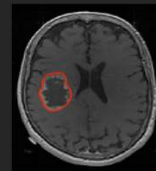
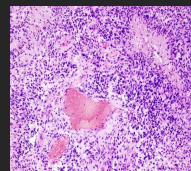
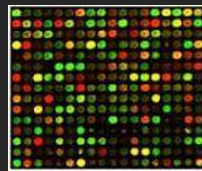




Multi-modal data fusion in biomedicine

and my road from PhD student to Stanford

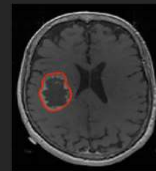
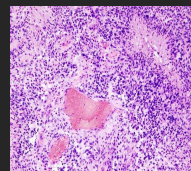
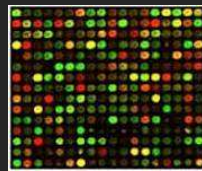


Olivier Gevaert, Ph.D.
Associate Professor

Stanford Center for Biomedical Informatics Research (BMIR)
Department of Medicine
& Department of Biomedical Data Science
Stanford University



Multi-modal data fusion in biomedicine and my road from **PhD** student to Stanford

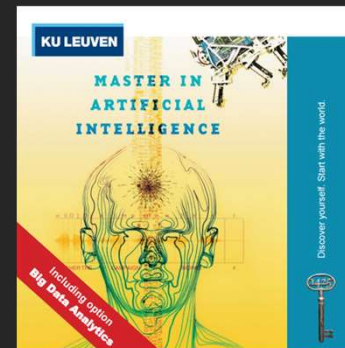
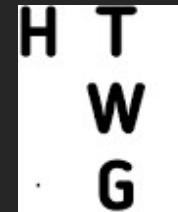


Olivier Gevaert, Ph.D.
Associate Professor

Stanford Center for Biomedical Informatics Research (BMIR)
Department of Medicine
& Department of Biomedical Data Science
Stanford University

My trajectory to PhD with BDM

- Industrial engineering ICT – KAHO Sint-Lieven Ghent
- Erasmus – Konstanz
Fachhochschule - EuropaHaus
- MaNaMa - Master of Artificial Intelligence – KU Leuven



Phd @ Esat

- I learned the concept of a PhD from classmates during master of AI and started contacting faculty.
- I met Bart in the summer of 2004 one day before a two week holiday
 - I did not say much 😊
- When I came back I had a voice mail from Bart
 - “You can start October 1st ...”

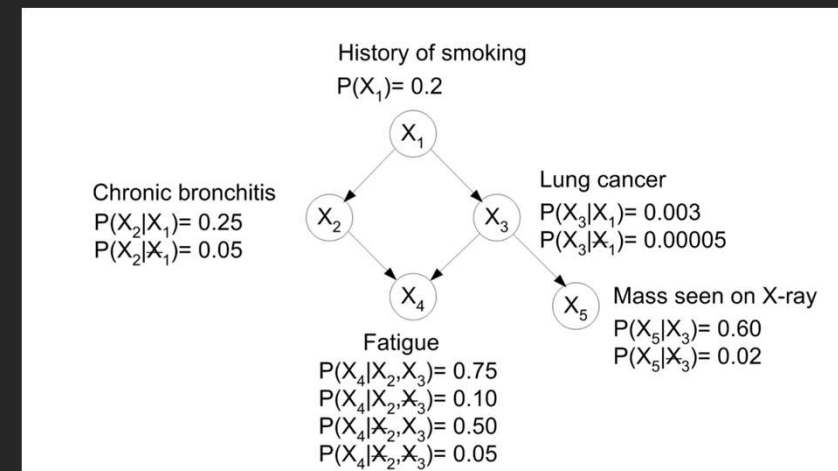
Bioinformatics @ ESAT

- 2nd generation of Bioinformatics students at ESAT
 - Co-advised by Dr. Dirk Timmerman, UZ Leuven
 - Co-mentored by 1st generation student: Frank De Smet (CM)
- Other 1st generation students who had just finished PhD
 - Gert Thijs (#22, Silicos, Cartagena, Agilent, **first bioinformatics student**)
 - Stein Aerts (#23, VIB/KUL)



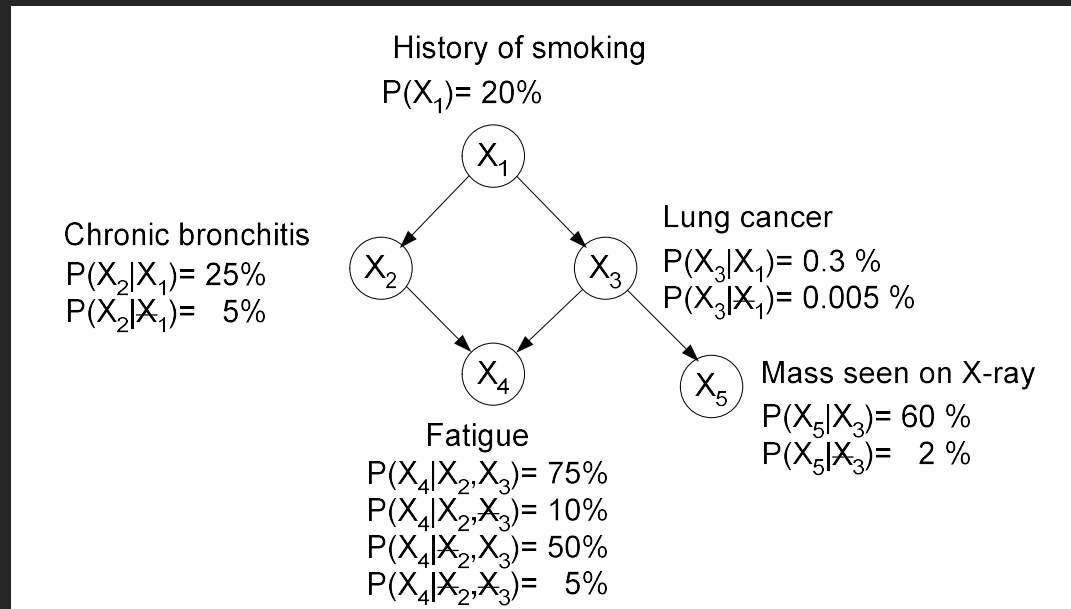
Initial steps in data fusion

- **Thesis topic:** Bayesian networks for fusion of clinical, ultrasound and microarray data
 - White box model with probabilistic interpretation
 - Classification based on an arbitrary instantiation of the genes
 - Bayesian networks are popular for modelling gene networks
 - Not too abstract (vs. boolean models) and not too detailed (vs. differential equations)

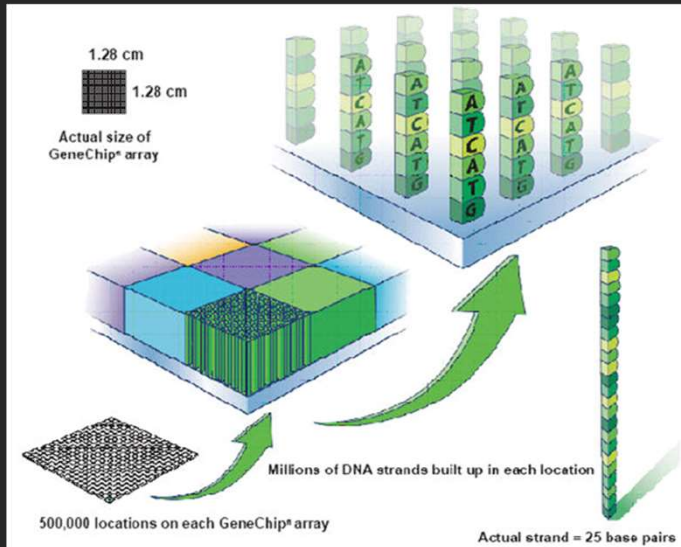


Definition

- A Bayesian network consists of two parts
 - Structure: directed acyclic graph
 - Parameters: conditional probability tables (CPT)



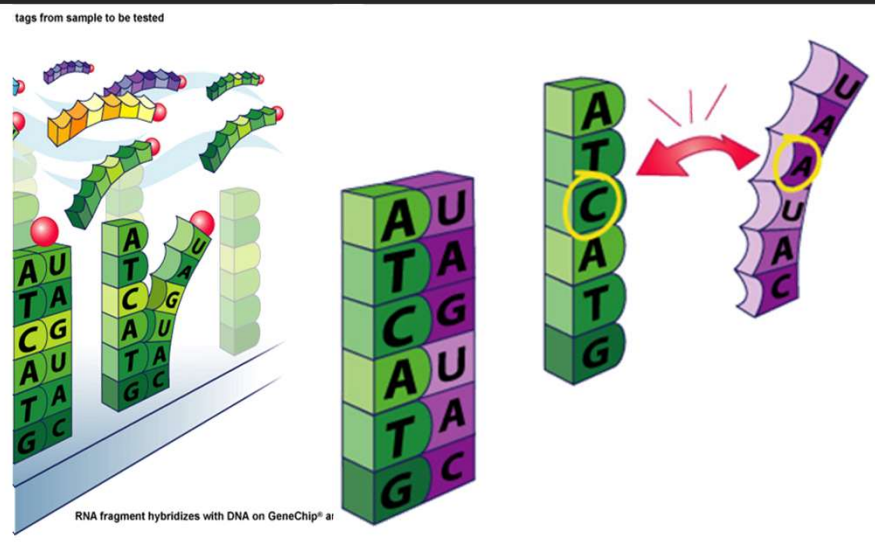
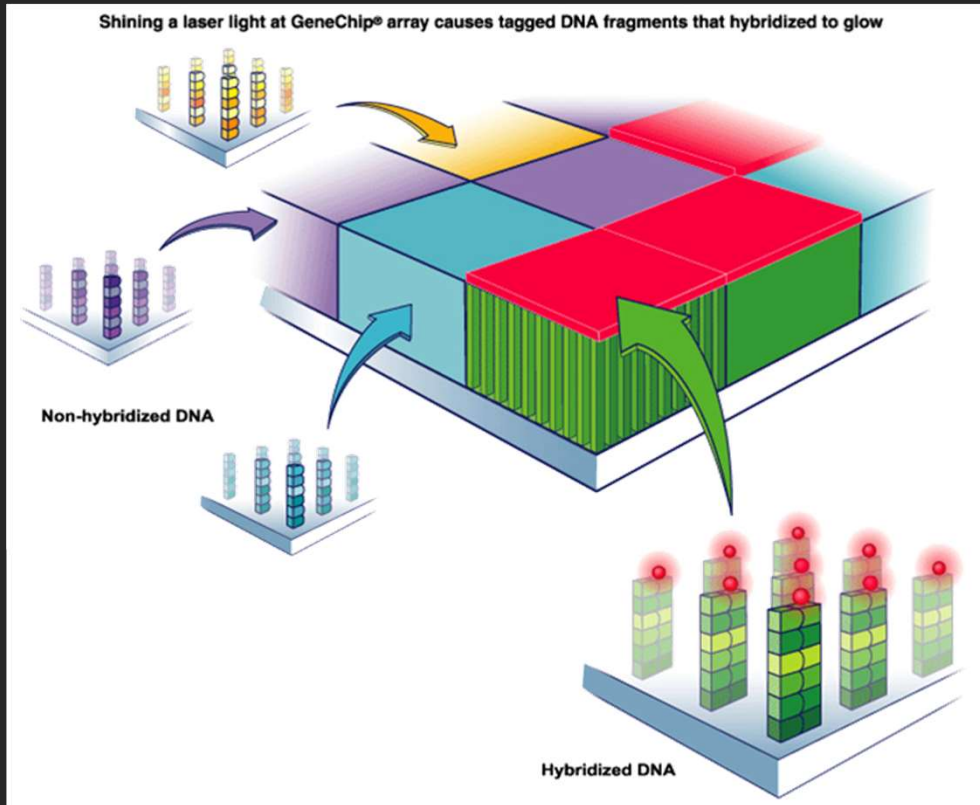
Microarray technology



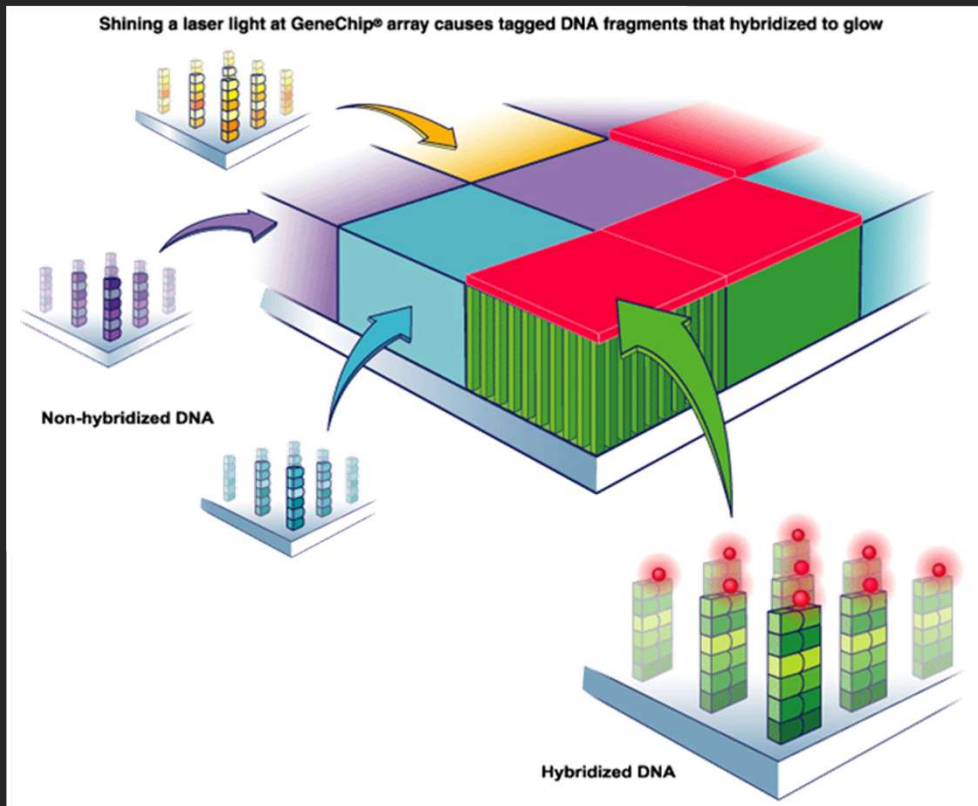
°cDNA microarray technology –
Patrick Brown - Stanford University
1981

Commercial technology by
Affymetrix.

Microarray technology



Microarray technology

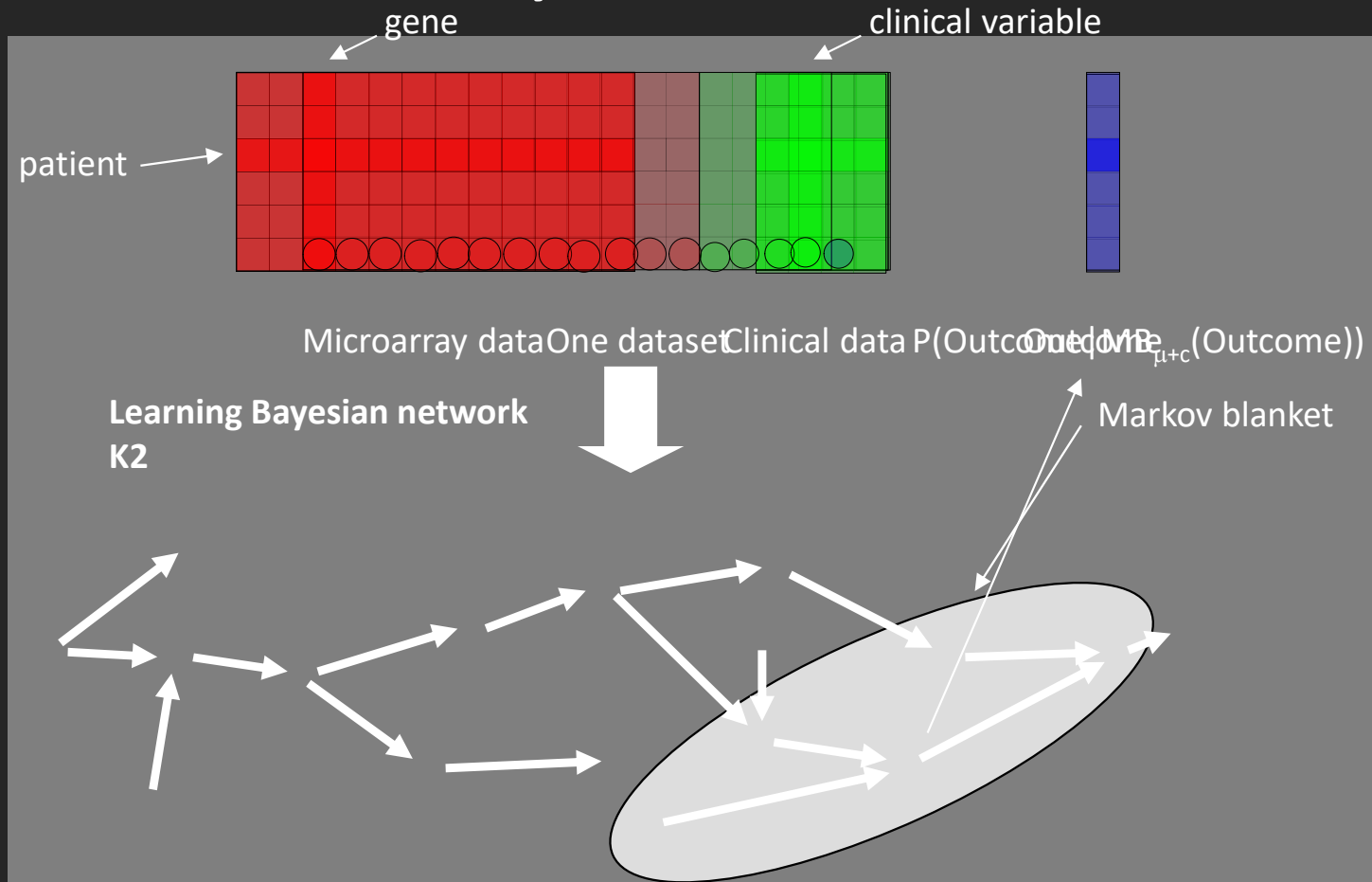


- High dimensional data
- No tools & methods to analyze them
- Accelerated the field of bioinformatics
- Supervised & unsupervised machine learning of RNA expression data became very popular

Bayesian data fusion

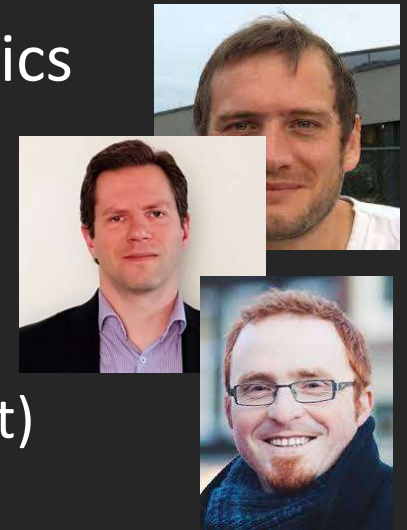
- Several “types” of data fusion
 - Early vs. late data fusion
 - Supervised vs. unsupervised
 - Multi-modal & multi-scale
- My thesis was focused on data fusion implementation with **Bayesian networks**

Early data fusion



Bioinformatics on the rise @ Biol/SISTA/ESAT

- Group of several students working on similar topics
 - Thomas Dhollander (Dsquare, Trendminer)
 - Tim Van Den Bulcke (#56, UA, J&J & Galapagos)
 - Karen Lemmens (#50, CropDesign, Bayer, BASF)
 - Raf Van de Plas (#62, Vanderbilt University & TU Delft)

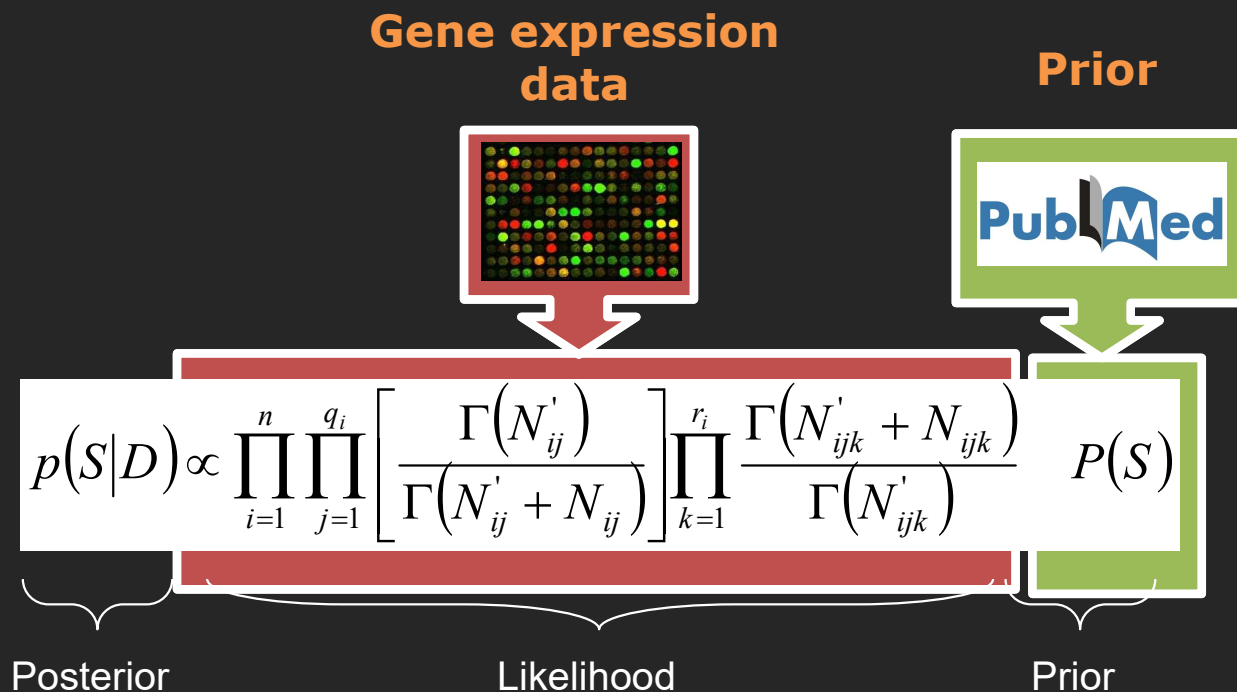


Bioinformatics on the rise @ Biol/SISTA/ESAT

- Group of several students working on similar topics
 - Thomas Dhollander (Dsquare, Trendminer)
 - Tim Van Den Bulcke (#56, UA, J&J & Galapagos)
 - Karen Lemmens (#50, CropDesign, Bayer, BASF)
 - Raf Van de Plas (#62, Vanderbilt University & TU Delft)
 - Peter Van Loo (#52, KUL, Francis Crick, MD Anderson)



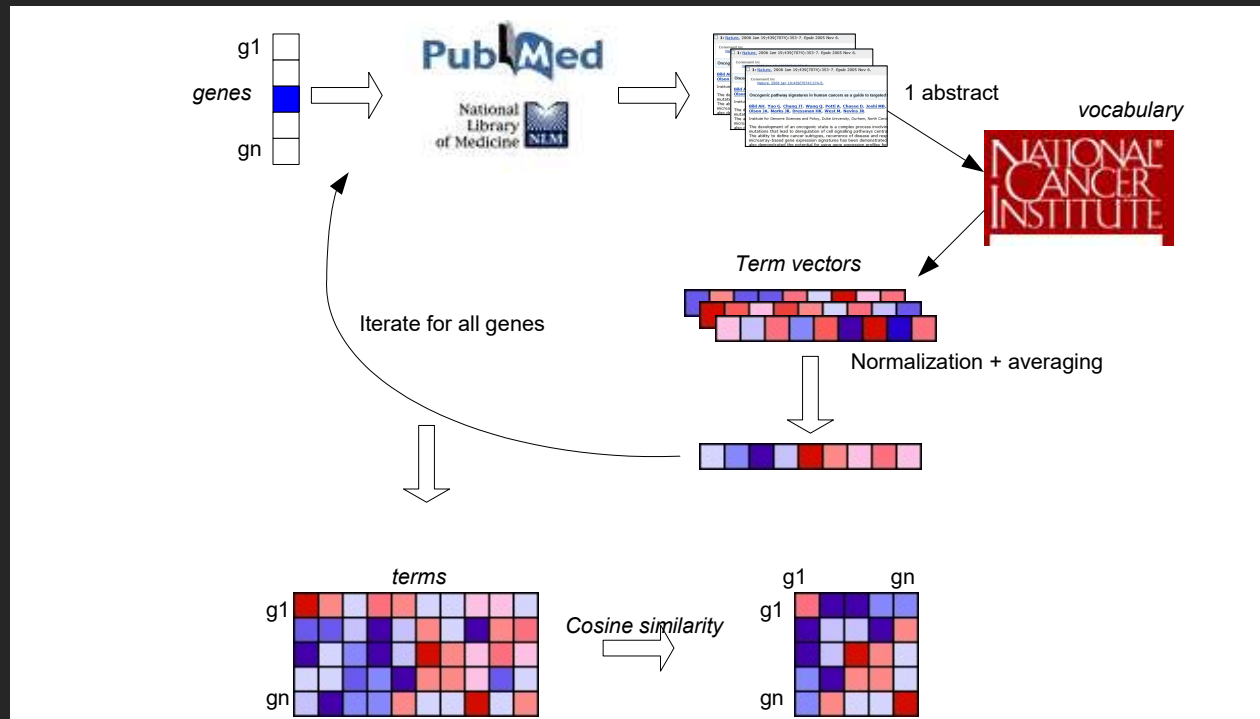
Integration of Bayes nets with Text Mining



with Steven Van Vooren (#57, Cartagena)

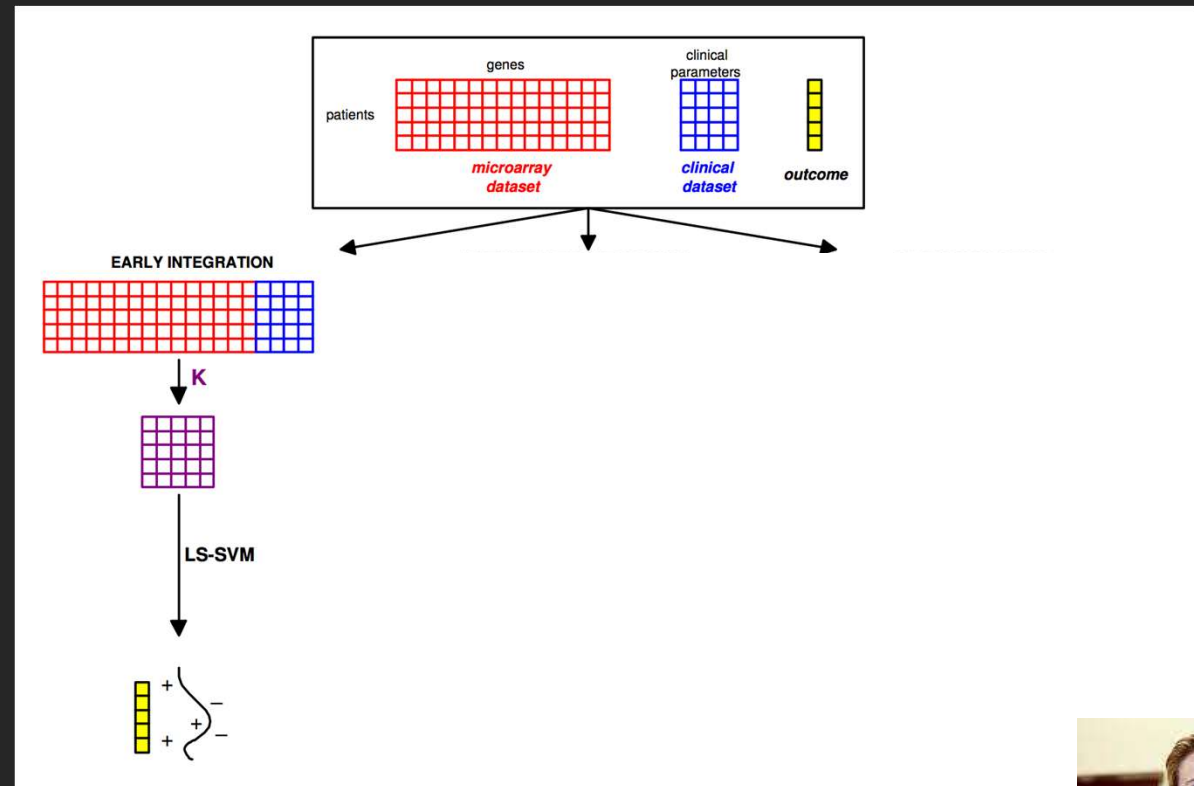


Structure prior



Alternative framework: Kernel Data Fusion

- Kernel data fusion as alternative to **Bayes** nets

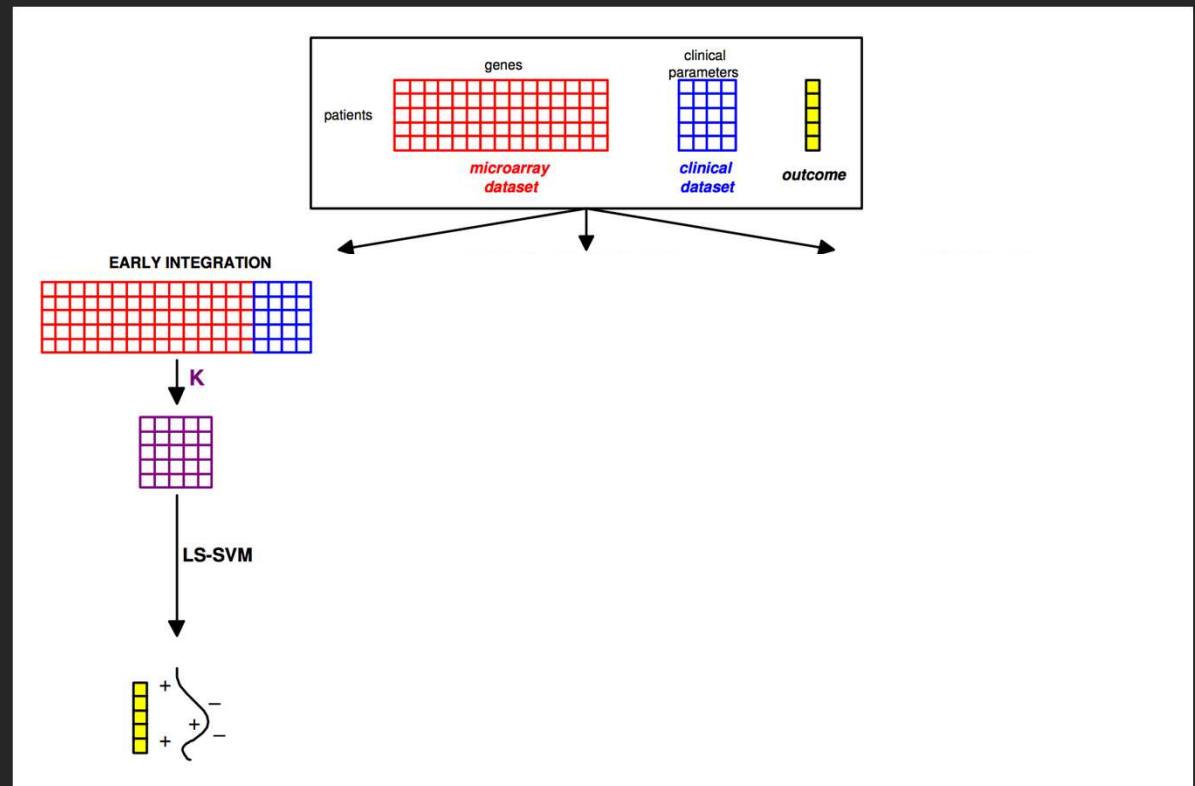


Anneleen Daemen (#60, UC Berkeley, Genentech)



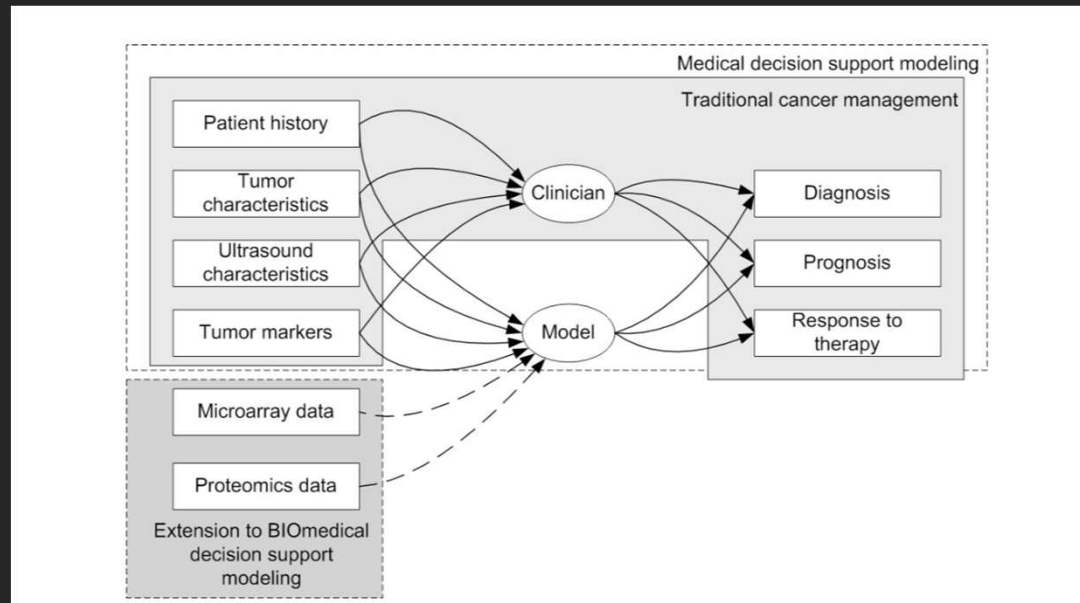
Early kernel fusion

- Kernel data fusion as alternative to **Bayes** nets
- Parallel implementation of early, intermediate & late integration



End of Phd

- PhD Defense December 2008
“Framework for Bayesian network integration of biomedical data”



Road to Stanford

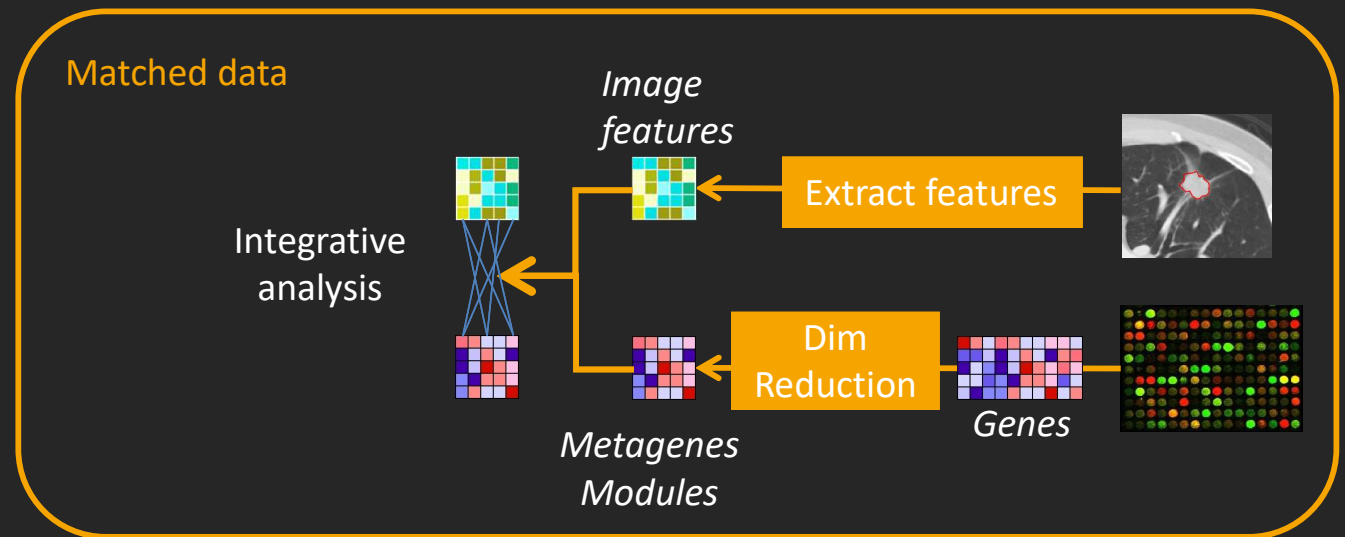
- After postdoc ~1 year with Bart
- Joined Stanford radiology dept in January 2010.
Support from Belgian American Educational Foundation (BAEF) & FWO



Journey @ Stanford

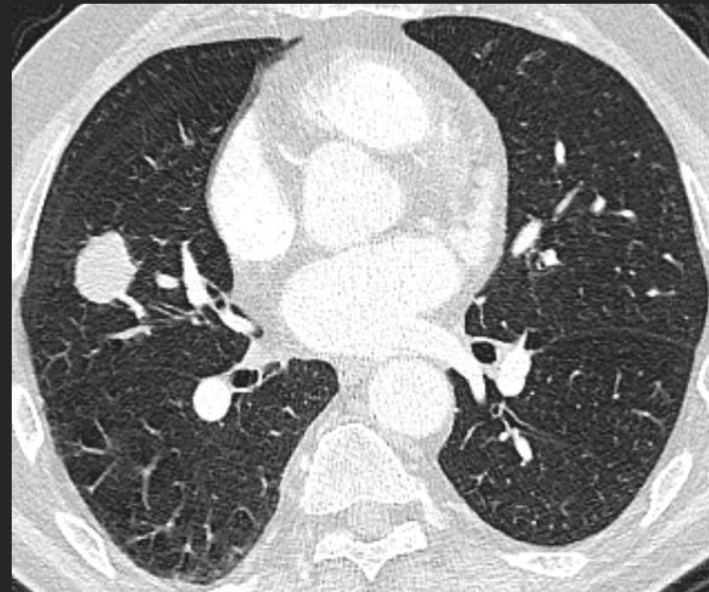
- Postdoc: 2010-2013 with Sylvia Plevritis – Stanford Radiology
- Staff: 2013-2015

Radiogenomics &
quantitative
imaging



Medical imaging

- CT, MRI data have complementary information
 - Size, location and morphology
- Part of diagnostic routine



Potential to describe anatomical, functional and physiological properties of patients

Faculty @ Stanford

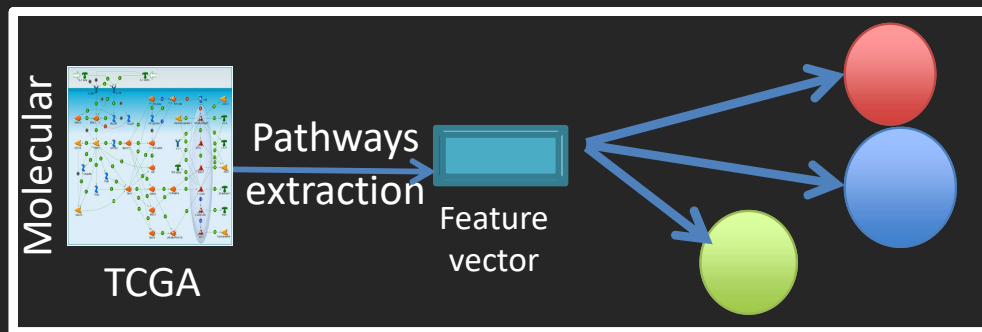
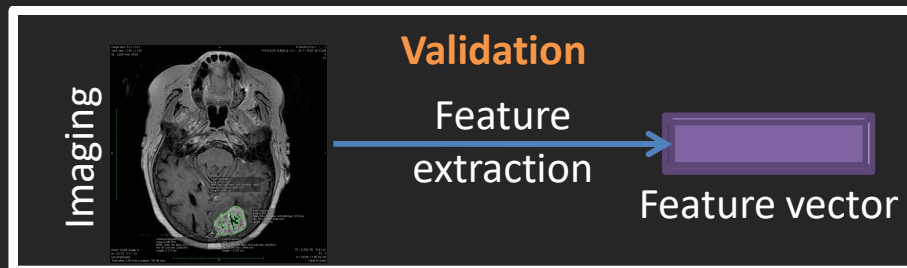
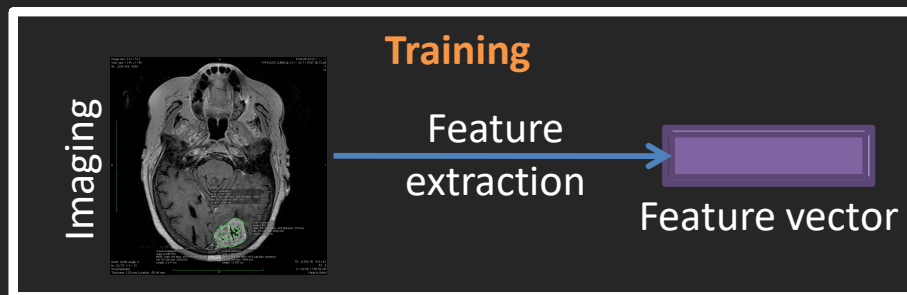
- Stanford Center for Biomedical Informatics Research @
Department of Medicine
 - Assistant Professor: 2015-2022
 - Associate Professor: 2022-now

First phd Student

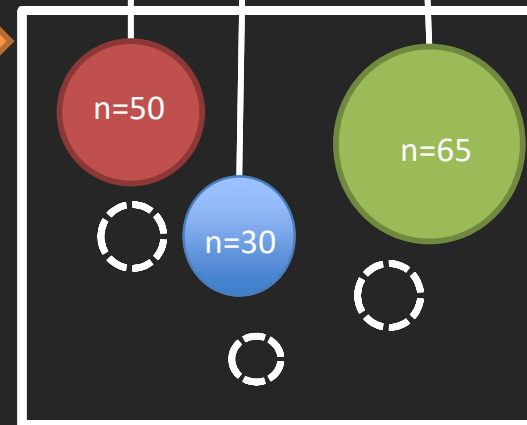
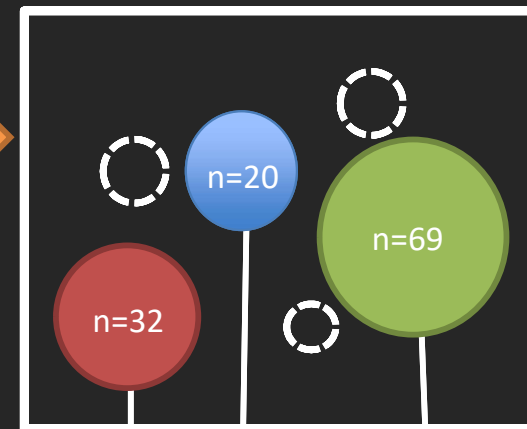
- Haruka Itakura (#1, Stanford faculty)
- Quantitative imaging as basis for subtype discovery



Subtype Discovery



Map subtypes to pathways



Validate subtypes



- Cluster 1:
 - Irregular shape
 - Many concave edges
- Cluster 2:
 - Spherical lesions
 - Regular edges
- Cluster 3:
 - Blurry edges
 - Rim enhancing

RESEARCH ARTICLE

CANCER

Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities

Haruka Itakura,¹ Achal S. Achrol,² Lex A. Mitchell,³ Joshua J. Loya,² Tiffany Liu,¹ Erick M. Westbroek,⁴ Abdullah H. Feroze,² Scott Rodriguez,² Sebastian Echegaray,⁵ Tej D. Azad,² Kristen W. Yeom,³ Sandy Napel,³ Daniel L. Rubin,^{1,3} Steven D. Chang,² Griffith R. Harsh IV,^{2*} Olivier Gevaert^{1*†}

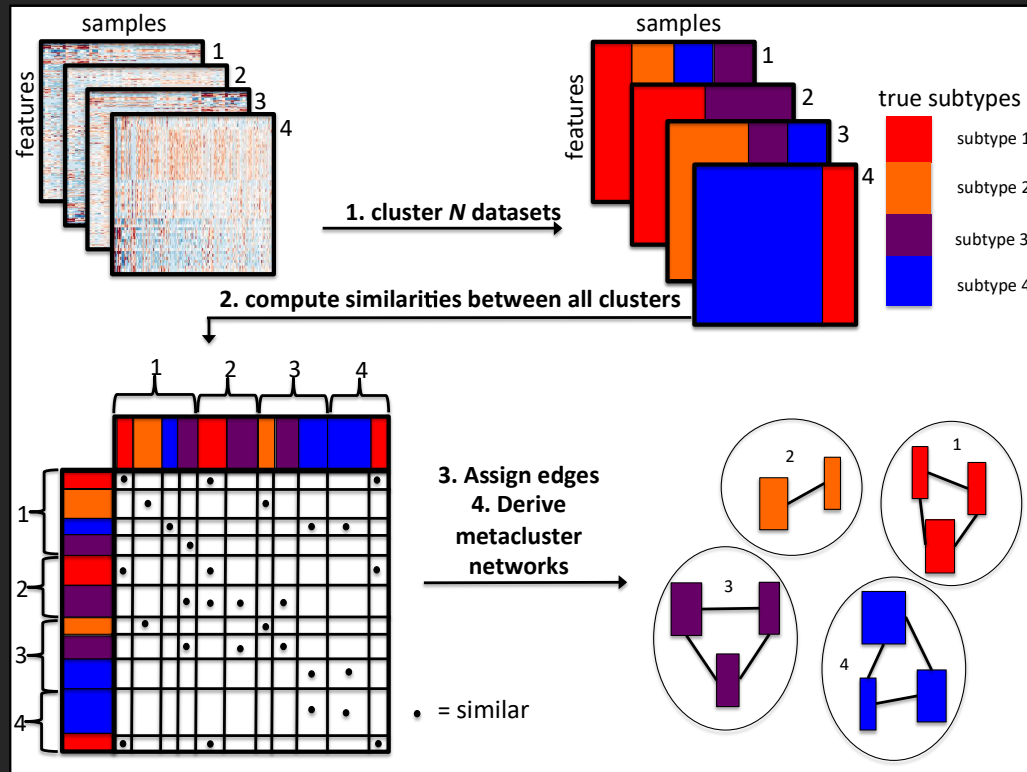
Glioblastoma (GBM) is the most common and highly lethal primary malignant brain tumor in adults. There is a dire need for easily accessible, noninvasive biomarkers that can delineate underlying molecular activities and predict response to therapy. To this end, we sought to identify subtypes of GBM, differentiated solely by quantitative magnetic resonance (MR) imaging features, that could be used for better management of GBM patients. Quantitative image features capturing the shape, texture, and edge sharpness of each lesion were extracted from MR images of 121 single-institution patients with de novo, solitary, unilateral GBM. Three distinct phenotypic “clusters” emerged in the development cohort using consensus clustering with 10,000 iterations on these image features. These three clusters—pre-multifocal, spherical, and rim-enhancing, names reflecting their image features—were validated in an independent cohort consisting of 144 multi-institution patients with similar tumor characteristics from The Cancer Genome Atlas (TCGA). Each cluster mapped to a unique set of molecular signaling pathways using pathway activity estimates derived from the analysis of TCGA tumor copy number and gene expression data with the PARADIGM (Pathway Recognition Algorithm Using Data Integration on Genomic Models) algorithm. Distinct pathways, such as c-Kit and FOXA, were enriched in each cluster, indicating differential molecular activities as determined by the image features. Each cluster also demonstrated differential probabilities of survival, indicating prognostic importance. Our imaging method offers a noninvasive approach to stratify GBM patients and also provides unique sets of molecular signatures to inform targeted therapy and personalized treatment of GBM.

Itakura et al. 2015 Science Translational Medicine



CoINcIDE: A framework for discovery of patient subtypes across multiple datasets

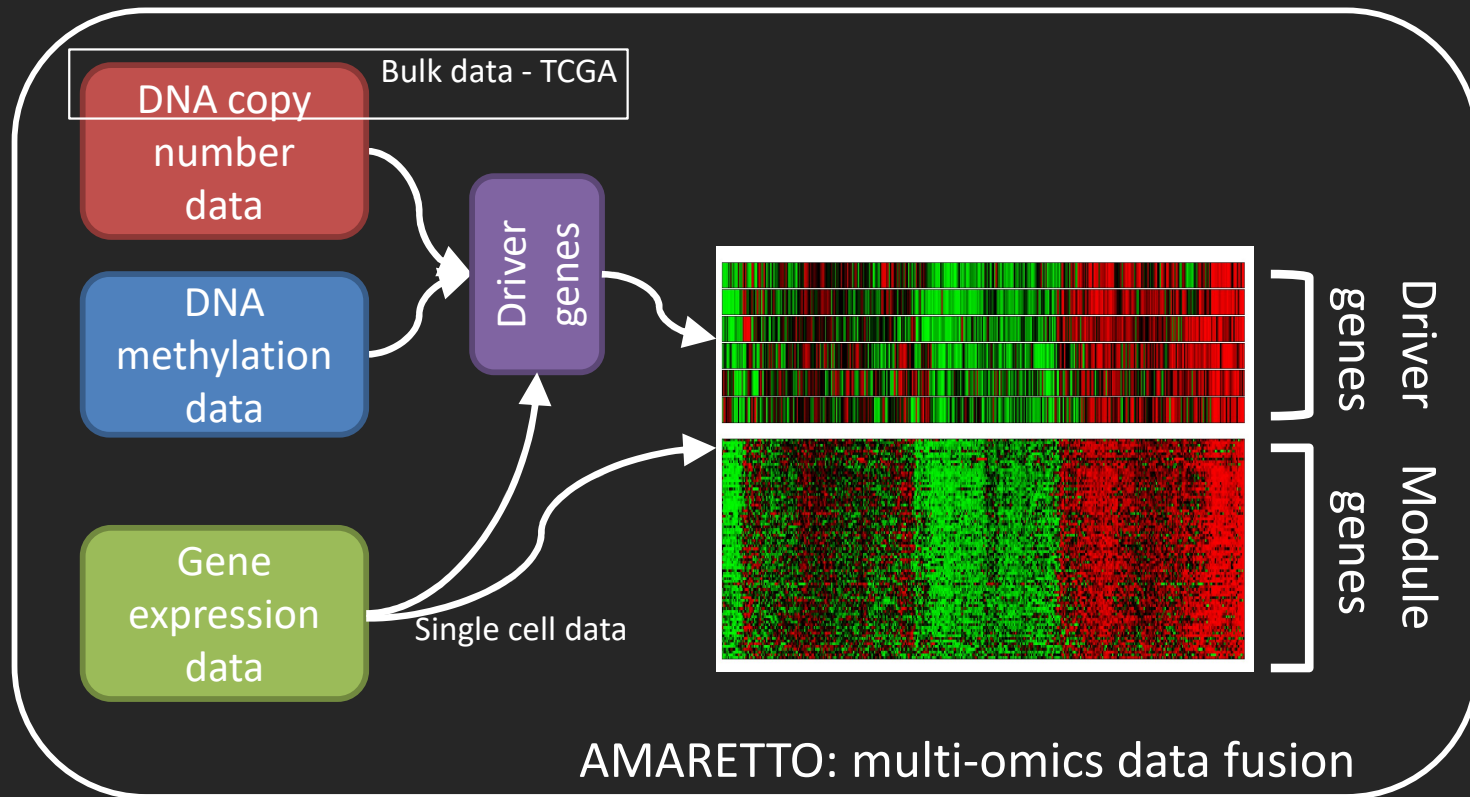
- Meta-analysis clustering
- Applied on set of ovarian cancer data sets
- Validation of subtypes across cohorts



Katie Planey (#2, CTO Mantra Bio)



AMARETTO: module networks

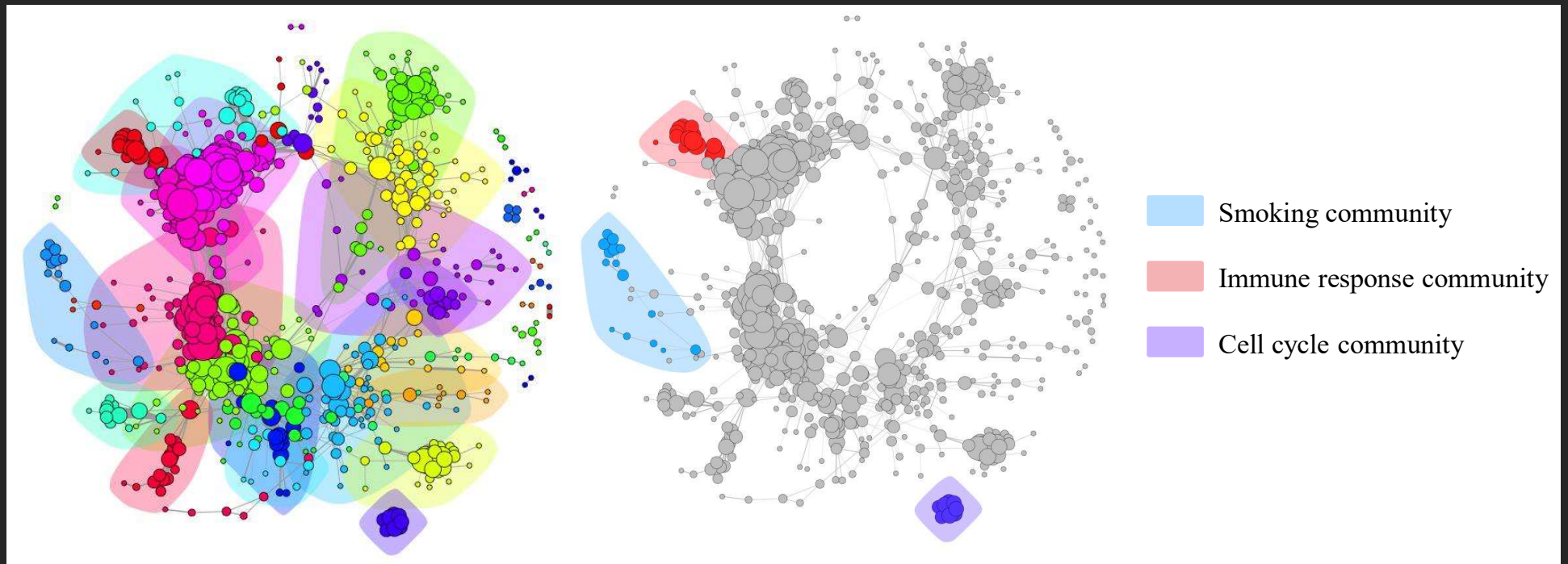


Gevaert O, et al. Interface Focus 2013
Manolakos et al. BMC Genomics 2014

Pancancer AMARETTO



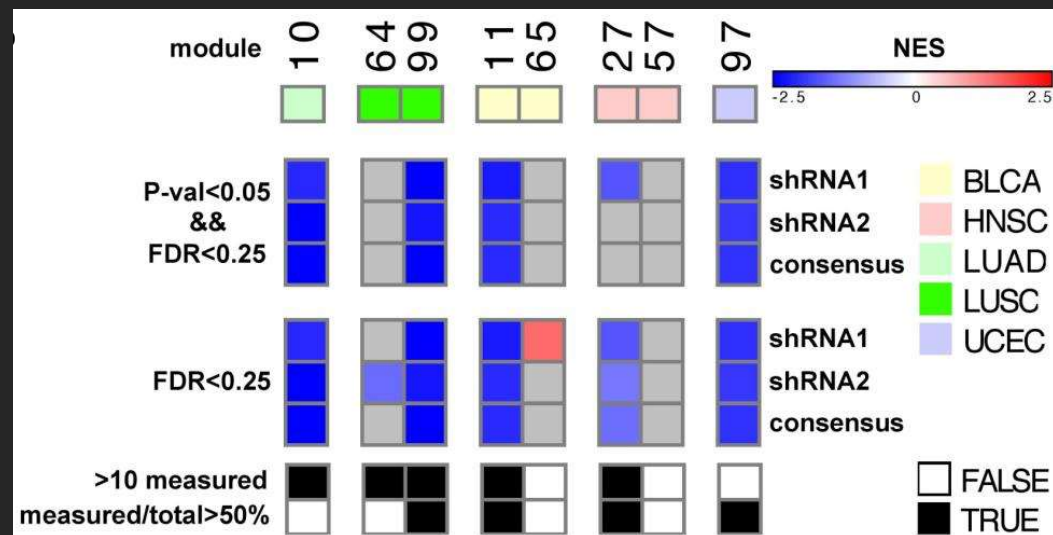
- with Nathalie Pochet (#43, Harvard & Broad institute)



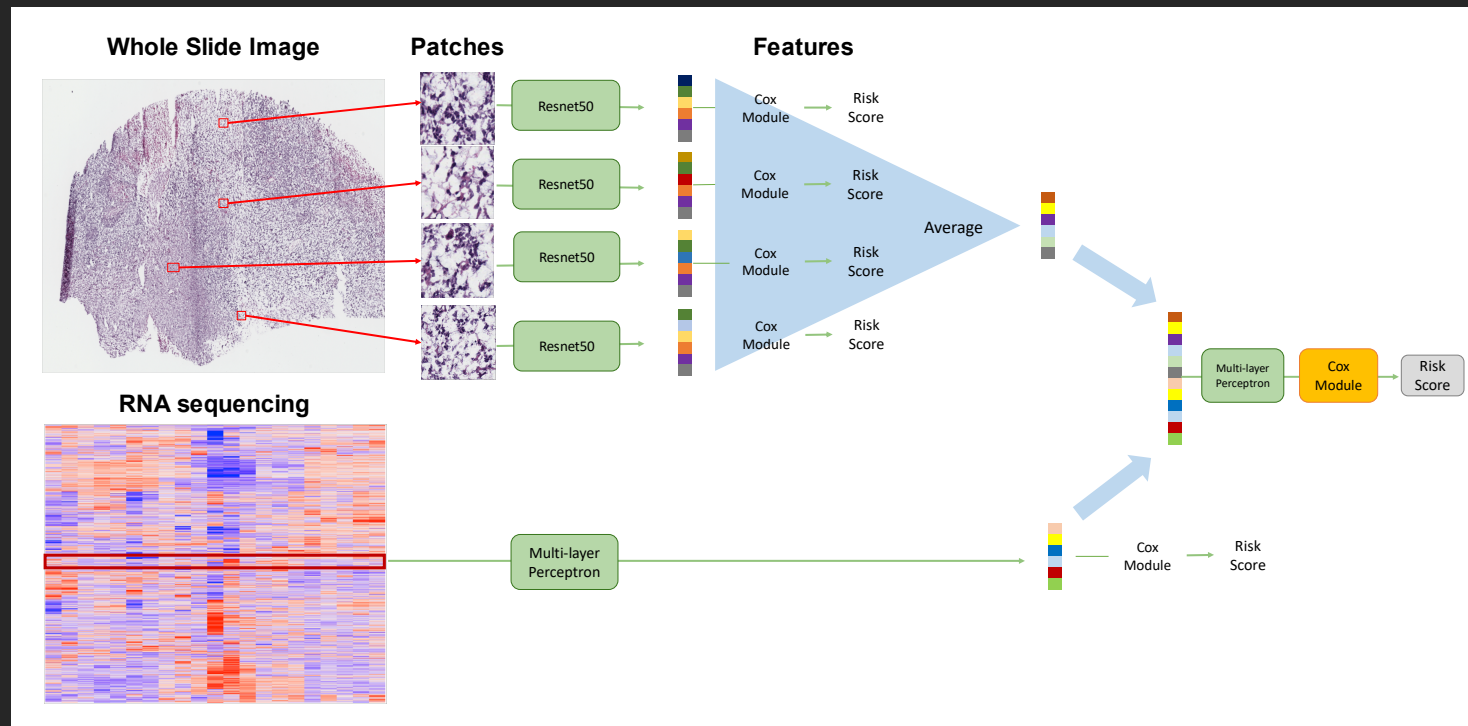
Champion et al. EBioMedicine, 2018

Validation of GPX2

- GPX2 is pancancer driver of smoking community
- LINCS project evaluates genetic perturbations and measures its effect on gene expression
- GPX2 is part of the LINCS study
- Measured in a lung cancer cell line



Data fusion frameworks with deep learning

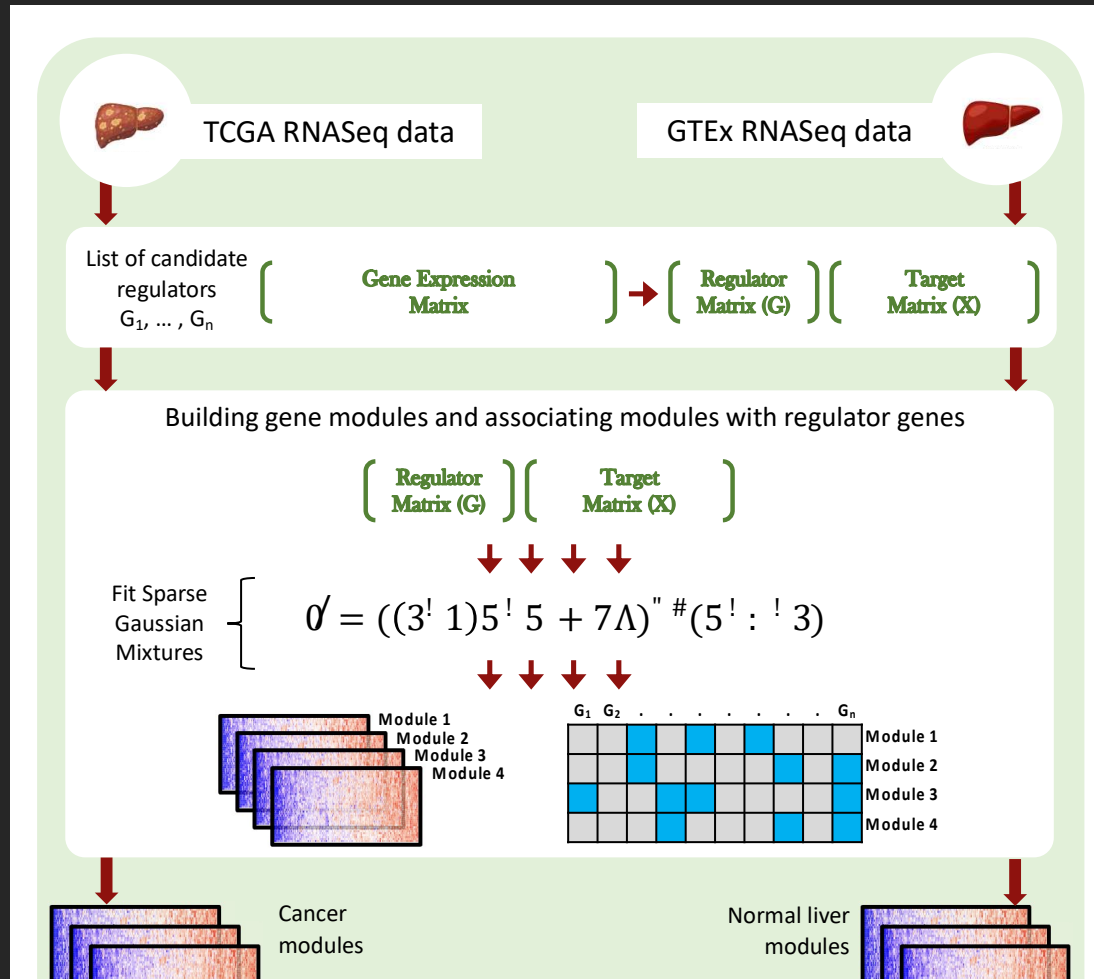


Lina Qiu (#3, Remesh Inc.)



SparseGMM

- Regulatory networks'
- Sparse Gaussian Mixture Model
- Applied on normal liver & liver cancer data
- Entropy of genes



Shaimaa Bakr (#4)

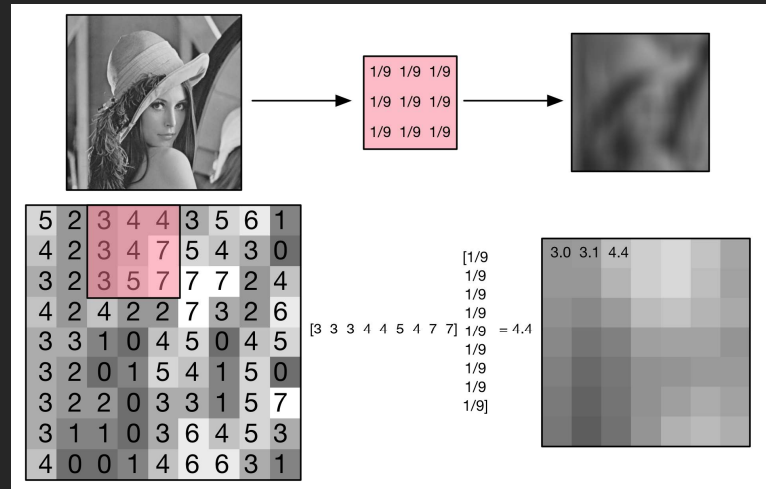
Visiting students @ Stanford

- Ongoing interactions with BDM bioinformatics graduates:
 - Spring 2020: visit Robin Van Daele – PhD student Dr. Tijn De Bie (#24).
 - Spring 2020: visit Tina Smets (#86, UgenTec, **currently last student**)
 - 2022: visit Marija Pizurica – PhD student Dr. Kathleen Marchal (Postdoc)

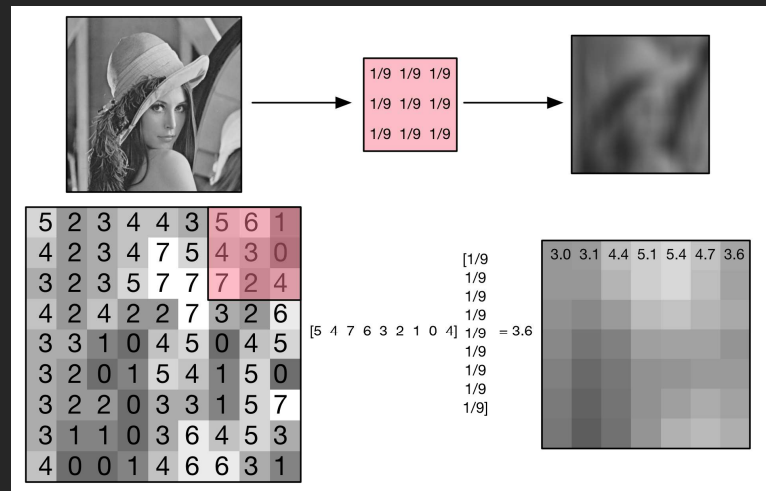
Deep learning & Convolutional Neural Networks

- Convolution operation done by using a kernel on an image
- Often used for blurring, sharpening, edge detection, and more

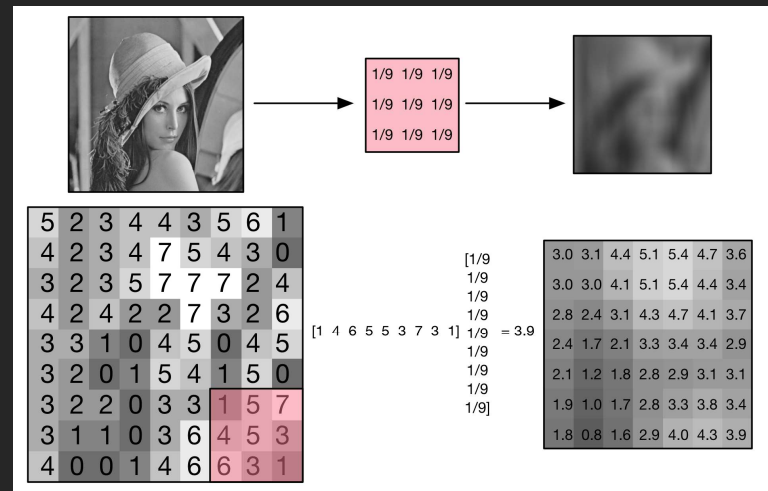
Convolutional Neural Network



Convolutional Neural Network



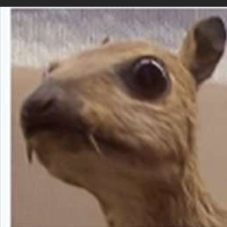
Convolutional Neural Network



Convolutions

Identity

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$



Edge detection

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$



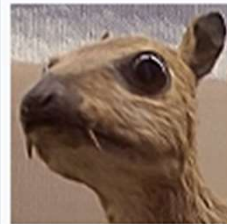
Sharpen

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$



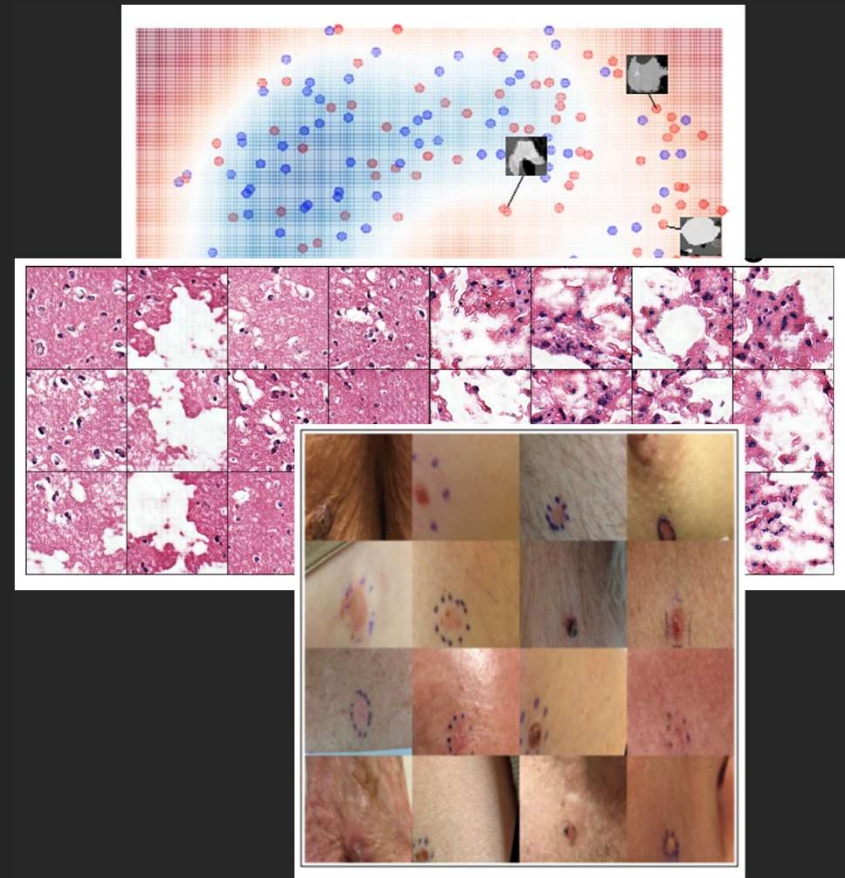
Gaussian blur

$$\frac{-1}{256} \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & -476 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}$$



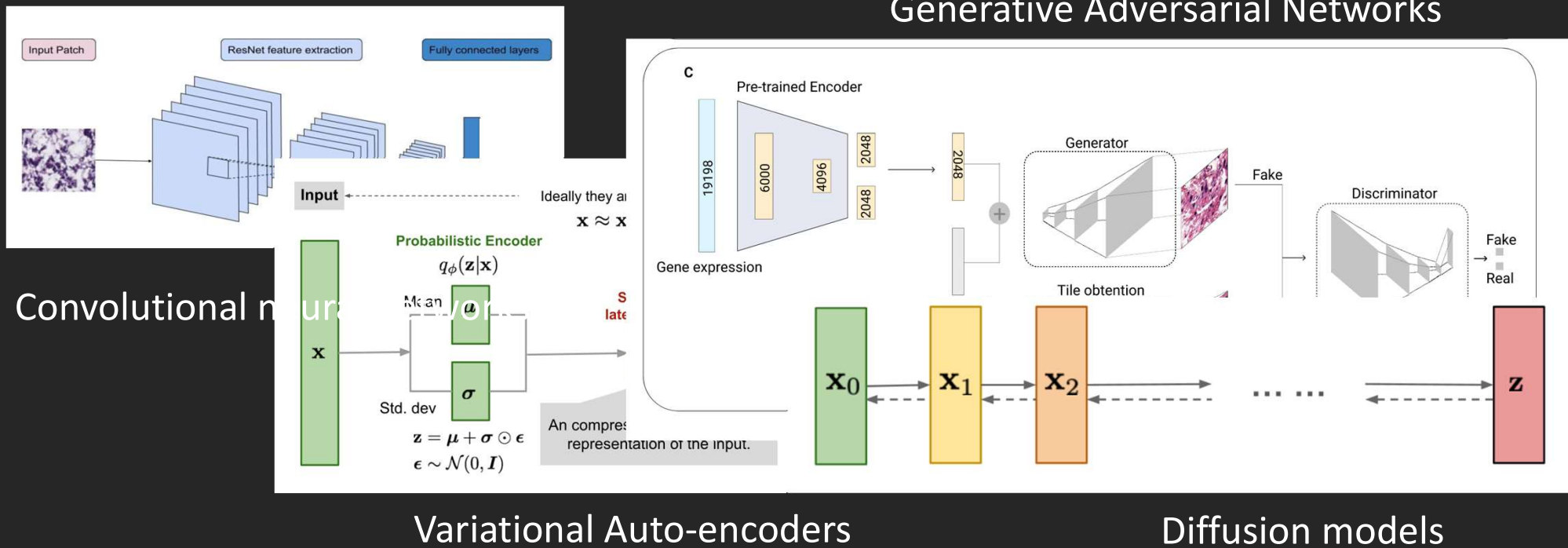
Deep learning

- Successfully used on medical images
 - Radiographic images (Mukherjee et al. 2021 Nat Mac Intel)
 - Digital pathology (Perez et al. In prep.)
 - Dermatology images (Sadee et al. In Prep)



Deep learning (r)evolution

- From CNNs, to VAE, GANs and Diffusion models.



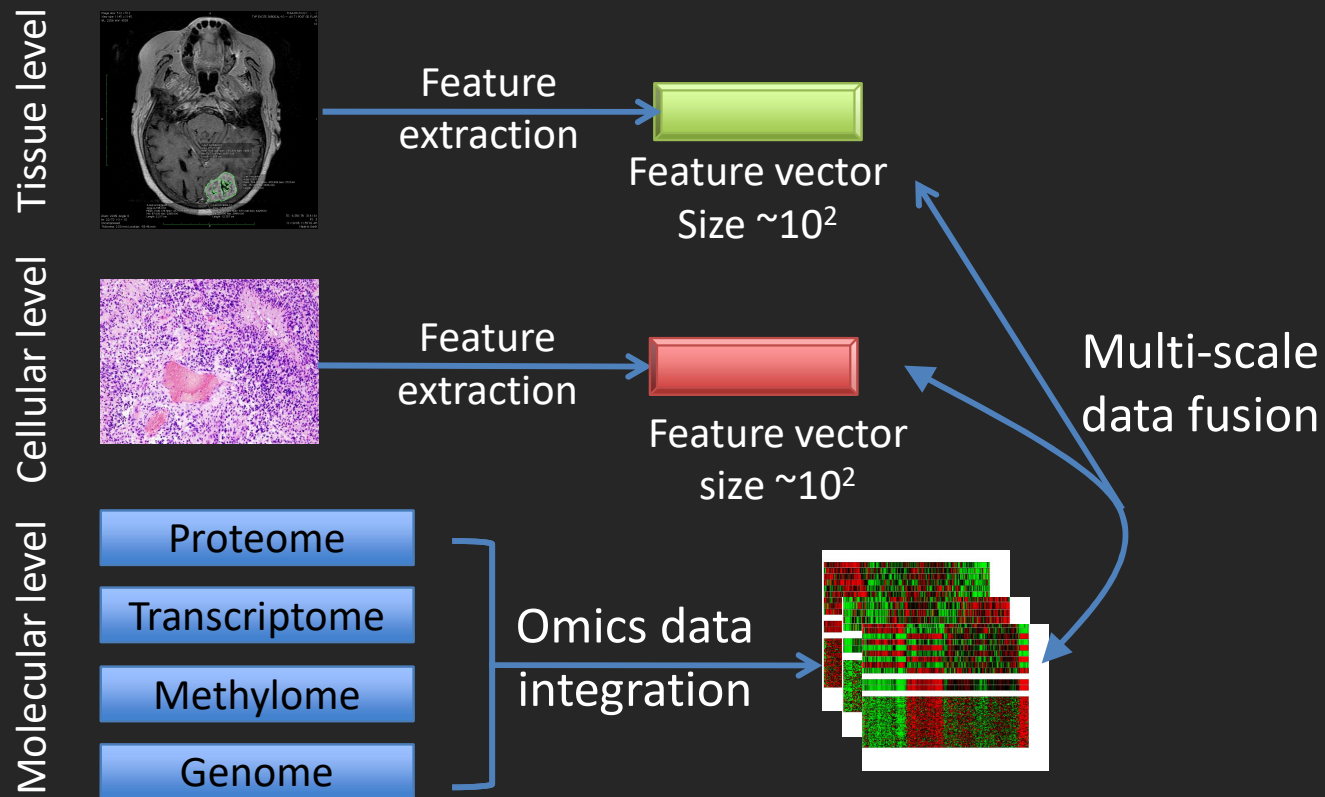
Biomedical data fusion & Multi-modal data



- Molecular data
 - Genome
 - Transcriptome
 - Methylome
 - Proteome
- Tissue data
 - Radiographic images
- Cellular data
 - Histopathology
- Clinical data
 - EHR
 - Clinical notes

Can we use multi-modal data for cancer patients to predict overall survival?

Multi modal data fusion



<https://github.com/gevaertlab/MultimodalPrognosis>

Summary

- Entire cohort of bioinformatics students have been trained by **Bart De Moor**
 - Ended up in key academic, spin-off and industry positions.
- Bioinformatics, biomedical data has evolved tremendously in the past ~20 years
- Amount of data biomedical data keeps exploding
 - Molecular Single cell
 - Routinely collected imaging data
 - Spatial technologies: spatial transcriptomics & proteomics

Summary

- Our research themes
 - Data Representation
 - Data fusion
 - Visualization & Interpretation
- More info: <http://gevaertlab.stanford.edu>
- Disclaimers
 - Apologies for anyone who I did not mention
 - Rankings of BDM PhDs are based on the “Mathematics Genealogy Project”