springer LLNC,* 2015

Klaus H. Maier-Hein
Medical Imaging
Computing Group

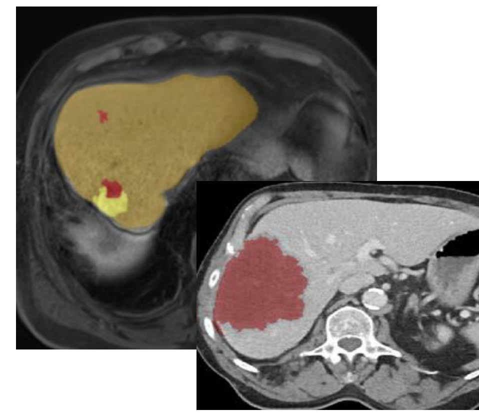
One method, many applications



DECT vs. Perfusions-CT

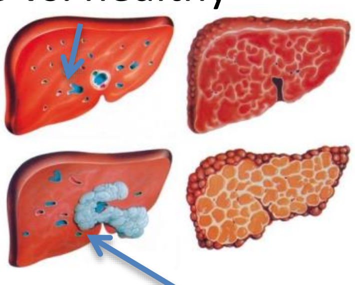


Tumor load/location



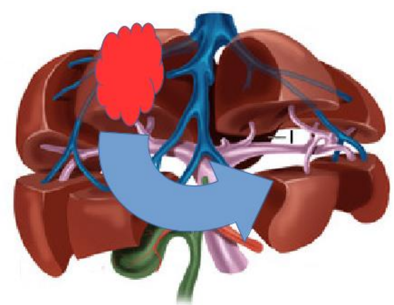
Tissue characterization

HCC vs. healthy



CCC vs. Fatty liver

Predictions

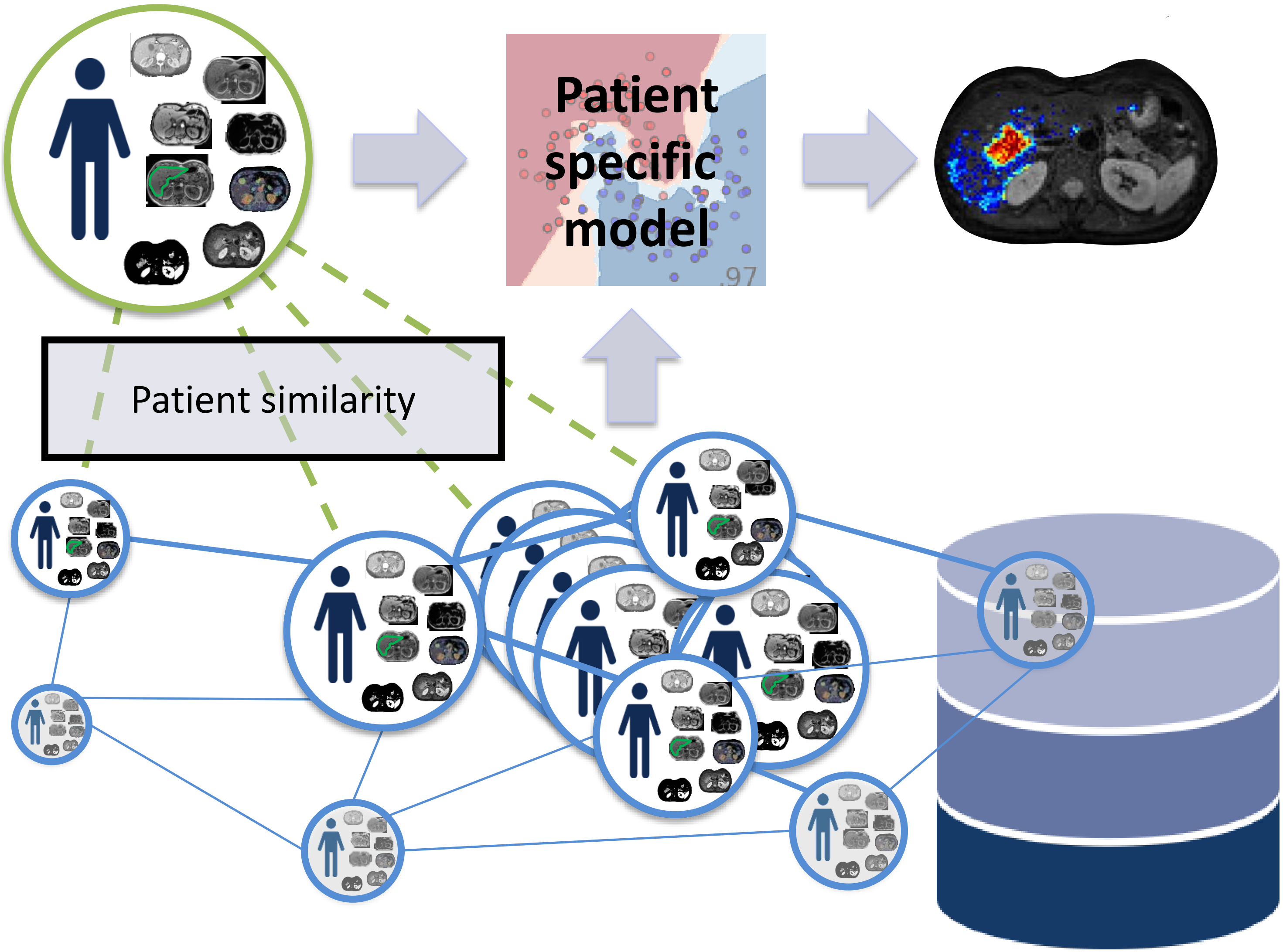


[1] Götz M, et al. **IEEE TMI 2015.** DALSA: Domain Adaptation for Learning from Sparsely Annotated MR Images.
[2] Götz M, et al. **Proc. MICCAI 2014 BraTS Challenge, 2014.** Extremely randomized trees for tumor segmentation.
[3] Götz M, et al. **Springer LLNC, 2015.** Input Data Adaptive Learning (IDAL) for Lesion Segmentation.

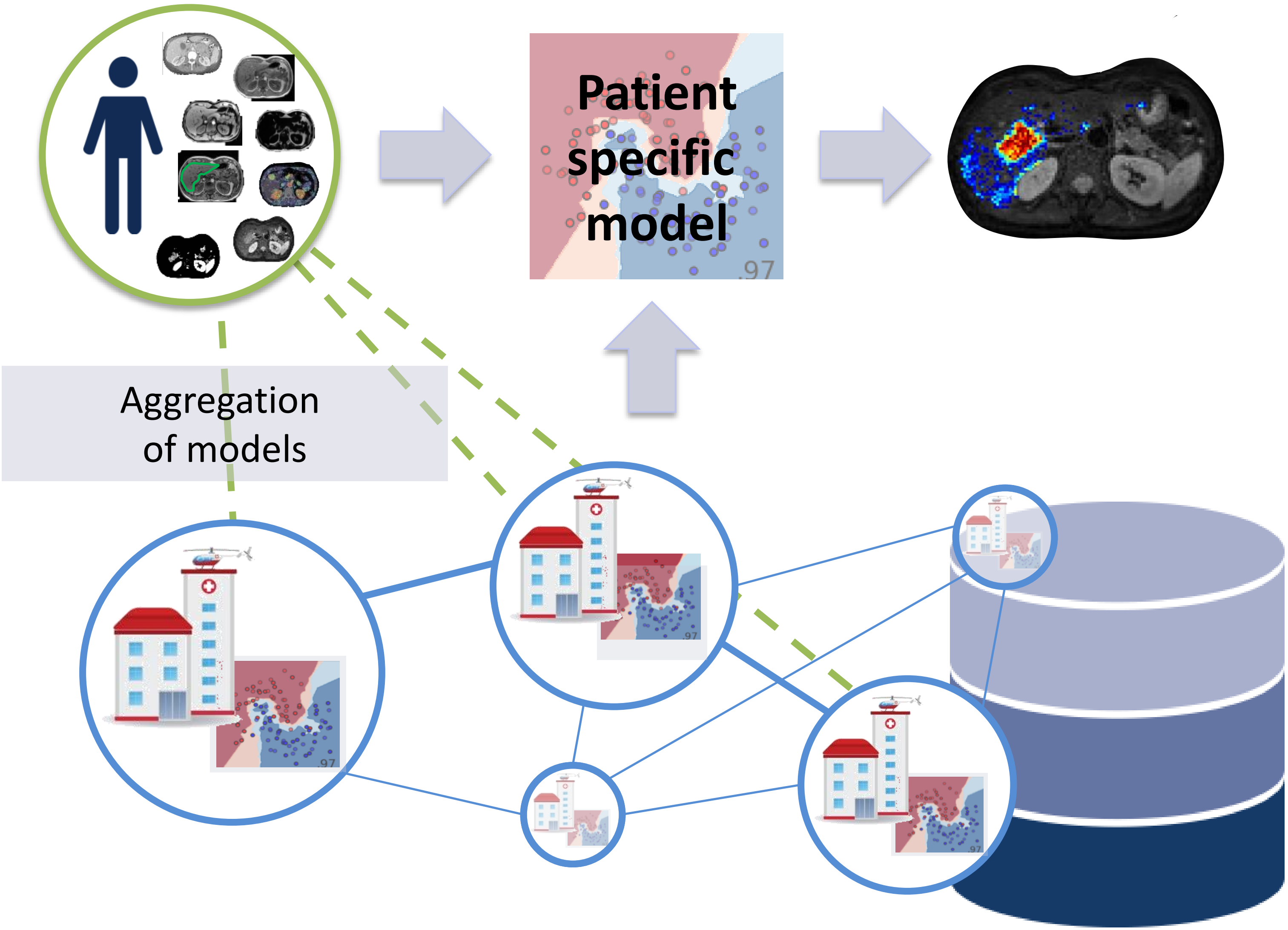
Big data challenges

- Capture
- Manage
- Process
- Share
- Integrate
- Analyze
- Interpret

Patient specific models



Distributed knowledge



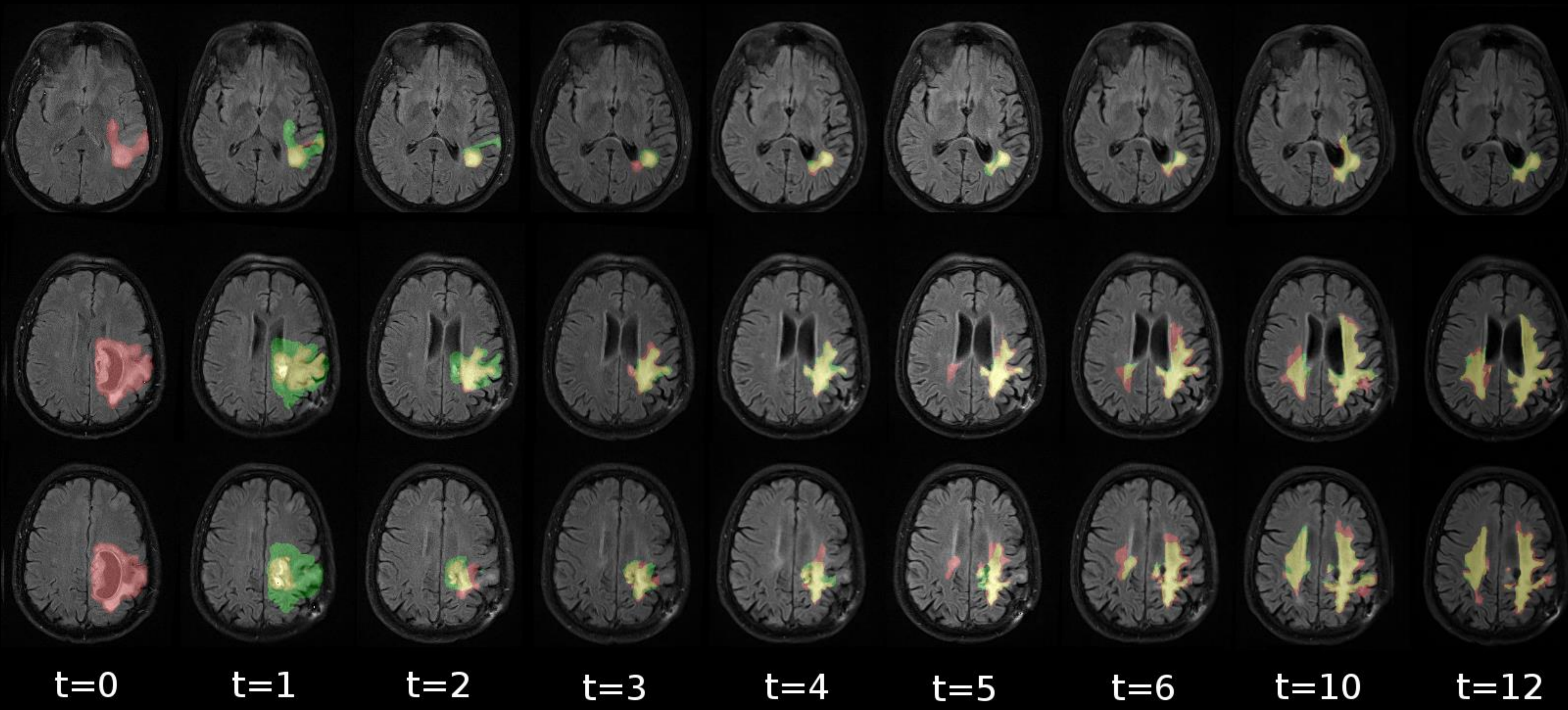
Big data challenges

- Capture
- Manage
- Process
- Share
- Integrate
- Analyze
- Interpret

Big data challenges

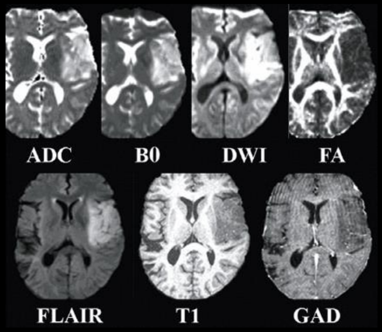
- Capture
- Manage
- Process
- Share
- Integrate
- Analyze
- Interpret

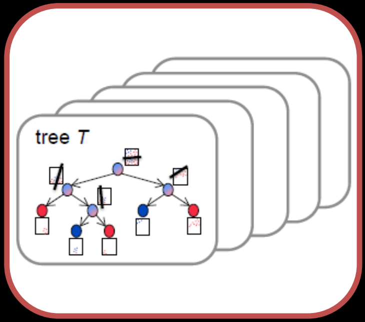
Prognostic models?

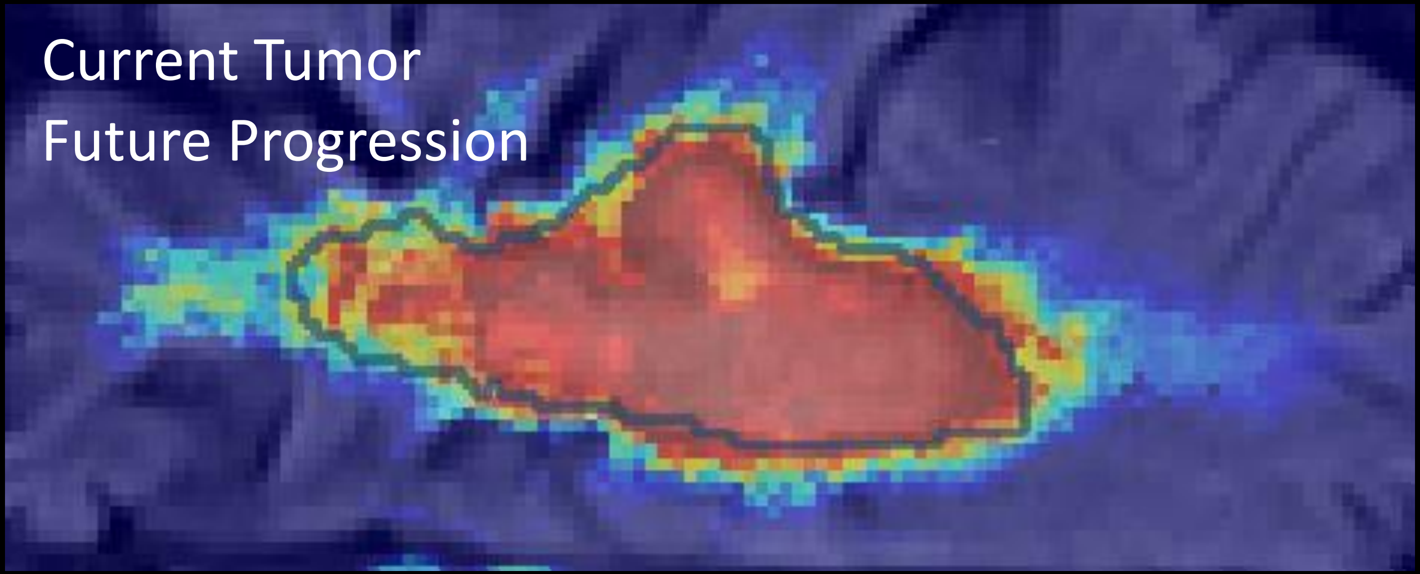


- Weber et al..Tumor Progression Mapping: An Intuitive Visualization of Glioblastoma Progression in MR Follow-ups. ISMRM (2014)

Machine learning & disease mechanisms

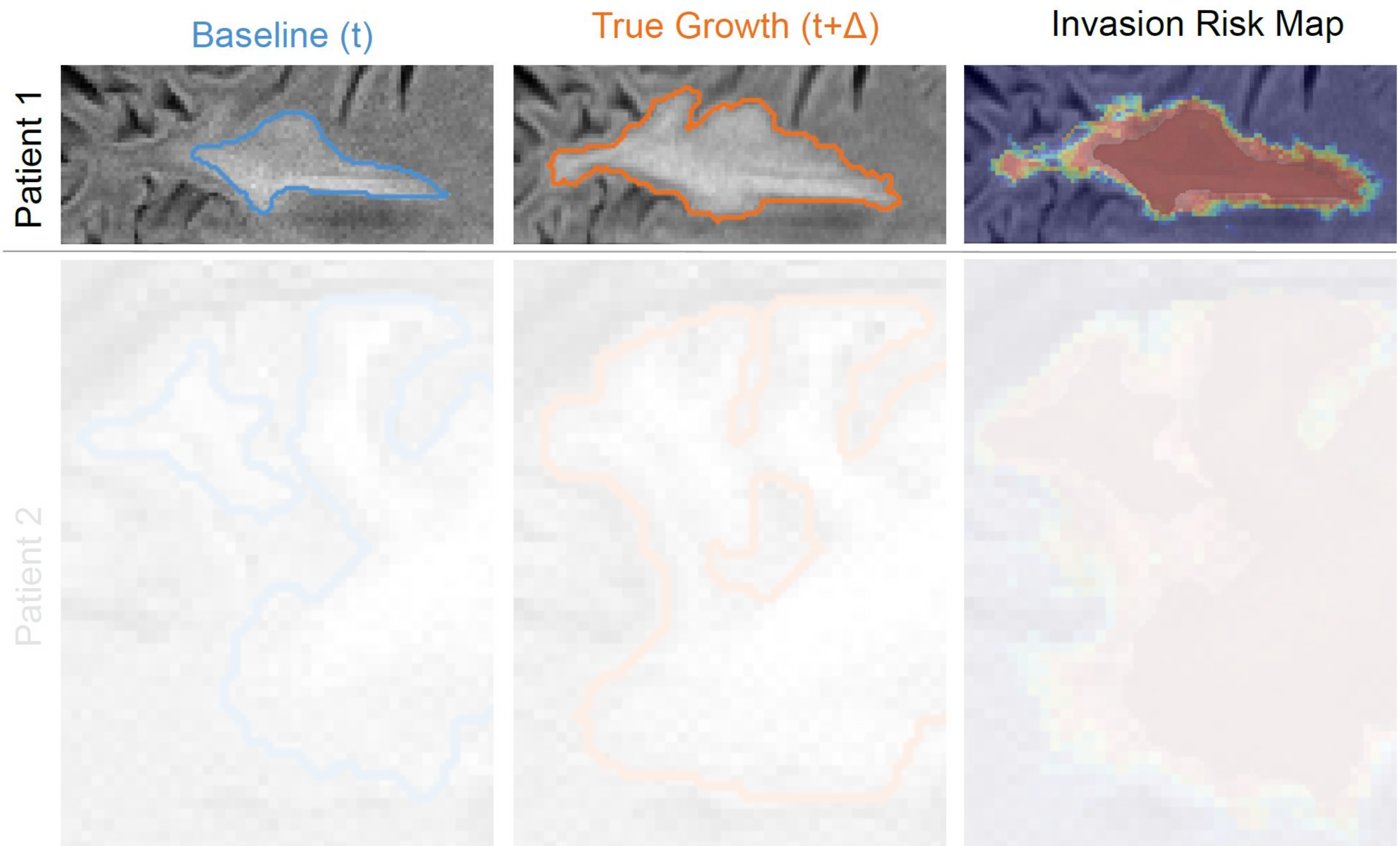




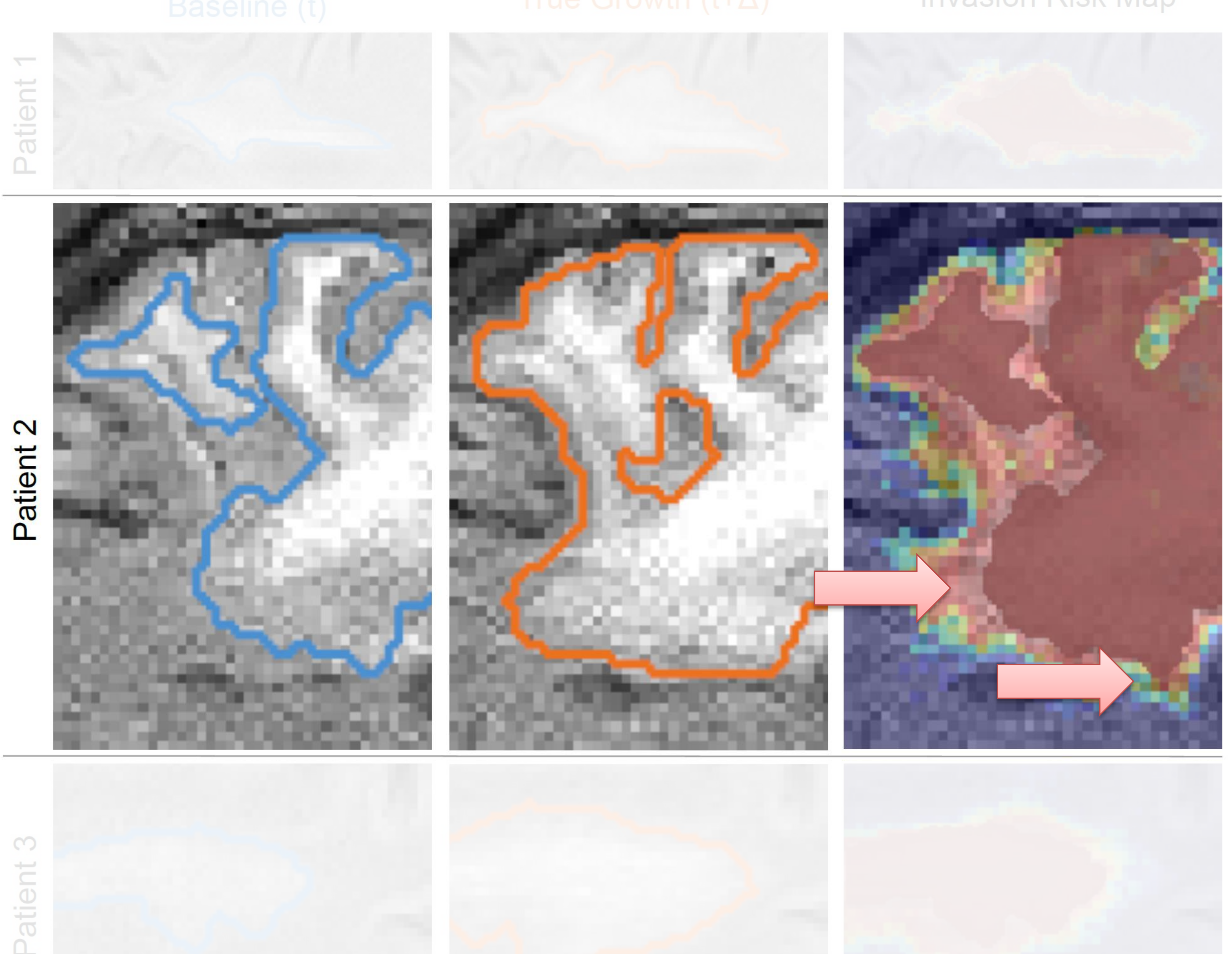


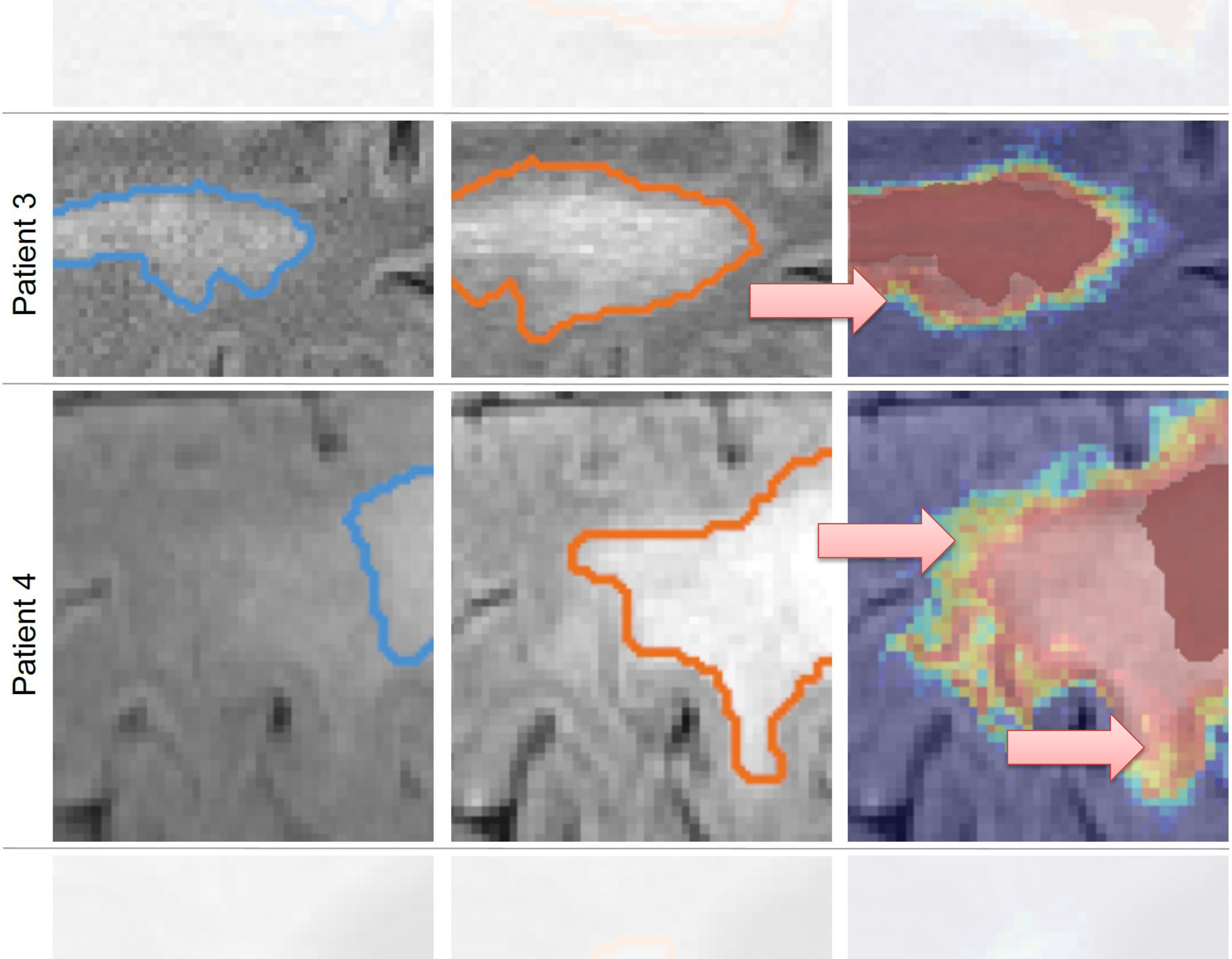
- Unpublished data: Weber et al. Individual estimation of brain tumor invasion margin using data-driven glioma growth models. (in preparation)

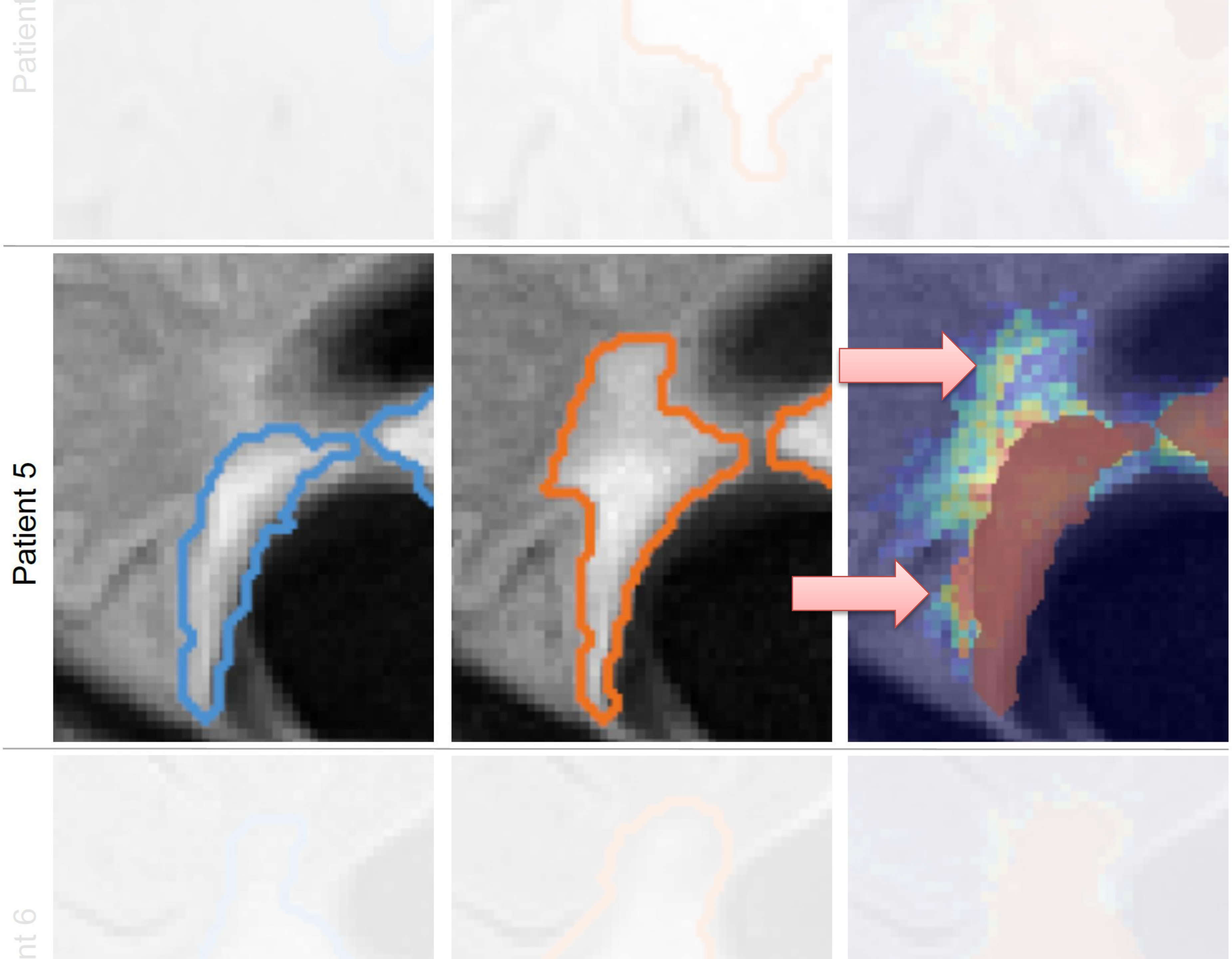
Results

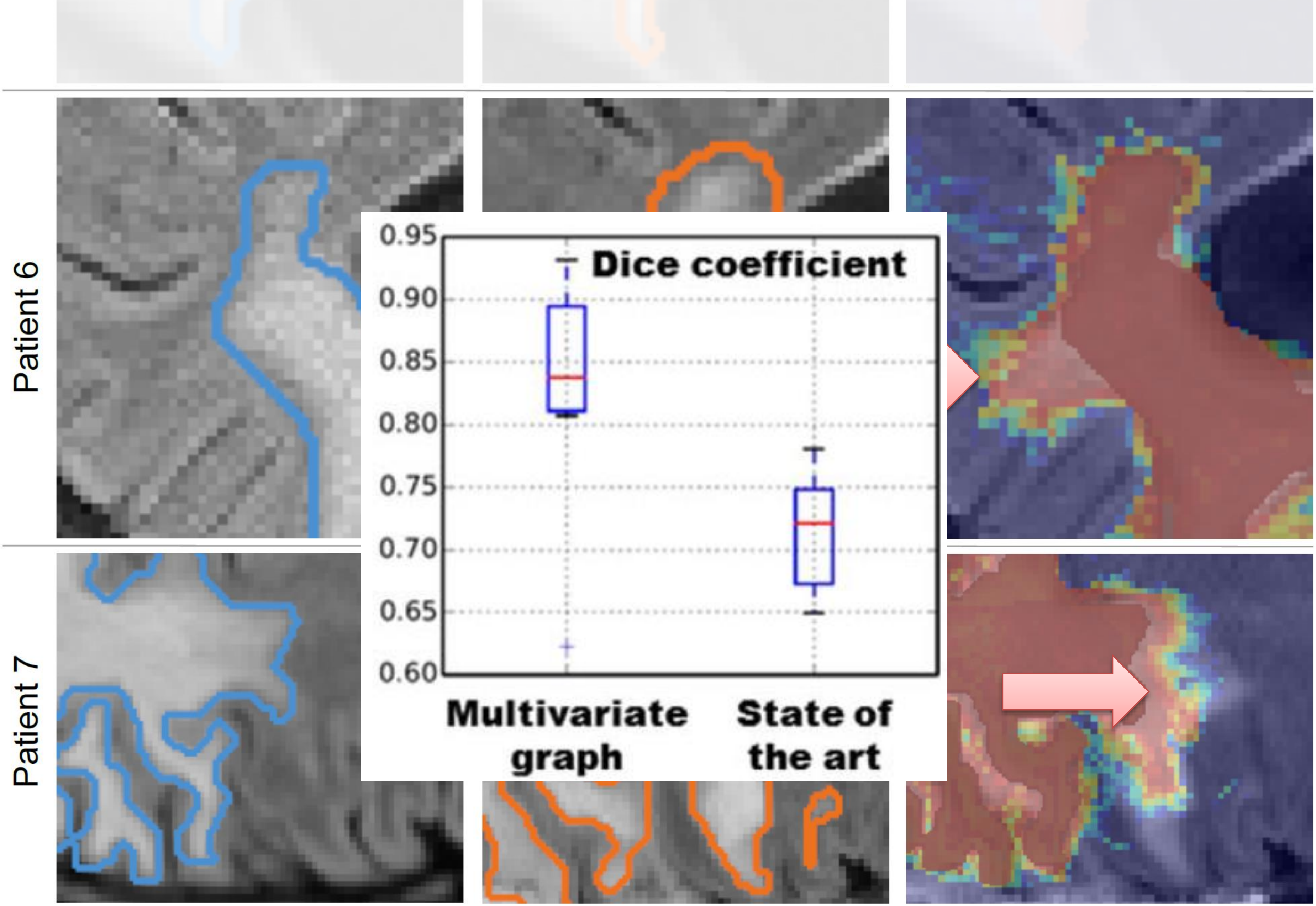


Baseline (t)
True Growth (t+Δ)
Invasion Risk Map
Patient 1
Patient 2

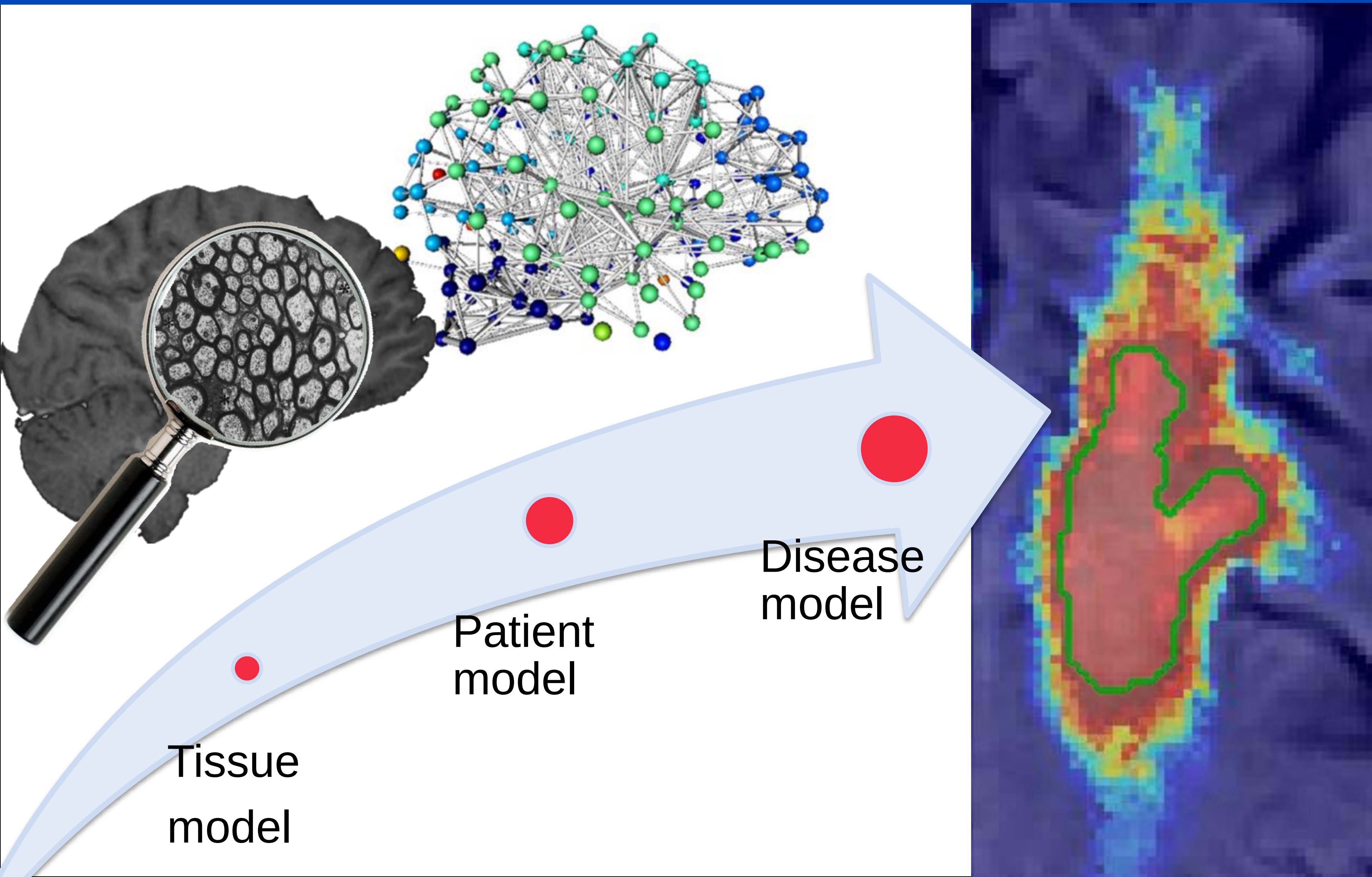

Baseline (t)
Invasion Risk Map
Patient 1
Patient 2
Patient 3


Patient 3
Patient 4

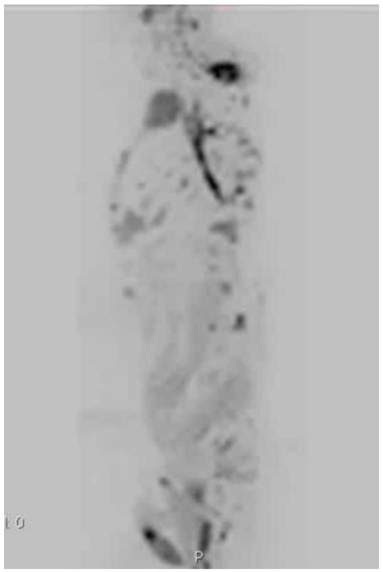

Patient 5

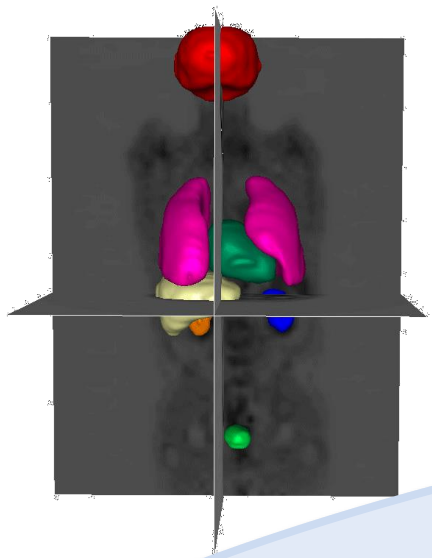

Patient 6
Patient 7
Dice coefficient
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
Multivariate graph
State of the art

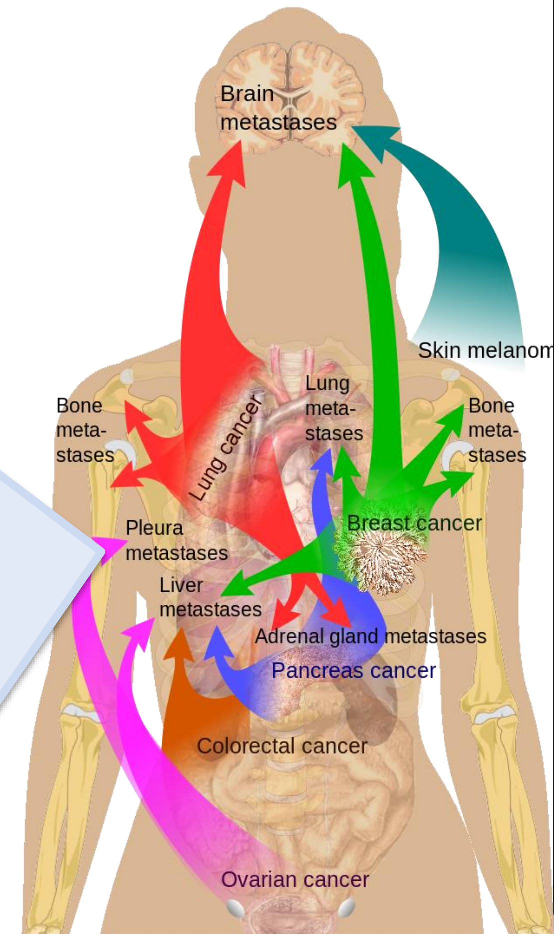
Prediction of brain tumor growth



Modeling patterns of metastases







Tissue model

Patient model

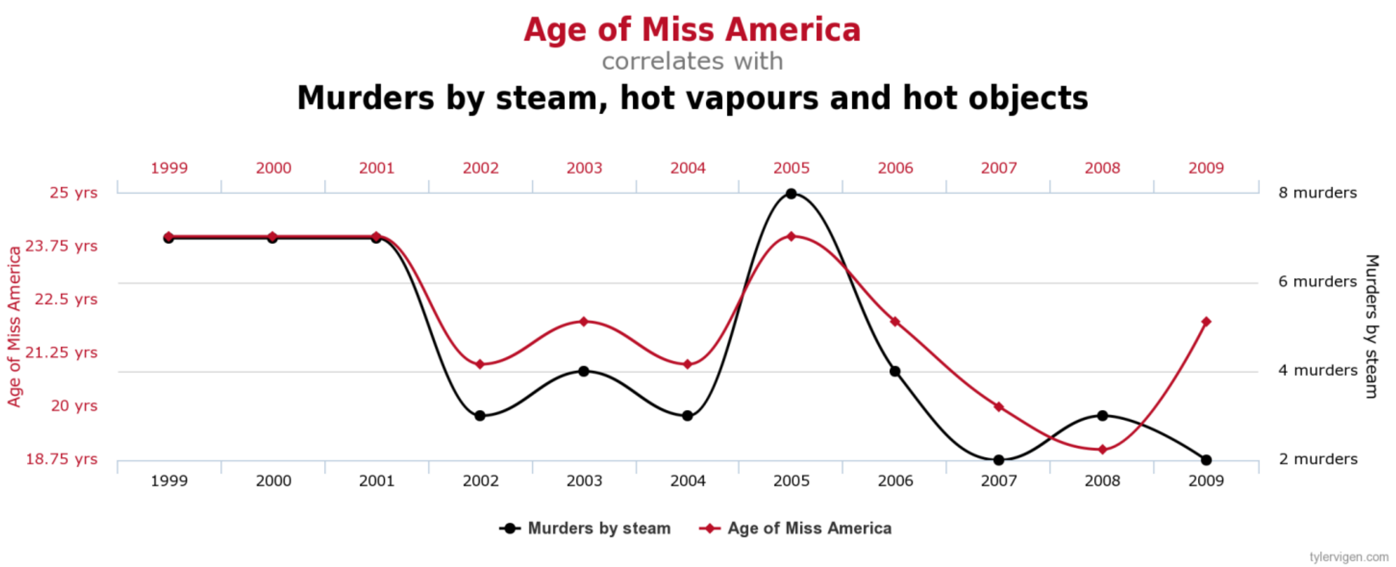
Disease modell

Big data challenges

- Capture
- Manage
- Process
- Share
- Integrate
- Analyze
- Interpret

Is the scientific method itself becoming obsolete?

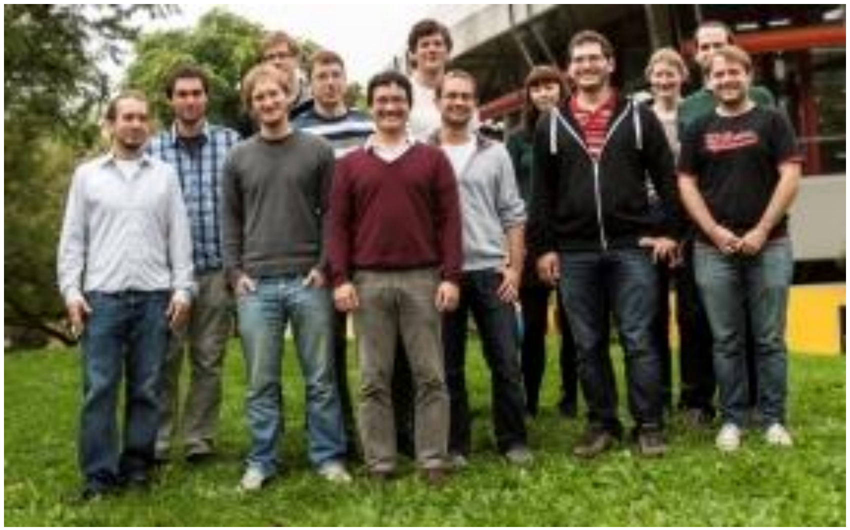
- Proportion of false alarms among proposed “findings” may increase when one can measure more things
- Big Data’s is a new tool, finds associations, does not in show whether these have meaning
- Does not replace randomized clinical trials and other experimental designs



Summary

- More in the data than expected: Use data-driven approaches!
- Radiomics for image-based personlized medicine
- Potential impact for radiation theraphy
- Many Challenges ahead

Thank you!



Dr. Peter Neher	Jens Petersen
Dr. Dr. J. Kleesiek	Fabian Isensee
Dr. Caspar Goch	Joseph Görres
Michael Götz	Jan Hettich
Jan Hering	Jonas Cordes
Christoph Kolb	M. Eisenmann
Tobias Norajitra	Vincent Reuter
Christian Weber	Alex Tschlatscher

- Bram Stieltjes, Univ. Basel
- Ofer Pasternak et al., Harvard, Boston
- Gary Zhang, UCL, London
- Ashish Raj, Weill Cornell, New York
- Bülent Yener, Rensselaer Polyt.
- Alex Leemans et al., Utrecht

- Collaborators @DKFZ Heidelberg
- Collaborators @Uni/NCT Heidelberg
- Collaborators @KIT Karlsruhe
- Collaborators @IWR Heidelberg
- Collaborators @SFB/Transregio 125

The MITK team
The MBI team