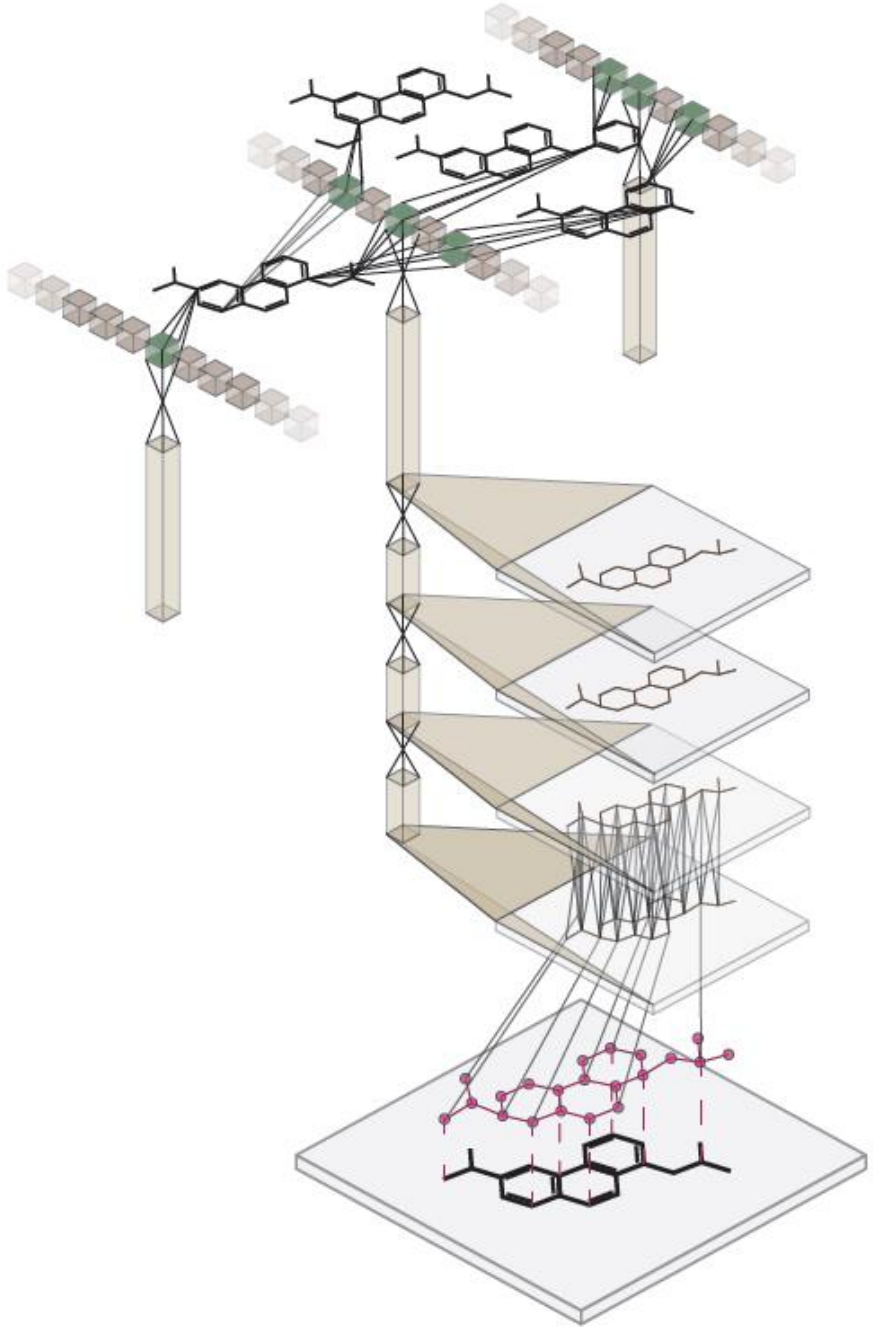
Deep Learning and Chemistry



Robert Glen
University of Cambridge
Imperial College London

The Hype

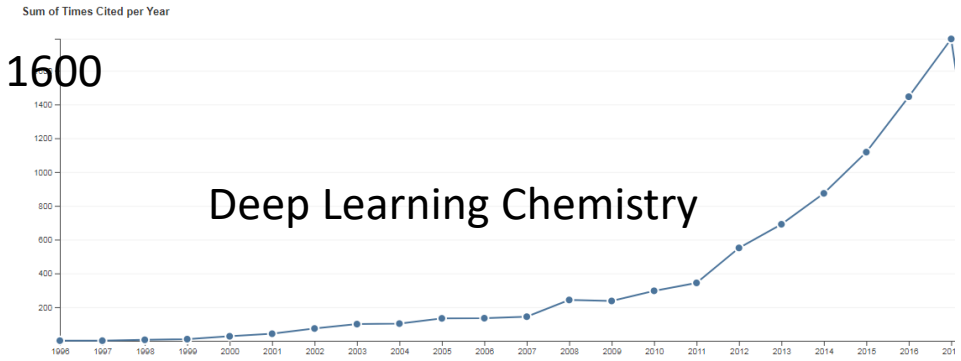
Have you heard that Deep Learning can do everything?

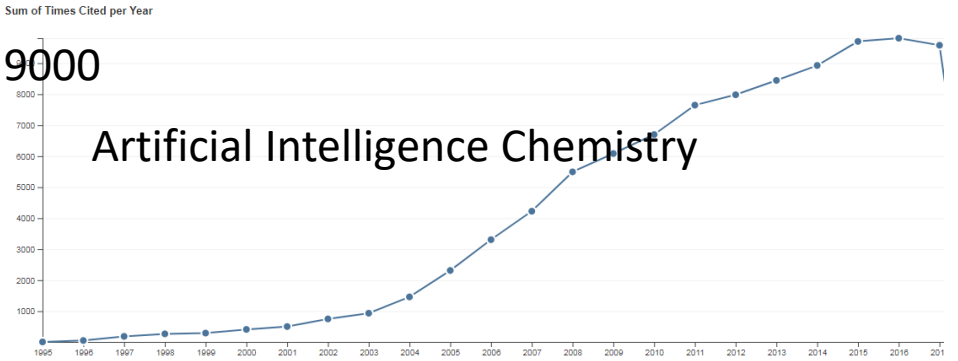
62 startups in AI/Chemistry segment

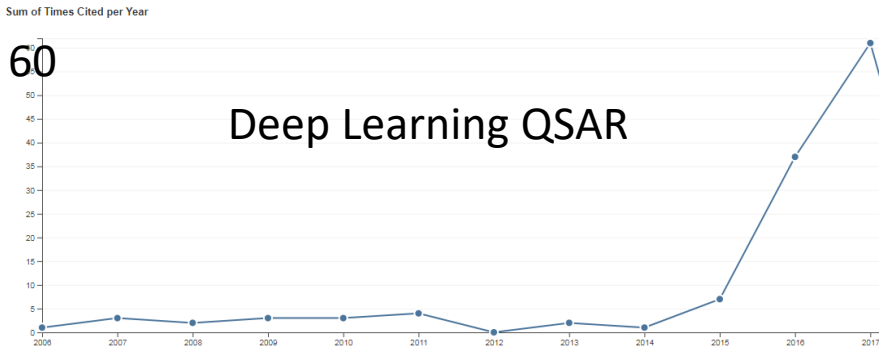
- Aggregate and Synthesize Information
- Understand Mechanisms of Disease
- Repurpose Existing Drugs
- Generate Novel Drug Candidates
- Validate Drug Candidates
- Design Drugs
- Design Preclinical Experiments
- Run Preclinical Experiments
- Design Clinical Trials
- Recruit for Clinical Trials
- Optimize Clinical Trials
- Publish Data
- One of my students – Sam Cooper has started Phenomics AI (Canada)

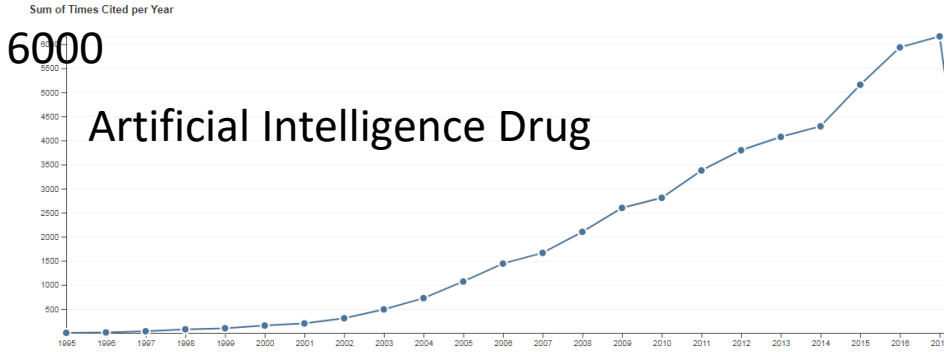
https://blog.benchsci.com/startups-using-artificial-intelligence-in-drug-discovery

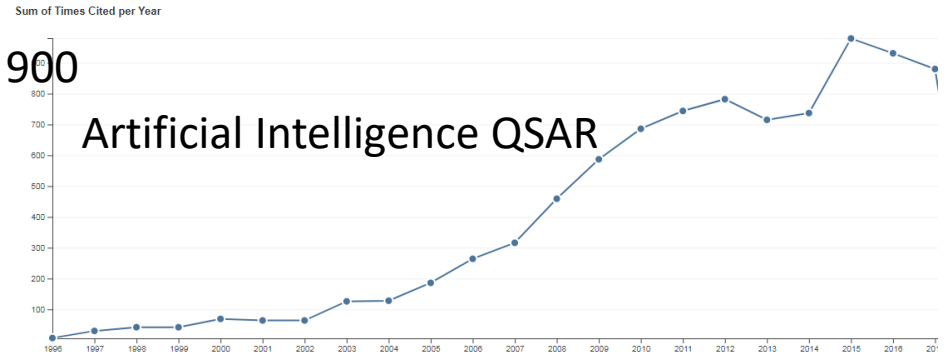
Deep Learning/Artificial Intelligence – citations (WoS, 1995-2018)



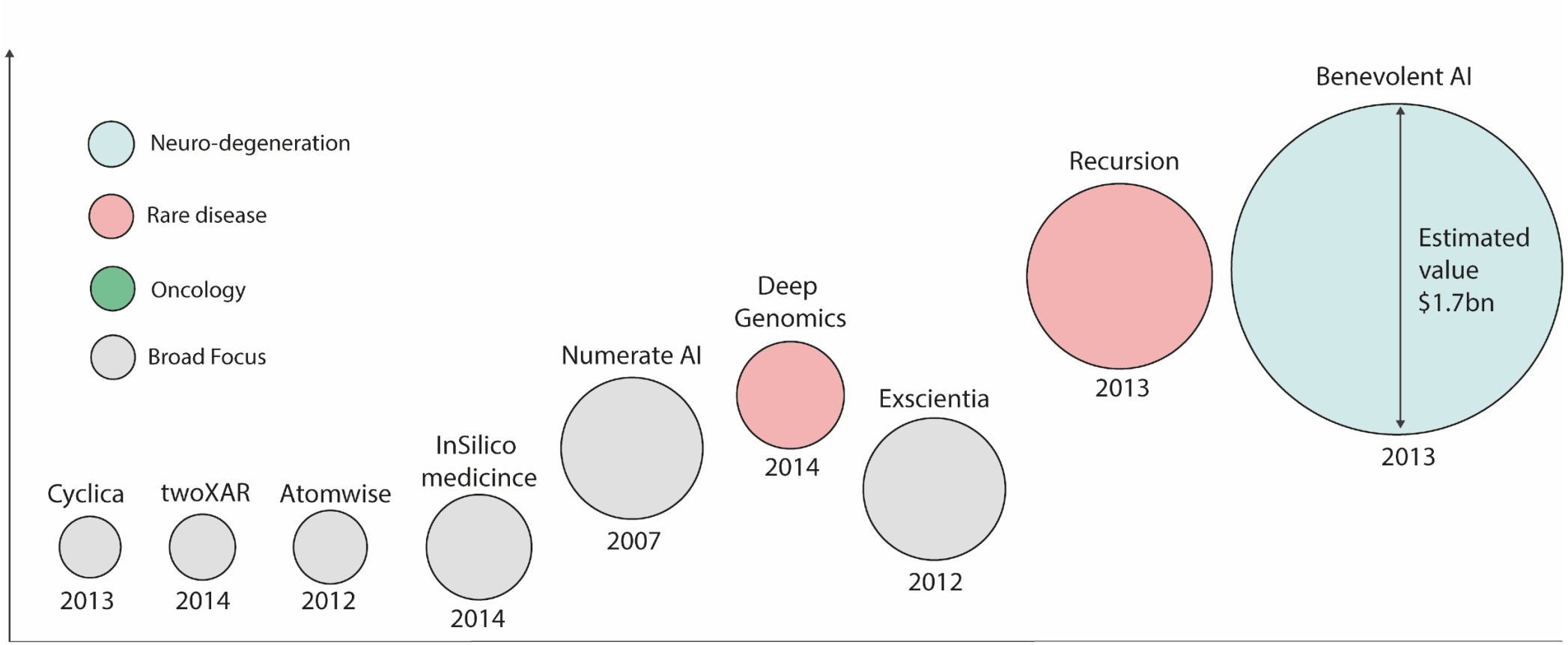






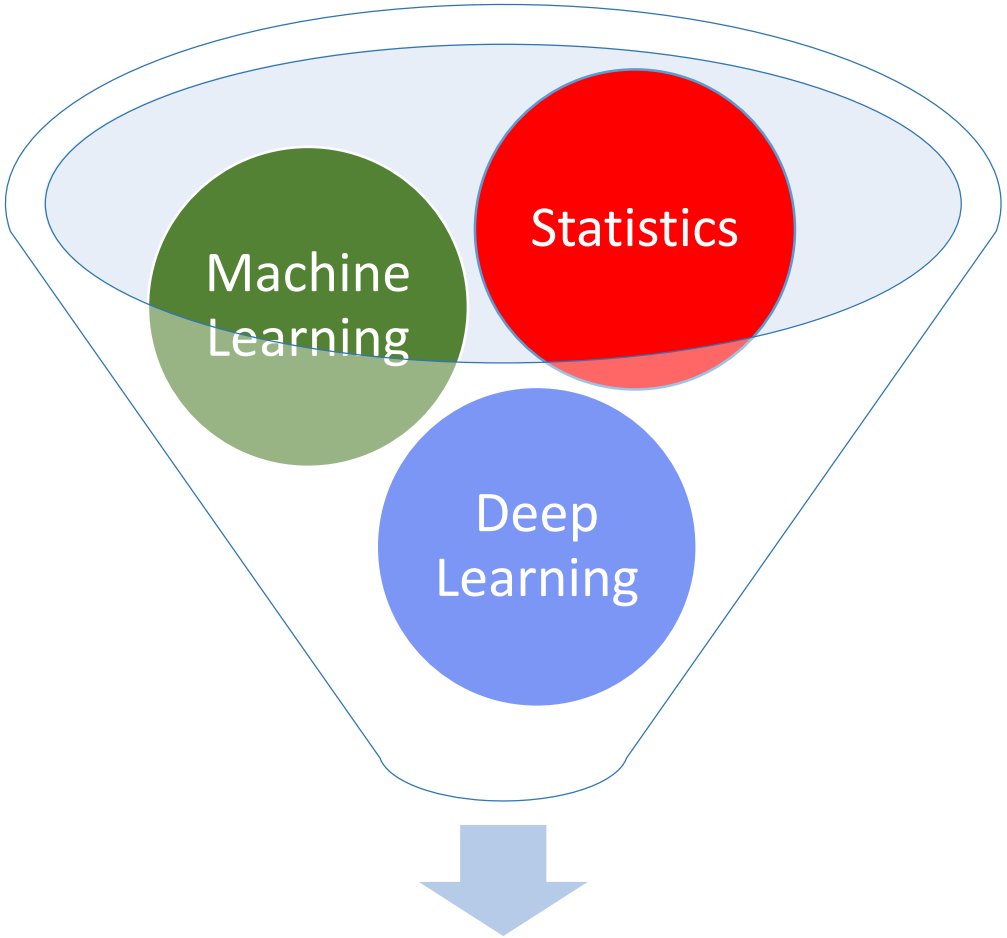


Some companies in AI/Drug discovery



What's the difference between Statistics, Machine Learning and 'Deep Learning'

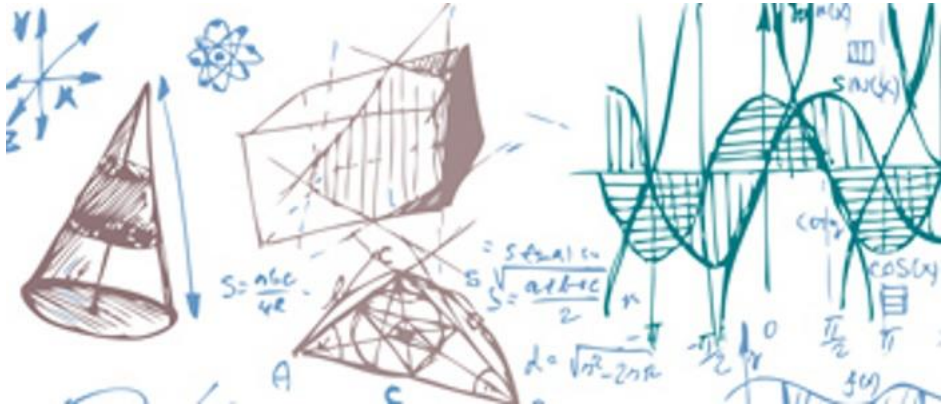
- There is significant overlap in the methodologies.
  - Quite often we mix and match to obtain good results – there is a place for each approach
  - I'd look at it as a spectrum of tools that can be combined to get the best results for particular problems



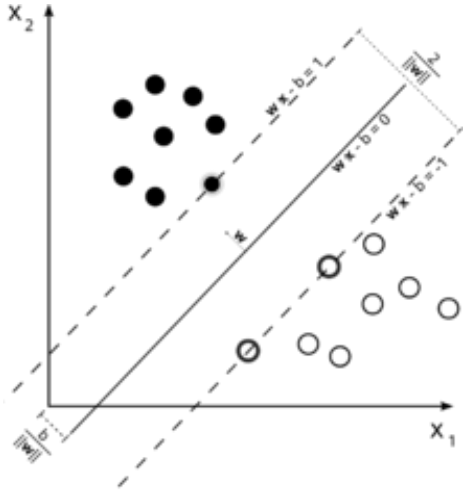
Understanding/Prediction

What's the difference between Statistics, Machine Learning and 'Deep Learning'

- Statistics has a long history in data analysis – many methods to analyse, predict and model data.
  - Based mostly on analysis of variance, expected distributions and mathematical formalism. Mature.

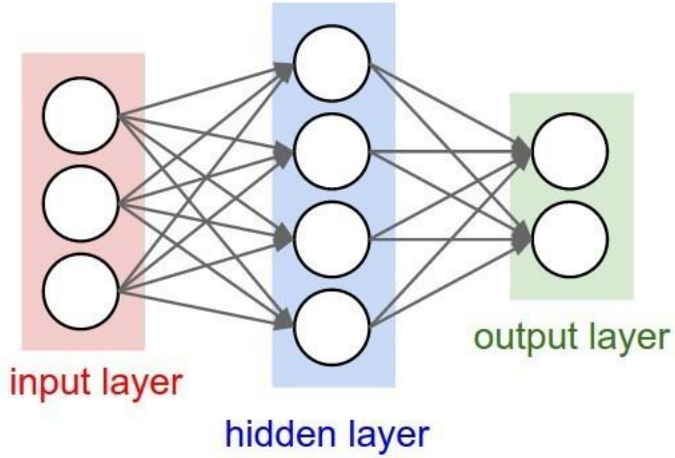


- Machine Learning grew out of a desire to cope with larger unstructured, disjointed data.
  - More of a black box (or even a complete black box). Suffers from 'the **curse of dimensionality' – need to reduce the descriptor space**

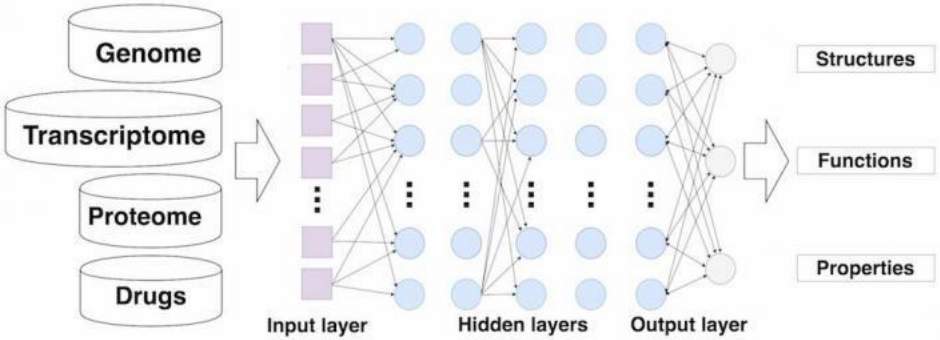


What's the difference between Statistics, Machine Learning and 'Deep Learning'

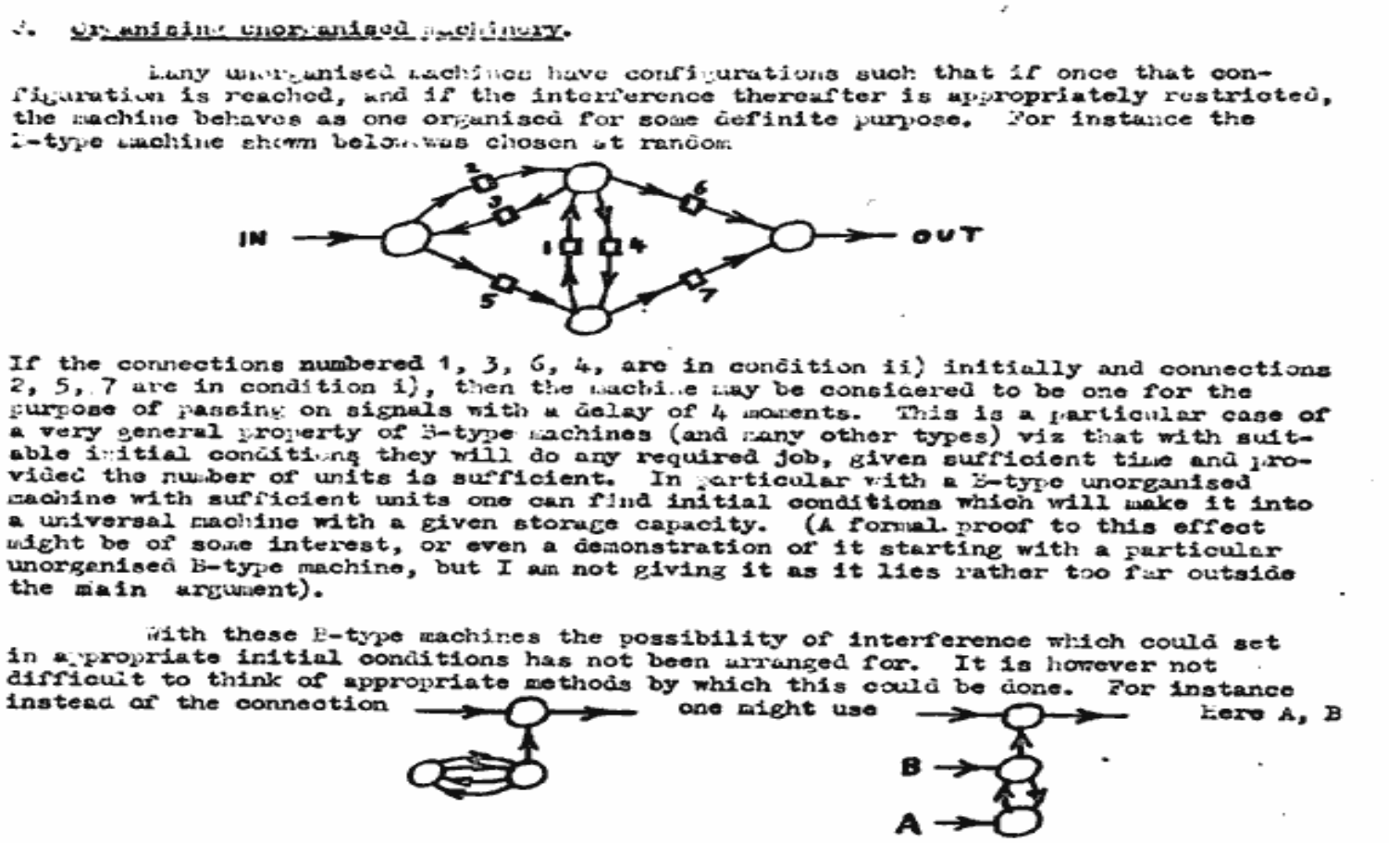
- Neural Networks were an initial approach to mimic brain architecture.
  - However, training 'deeper' feedforward neural networks tends to yield worse results (both in training/test error) than shallow ones (with 1 or 2 hidden layers).



- 'Deep' architectures for learning were recently developed – in 2006 some breakthroughs were made
  - Humans organize their ideas and concepts hierarchically.
    - Brain architecture is 'deep' , we can copy that
  - Humans first learn simpler concepts and then compose them to represent more abstract ones.
    - Pre-training can significantly improve prediction. Deep NNs abstract layers of information at different levels of abstraction



Alan Turing (1948) – idea of an ‘e-machine’ – first idea of a neural computer



9. Organising unorganised machinery.

Many unorganised machines have configurations such that if once that configuration is reached, and if the interference thereafter is appropriately restricted, the machine behaves as one organised for some definite purpose. For instance the P-type machine shown below was chosen at random

If the connections numbered 1, 3, 6, 4, are in condition ii) initially and connections 2, 5, 7 are in condition i), then the machine may be considered to be one for the purpose of passing on signals with a delay of 4 moments. This is a particular case of a very general property of B-type machines (and many other types) viz that with suitable initial conditions they will do any required job, given sufficient time and provided the number of units is sufficient. In particular with a B-type unorganised machine with sufficient units one can find initial conditions which will make it into a universal machine with a given storage capacity. (A formal proof to this effect might be of some interest, or even a demonstration of it starting with a particular unorganised B-type machine, but I am not giving it as it lies rather too far outside the main argument).

With these B-type machines the possibility of interference which could set in appropriate initial conditions has not been arranged for. It is however not difficult to think of appropriate methods by which this could be done. For instance instead of the connection one might use Here A, B

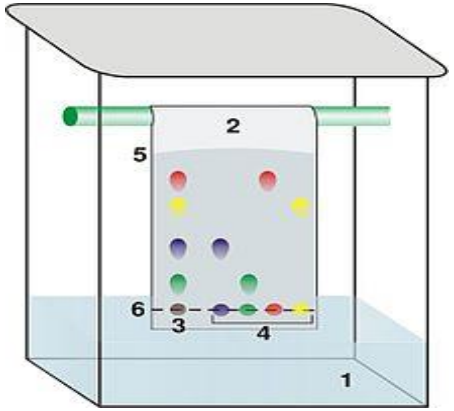
Written while Turing was working for the National Physical Laboratory in London, the paper did not meet with his employers' approval. Sir Charles Darwin, the director of the Laboratory, called it a 'schoolboy essay' and wrote to Turing complaining about its 'smudgy' appearance. In reality this far-sighted paper was the first outline of a ‘neural network’, but sadly Turing never published it.

http://www.alanturing.net/turing_archive/pages/Reference%20Articles/connectionism/Turing%27s%20neural%20networks.htm

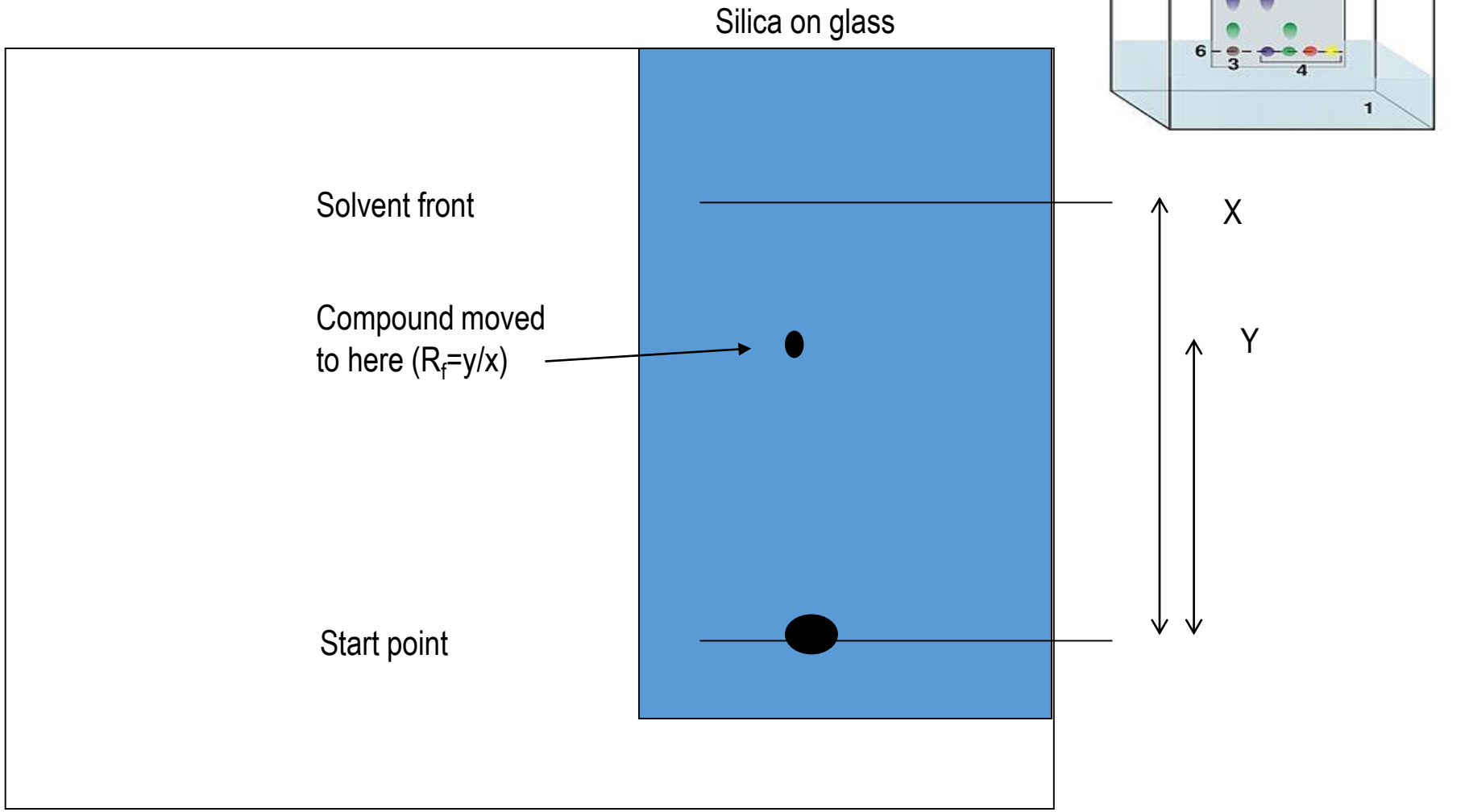
In 1943, neurophysiologist Warren McCulloch and mathematician Walter Pitts wrote a paper on how neurons might work. In order to describe how neurons in the brain might work, they modelled a simple neural network using electrical circuits.

In 1949, Donald Hebb wrote *The Organization of Behavior*, a work which pointed out the fact that neural pathways are strengthened each time they are used, a concept fundamentally essential to the ways in which humans learn. If two nerves fire at the same time, he argued, the connection between them is enhanced. https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/History/history1.html

Example of a Neural Network
Predicting TLC (Thin Layer Chromatography)

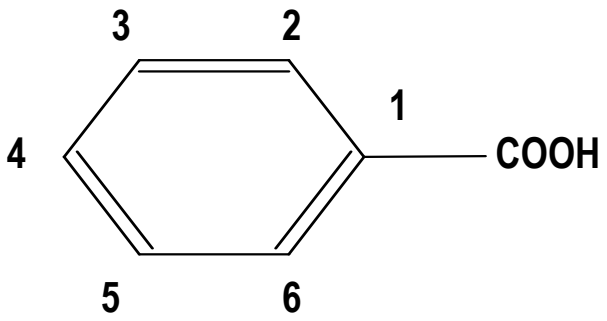


- Compounds move up the plate depending on the solvent, their properties etc.
- We can predict the R_f's (retention times) using details of the molecules and the solvent.
- Separate mixtures, identify compounds etc.



Data

- 22 substituted benzoic acids



1	4-F	8	4-F,3-CF_3	15	2-OH
2	3-F	9	4-F,2-CF_3	16	3-OH,6-OH
3	2-F	10	4-CH_3	17	2-OH,6-OH
4	CF_3	11	2-CH_3	18	2-OH,3-OH
5	3-CF_3	12	3-CH_3	19	2-OH,5-OH
6	4-CF_3	13	4-NH_2	20	2-OH,4-OH
7	2-F,4-CF_3	14	H	21	3-OH,4-OH
				22	2-COOH

Data

- 2 solvent systems
- 6 - mixtures

1	Acetonitrile - Water	30 : 70
2	Acetonitrile - Water	40 - 60
3	Acetonitrile - Water	50 - 50
4	MeOH - Water	40 - 60
5	MeOH - Water	50 - 50
6	MeOH - Water	60 - 40

- 22 compounds x 6 mixtures = 132 experiments

Compound	MeOH:H_2O:TFA 60:40:1	CH_3CN:H_2O:TFA 40:60:1	THF:H_2O:TFA 25:75:1
4-F-benzoic acid	3.4	3.08	13.5
3-F-benzoic acid	3.9	3.09	14.3
2-F-benzoic acid	2.0	2.01	6.9
2-trifluoromethylbenzoic acid	3.8	4.83	19.2
3-trifluoromethylbenzoic acid	9.8	7.70	–
4-trifluoromethylbenzoic acid	11.3	14.47	–
2-F,4-trifluoromethylbenzoic acid	9.6	8.24	–
4-F,3-trifluoromethylbenzoic acid	12.8	9.54	–
4-F,2-trifluoromethylbenzoic acid	6.4	6.68	–
2,5-dichlorobenzoic acid	8.7	6.72	–
salicylic acid	4.1	3.50	19.8
benzoic acid	4.0	2.38	8.4
2-methoxybenzoic acid	1.5	1.64	2.8
4-methoxybenzoic acid	3.0	2.46	7.7
3-methoxybenzoic acid	3.4	2.74	9.4
3-toluic acid	5.5	4.16	14.2
2-toluic acid	5.4	4.01	13.4
4-toluic acid	4.9	3.92	15.1
L(+)-mandelic acid	0.8	0.72	2.6
4-aminobenzoic acid	0.4	0.42	1.9
2-chlorobenzoic acid	3.1	3.09	11.7
4-chlorobenzoic acid	8.0	5.54	28.0
4-(trifluoromethyl)mandelic acid	3.4	2.91	20.7
phthalic acid	0.7	0.45	2.6
3-(p-hydroxyphenyl)propionic acid	1.0	0.90	5.0
probenicid	14.5	17.58	–
4-(aminomethyl)benzoic acid	1.2	1.17	4.1
3,5-dichlorobenzoic acid	–	13.20	–
2,4,6-trihydroxybenzoic acid	0.3	0.35	4.3
2,6-dichlorobenzoic acid	2.9	3.99	16.2
3-hydroxybenzoic acid	1.0	0.71	5.5
3-chlorobenzoic acid	8.0	5.26	25.4
3,4,5-trimethylbenzoic acid	2.0	1.80	3.8
2,4-dichlorobenzoic acid	11.4	7.88	–
2,4,6-trihydroxybenzoic acid	0.4	0.36	4.1
4-hydroxybenzoic acid	0.7	0.50	4.5
4-ethylbenzoic acid	10.8	7.27	23.7
2,3,4-trimethoxybenzoic acid	1.6	1.92	1.7
2,4,6-trimethoxybenzoic acid	9.1	6.91	22.2
3,5-dihydroxybenzoic acid	0.2	0.32	2.9
2,3-dihydroxybenzoic acid	1.3	0.96	7.9
2,6-dihydroxybenzoic acid	0.8	0.79	3.8
2,5-dihydroxybenzoic acid	0.8	0.78	7.5
2,4-dihydroxybenzoic acid	1.1	0.95	10.7
3,4-dihydroxybenzoic acid	0.3	0.33	3.1
2,4-chlorophenoxy-2-methylpropionic acid	15.0	11.66	–
4-acetoamidophenol	0.2	0.32	0.9

Measurements

• Molecular properties were calculated for each of the molecules and tabulated in a spreadsheet (tlcdata.xls) e.g.

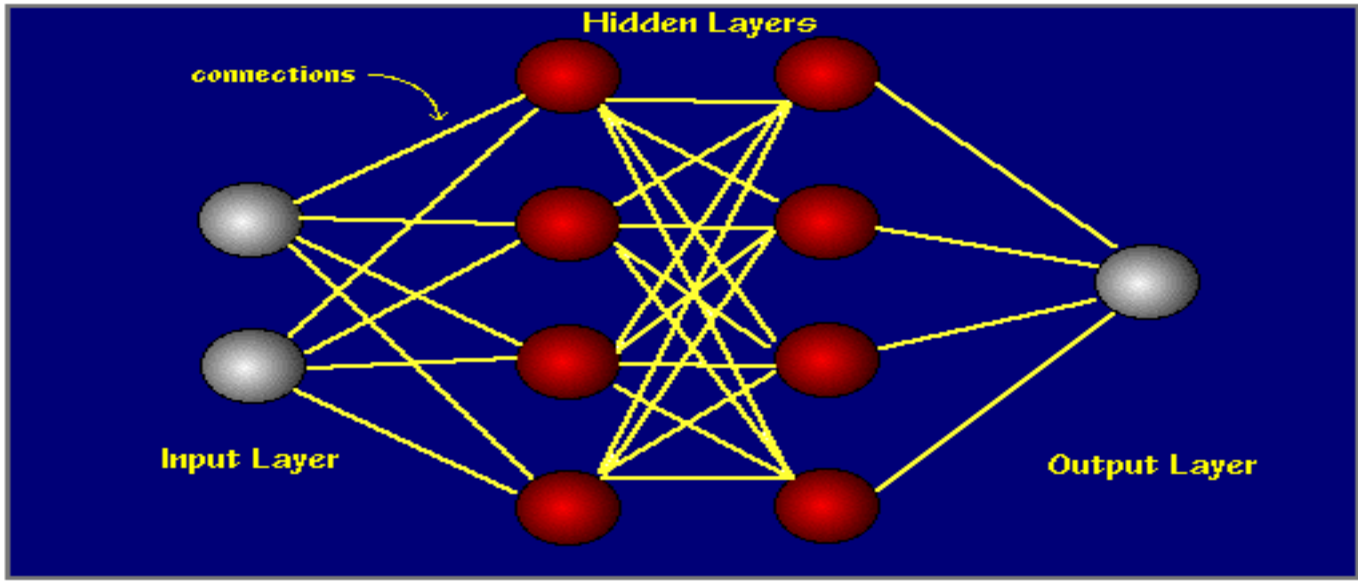
No. compound number
Cpd name of compound
Solvent water and acetonitrile/methanol
Rf retention time
Rm (log (1-Rf)/Rf))
S_Area surface area of molecule in A^2
clogp calculated partition coefficient octanol/water
volume molecular volume in A^3
MPolar polarizability of the molecule cm^{-25}
dipole dipole moment of the molecule (Debye)
dipsol dipole moment of the solvent (%solv1+%sol2)*100 Debye
PolSol polarizability of of the solvent (%pol1+%pol2)/100 Debye
Ovality: how removed from sperical

$$Ovality = S / \left[\frac{4}{3} \pi \left(\frac{3}{4} V \pi \right)^{2/3} \right]$$

water dipole is also given, 2.75Debye

Neural network

- Simulates the way that neurons are interconnected
- 'learns' by adjusting the connection weights between nodes taking an input set of parameters and attempting to fit the output measurements
- New data can then be entered and using the 'learned' model -> predict



This network has a 2:4:4:1 topology

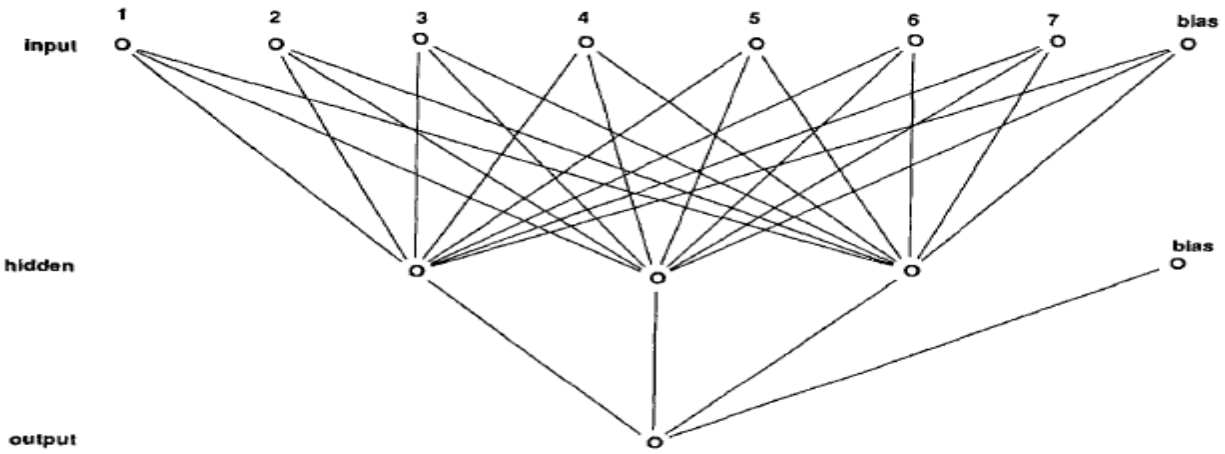
Like neurons, the connections are made when a threshold value is attained.

Use 'back propagation of errors' to adjust the connections

http://en.wikipedia.org/wiki/Backpropagation
http://en.wikipedia.org/wiki/Artificial_neural_network

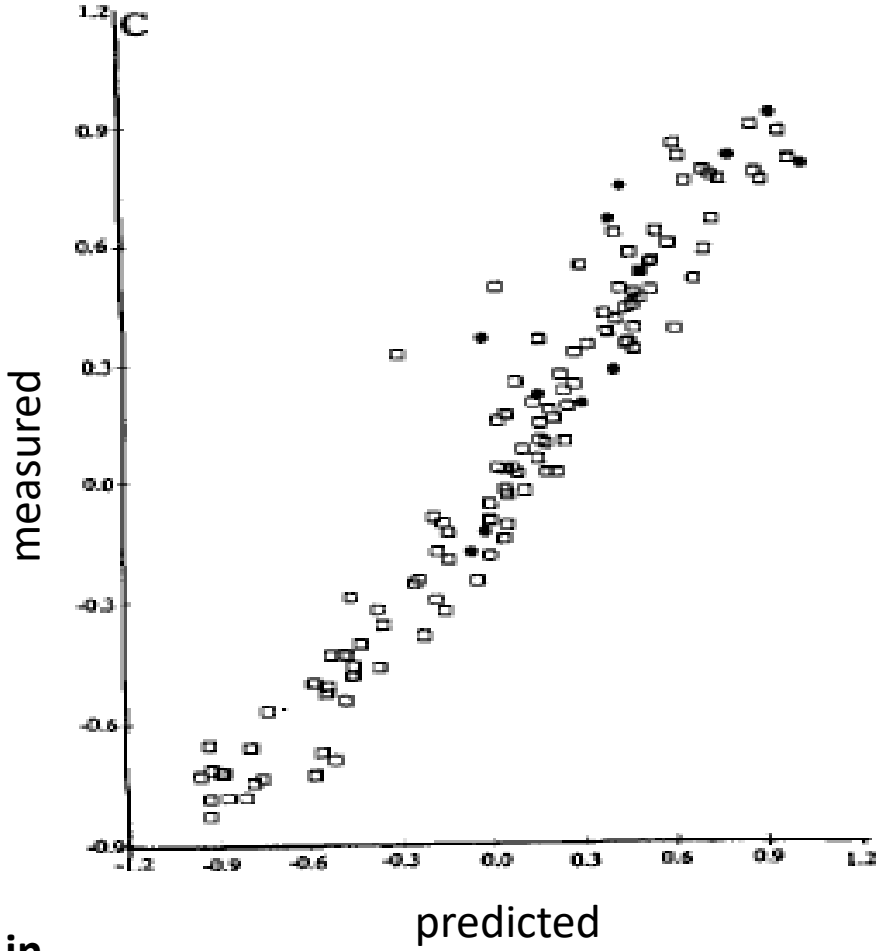
TLC Neural Network and plot of measured vs Predicted results

Table III Neurone-interconnection weights for the 7:3:1 network after 1000 cycles of training.



	I1 QON	I2 CLOGP	I3 DIP	I4 DIPMOM	I5 ESDL1	I6 CMR	I7 POLSOL	BIAS
H1	−0.139	1.217	0.612	0.022	−0.434	0.229	−0.282	0.065
H2	1.82	−3.616	−2.072	1.699	0.866	0.4033	0.153	0.054
H3	−1.402	−2.137	−2.302	1.255	0.177	−2.230	−2.427	−1.500

	H1	H2	H3	BIAS
Output	1.583	−4.995	−4.170	0.114



Quantitative Structure Chromatography Relationships in Reversed-phase High Performance Liquid Chromatography: Prediction of Retention Behaviour using Theoretically Derived Molecular Properties

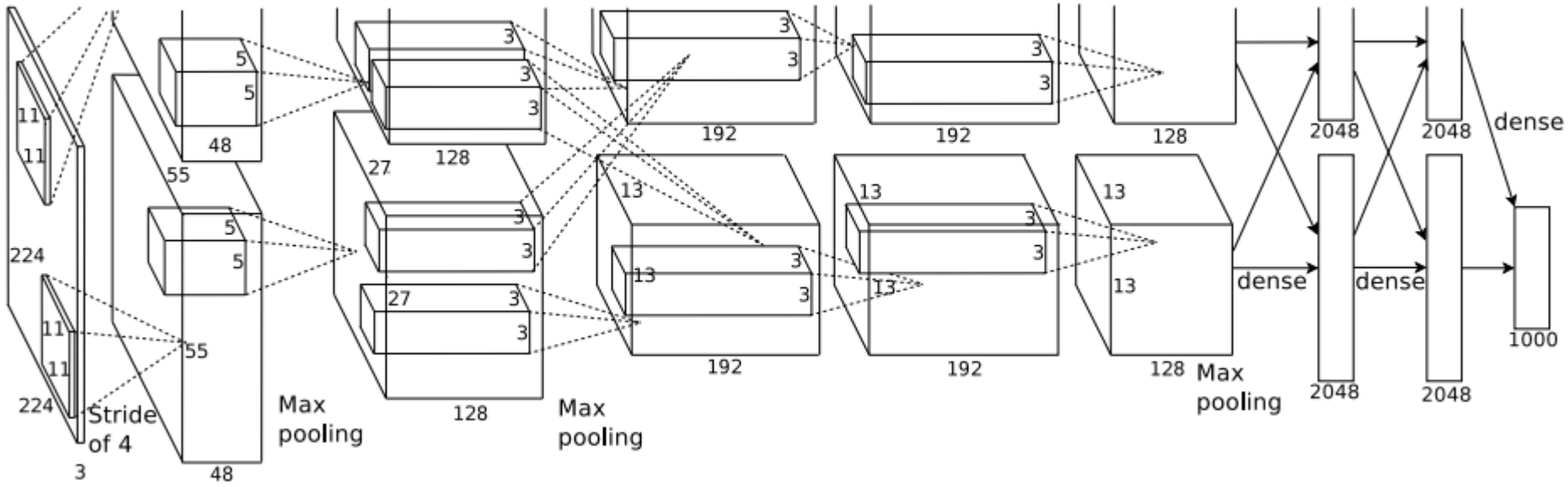
Cupid et al. Chromatographia, 1993, 37(5), 241-249

(24 years ago!)

What's changed? – much 'deeper' networks can now be optimised

The renaissance in NN started with "ImageNet Classification with Deep Convolutional Networks", cited over 6,000 times and is widely regarded as one of the most influential publications in the field.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton created a "large, deep convolutional neural network" that was used to win the 2012 ILSVRC (ImageNet Large-Scale Visual Recognition Challenge).



Using a Deep NN model to recognise images

- What caught attention – Hintons work on image recognition
- deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes
- considerably better than the previous state-of-the-art.
- NN has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax.
- GPU implementation of the convolution operation. Used “dropout” to avoid overfitting.

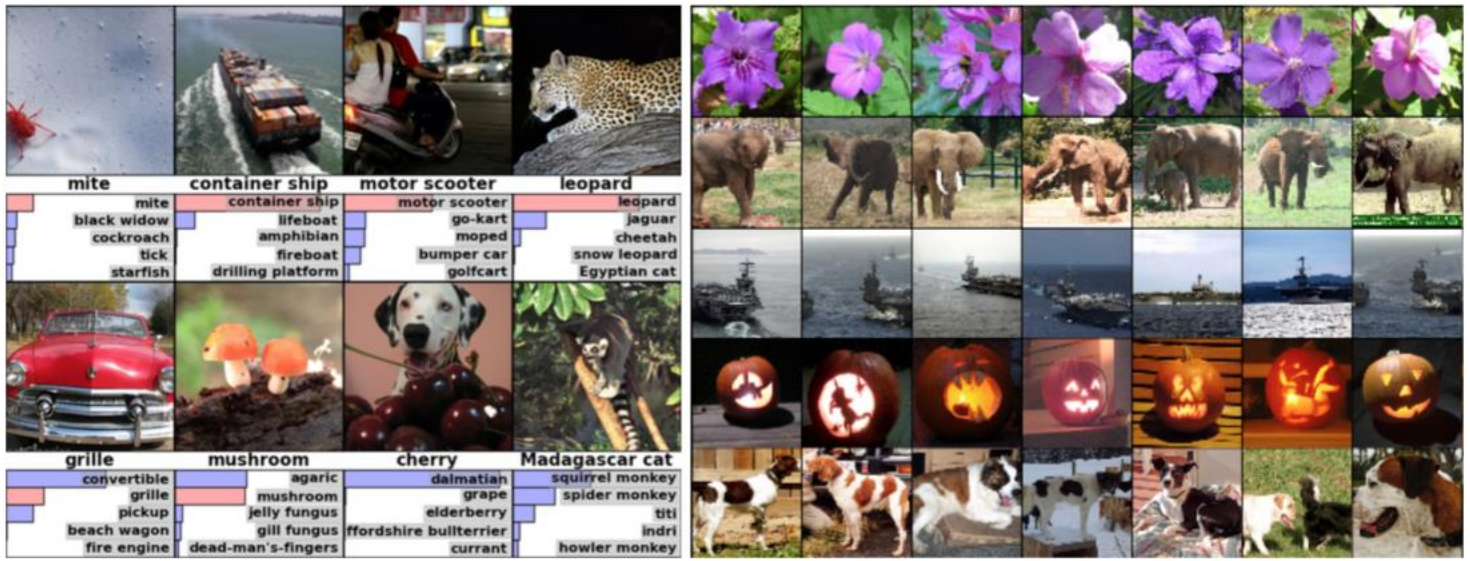


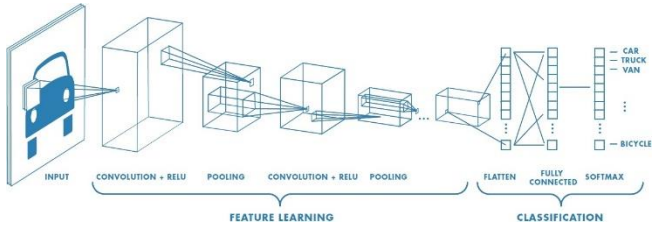
Figure 4: **(Left)** Eight ILSVRC-2010 test images and the five labels considered most probable by our model. The correct label is written under each image, and the probability assigned to the correct label is also shown with a red bar (if it happens to be in the top 5). **(Right)** Five ILSVRC-2010 test images in the first column. The remaining columns show the six training images that produce feature vectors in the last hidden layer with the smallest Euclidean distance from the feature vector for the test image.

What's changed?

- Hinton's revolutionary work on Deep Belief Networks (DBNs):
  - A fast learning algorithm for deep belief nets Neural Computation 2006, 18:1527-1554
- Bengio et al.
  - Greedy Layer-Wise Training of Deep Networks. Advances in Neural Information Processing Systems 2007, 19 (NIPS 2006), pp. 153-160, MIT Press
- Marc'Aurelio et al.
  - Efficient Learning of Sparse Representations with an Energy-Based Model, in J. Platt et al. (Eds), Advances in Neural Information Processing Systems (NIPS 2006), 2007, MIT Press

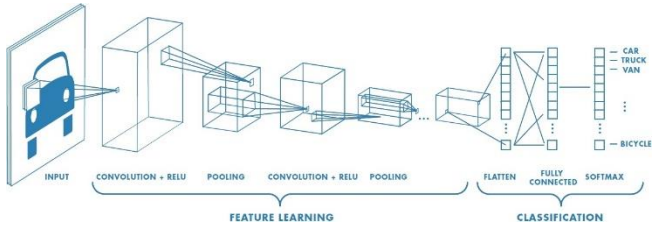
- These include the following key principles:
  - Unsupervised learning of representations is used to (pre-)train each layer.
  - Unsupervised training of one layer at a time, on top of the previously trained ones. The representation learned at each level is the input for the next layer.
  - Use supervised training to fine-tune all the layers (in addition to one or more additional layers that are dedicated to producing predictions).

The renaissance in Neural Networks



- Deep learning allows computational models that are composed of (many) multiple processing layers to learn representations of data with multiple levels of abstraction.
- These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics.
- Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer.

The renaissance in Neural Networks



- Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

- Deep-learning methods are representation-learning methods with multiple levels of abstraction, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level. With the composition of enough such transformations, very complex functions can be learned.

- LeCun et al. N AT U R E | VO L 5 2 1 | 2 8 M AY 2 0 1 5

What is ‘Deep Learning’ ?

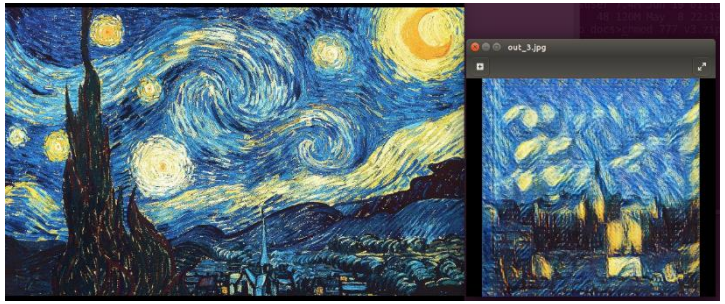
- It belongs to the class of machine learning methods
- Typically includes multiple layers of non-linear processes for feature extraction and a connected series of layers for processing, model building and information extraction.
- As in machine learning – Deep learning can be supervised (models) or unsupervised (classification).
- Different layers (or components) are essentially different layers of abstraction (representative, but not typically ‘real’). The layers can be simple connections, include transformation of the data or even include embedded generative models.
- The architecture of the system in some way follows that of the brain and artificial neural networks are the simplest in terms of topology.

What is ‘Deep Learning’ ?

- The transition from Neural Networks to Deep Learning probably arose from the idea of generative models (Learning multiple layers of representation, Hinton G. E., TRENDS in Cognitive Sciences 2007, 11(10), 428-434).

- I think the ‘Deep’ description is best categorised by the level of abstraction. This a move towards machine intelligence. Humans abstract knowledge, now machines do the same.
- A major advance is that feature selection (a major bug-bear in Machine Learning methods) is often no longer required, or at least can be handled.

Generative models

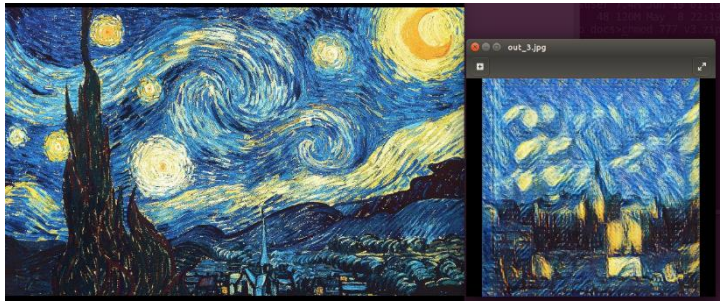


To learn multiple layers of feature detectors when labelled data are scarce or nonexistent, some objective other than classification is required

In a neural network that contains both bottom-up ‘recognition’ connections and top-down ‘generative’ connections it is possible to recognize data using a bottom-up pass and to generate data using a top-down pass.

If the neurons are stochastic, repeated top-down passes will generate a whole distribution of data-vectors (fantasies). This suggests a sensible objective for learning: adjust the weights on the top-down connections to maximize the probability that the network would generate the training data.

Generative models



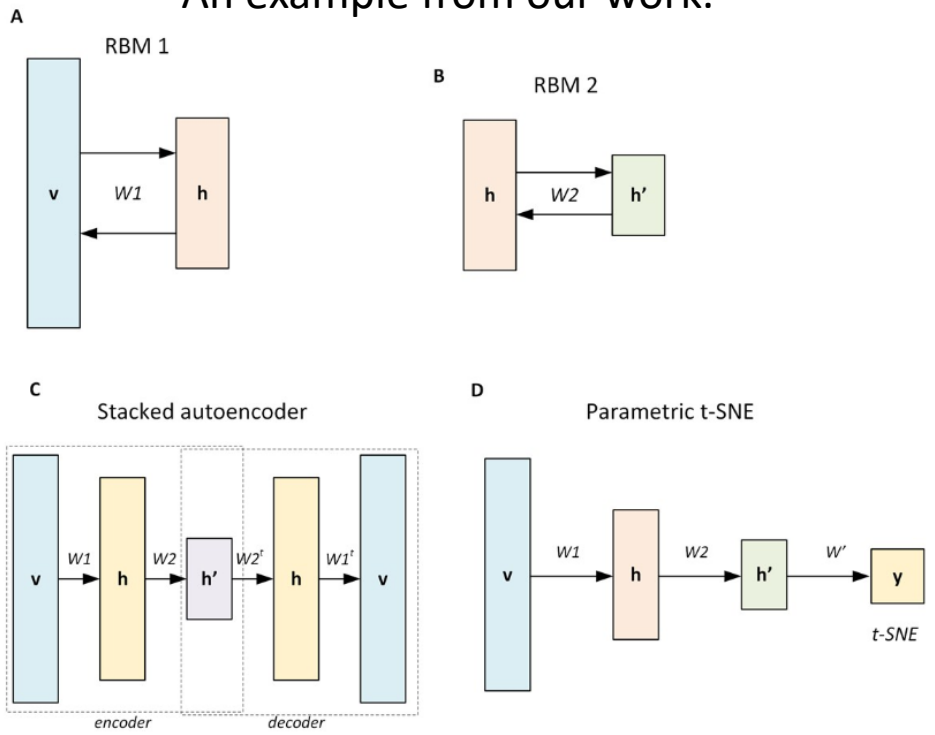
So, the objective is to find a hidden representation that allows generation of the training data. We're not fitting a model to an output, we're fitting an input and generating the input. The hidden layers tell us what is important to generate the model. It does feature selection.

The example developed by Hinton, using as the generative layers "Restricted Boltzmann Machines (RBMs)" – which is based on work by Somlenski (1986) on "Foundations of Harmony Theory" – a mathematical theory of information processing. The combination of approximate inference for learning the generative weights, and "fantasies" (fantasies are generated from the model by using the generative weights in a topdown pass) for learning the recognition weights is known as the 'wake-sleep' algorithm learn deep directed networks one layer at a time by stacking RBMs.
http://www.cs.toronto.edu/~hinton/

Lots of terminology:

Convolutional Networks
Recursive Nets
Autoencoders
Representational Learning
Structured Probabilistic Models
Deep Generative Models
Bi-Directional Recurrent Neural Networks
Recursive Boltzmann Machine
Deep Belief Networks
Generative Adversarial Networks
Long-Short Memory Units
Multi-Layer Perceptron
Rectified Linear Units
Recurrent Neural Networks
Recursive Neural Networks
Representation Learning
T-SNE
Transfer Learning
…

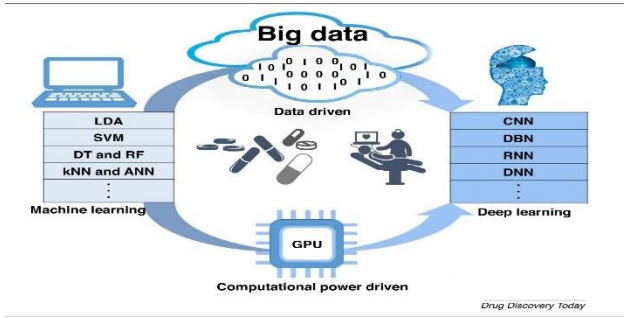
An example from our work.



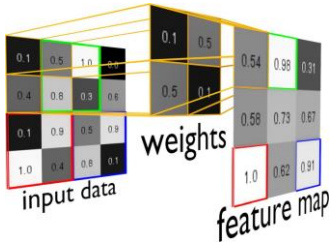
Supplementary Fig 2 - Example of a 4-layer parametric t-SNE model. Two RBMs (A-B) are pre-trained through contrastive divergence. The hidden layer of the first RBM is used as input for the second RBM (B). A stacked autoencoder is defined combining the RBMs (C). Fine-tuning of network weights and biases is performed after adding a t-SNE layer on top (D) of the encoder by backpropagation.

t-distributed stochastic neighbour embedding (**t-SNE**) is a machine learning algorithm for dimensionality reduction developed by Geoffrey Hinton and Laurens van der Maaten. We used a version called parametric t-SNE.

Several reviews in drug discovery and applications of Deep Learning

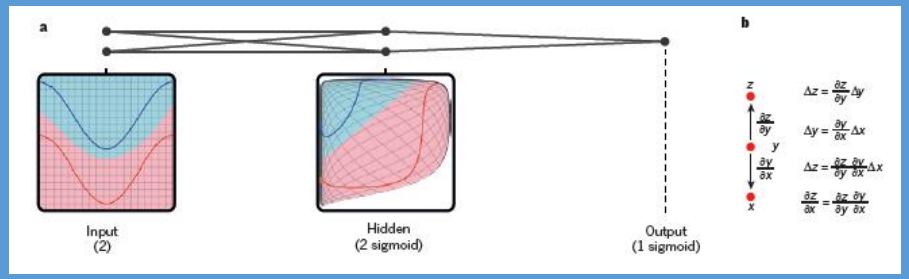


From machine learning to deep learning: progress in machine intelligence for rational drug discovery
Zhang et al. Drug Discovery Today Volume 22, Number 11 November 2017
'The most commonly used networks are convolutional neural networks (CNN), stacked autoencoders, deep belief networks (DBN), and restricted Boltzmann machines'



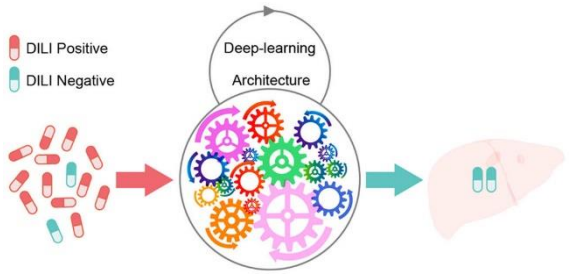
Deep Learning in Drug Discovery
Gawehn, Hiss and Schneider. Mol. Inf. 2016, 35, 3 – 14
'With the development of
new deep learning concepts such as RBMs and CNNs, the molecular modeler's tool box has been equipped with potentially game-changing methods.'

Is Multitask Deep Learning Practical for Pharma?
Ramsundar et al. Chem. Inf. Model., 2017, 57 (8), pp 2068–2076
' Our analysis and open-source implementation in DeepChem provide an argument that multitask deep networks are ready for widespread use in commercial drug discovery.'



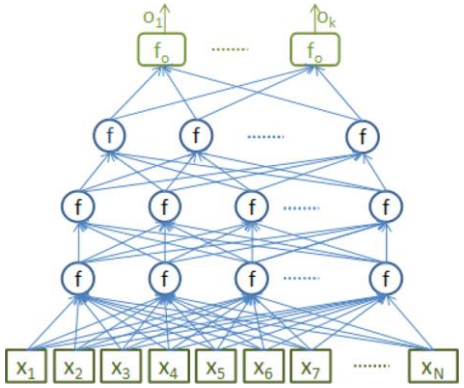
Deep learning. Le Cun (Hinton) et al. Nature, 2015, 521, 436-444.

Several reviews in drug discovery and applications of Deep Learning

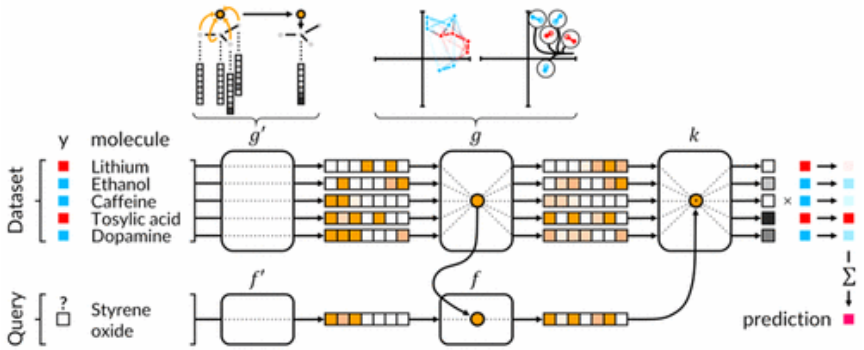


Deep Learning for Drug-Induced Liver Injury
Xu et al. J. Chem. Inf. Model. 2015, 55, 2085−2093

Protein−Ligand Scoring with Convolutional Neural Networks.
Ragoza et al. J. Chem. Inf. Model. 2017, 57, 942−957



Deep Neural Nets as a Method for Quantitative Structure−Activity Relationships.
Junshui et al. J. Chem. Inf. Model. 2015, 55, 263−274



Low Data Drug Discovery with One-Shot Learning.
Alte-Tran et al. ACS Cent. Sci., 2017, 3 (4), pp 283−293
'we demonstrate how one-shot learning can be used to significantly lower the amounts of data required to make meaningful predictions in drug discovery applications. We introduce a new architecture, the iterative refinement long short-term memory, that, when combined with graph convolutional neural networks, significantly improves learning of meaningful distance metrics'

Deep Learning and Big Data

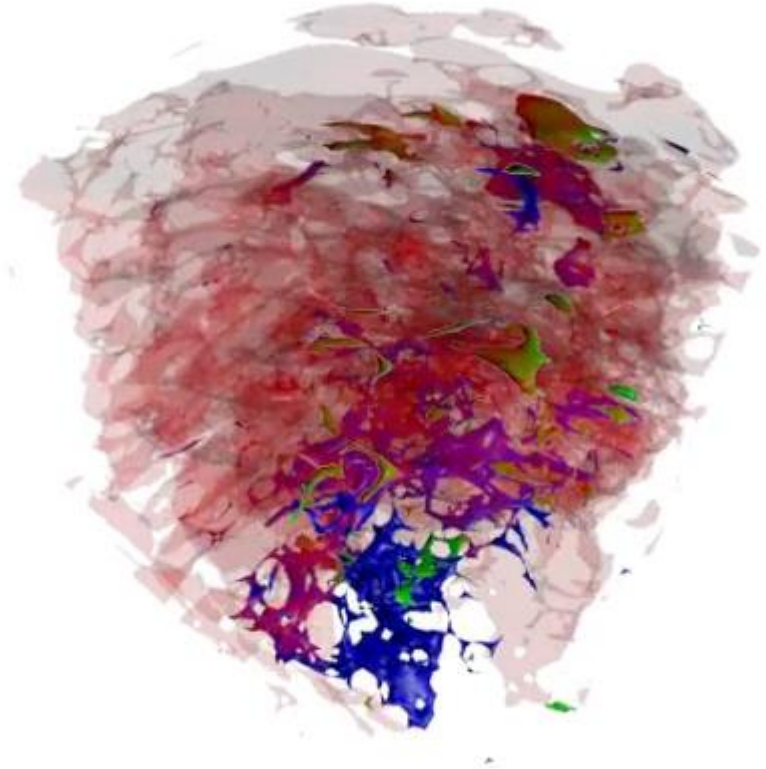
To a large extent, they have gone hand in hand. Along with massive computing power (GPUs) you typically need:

- A very large amount of data to refine what is essentially a very large model with huge numbers of variables.
  - Although this is changing see reference ACS Cent. Sci., 2017, 3 (4), pp 283–293 on one-shot learning.
    - Reinforcement learning, resampling, stepwise methods can be more efficient in smaller datasets.
    - The other problem can be that datasets are too big. We recently had a problem of clinical imaging data that was too big to compute.

- Tests of robustness that are up to finding the ‘ground truth’
  - L1 and L2 regularisation, Bayesian regularization with penalties, early termination, dropout, resampling

- Ways of finding the correct hyperparameters - it’s not just plug and play. Serious work is required to optimise.

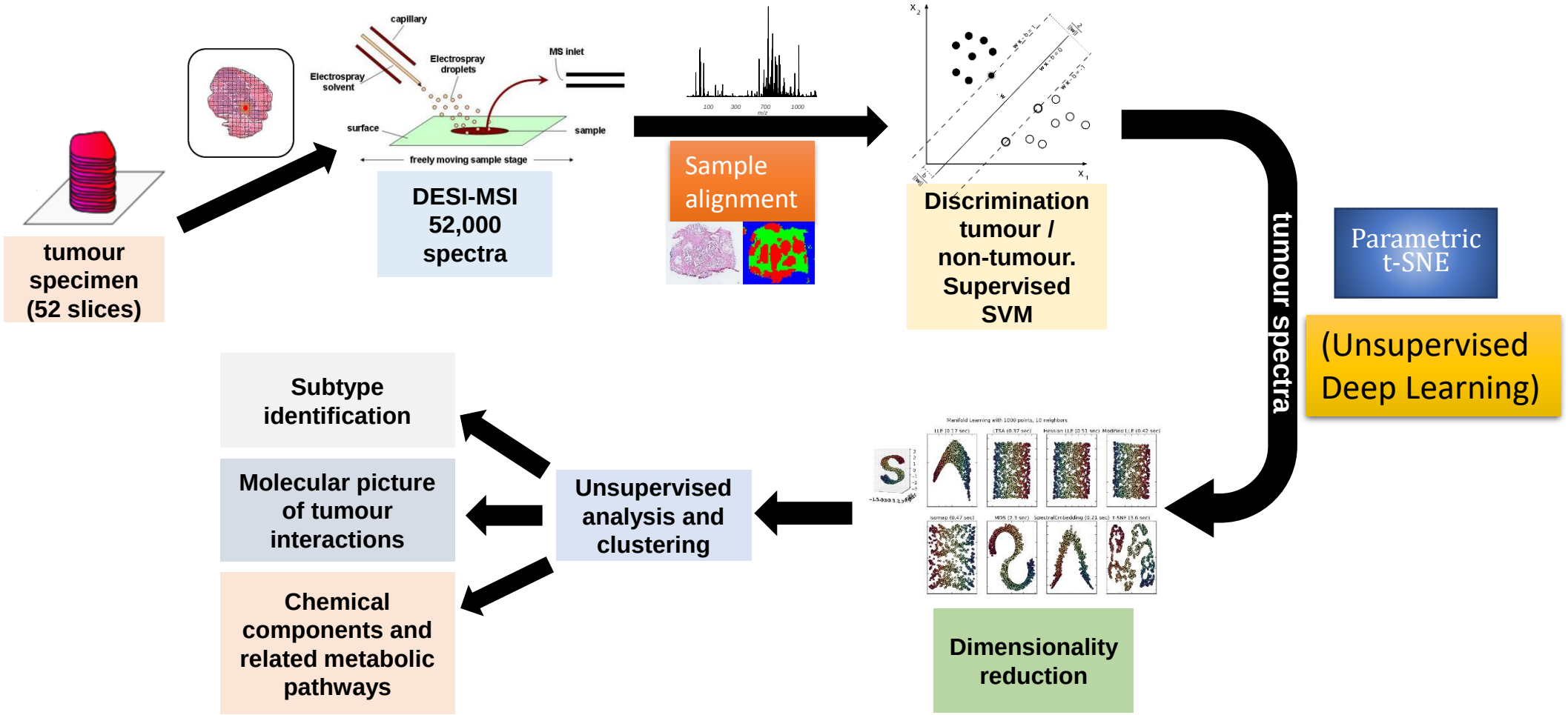
Data Driven analysis of tumours using DESI-MS



***Paolo Inglese*, Robert Glen, Zoltan Takats, Jeremy Nicholson**

Data Driven Identification and analysis of tumour sub-types

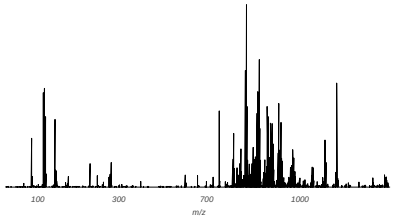
*The tumour microenvironment is **3-dimensional**.*
More opportunity to capture the biological interactions.



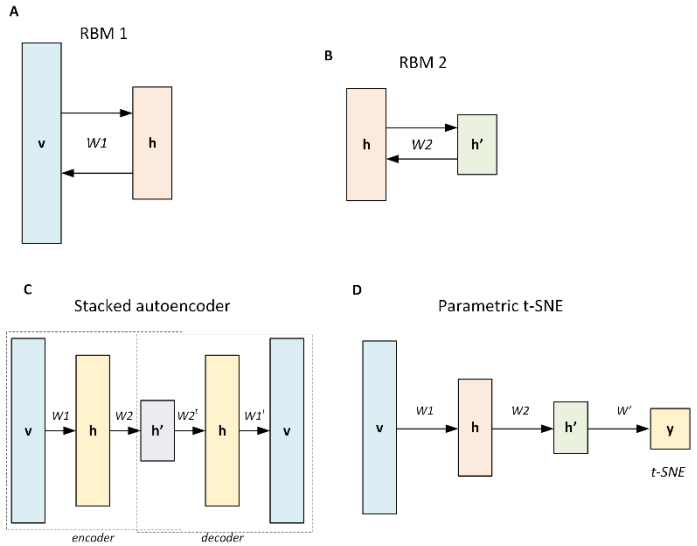
Inglese, Paolo, et al. "Deep learning and 3D-DESI imaging reveal the hidden metabolic heterogeneity of cancer." *Chemical Science* (2017), 5, 3500-3511

The spectra classified as belonging to the tumour class can be analysed through unsupervised techniques.

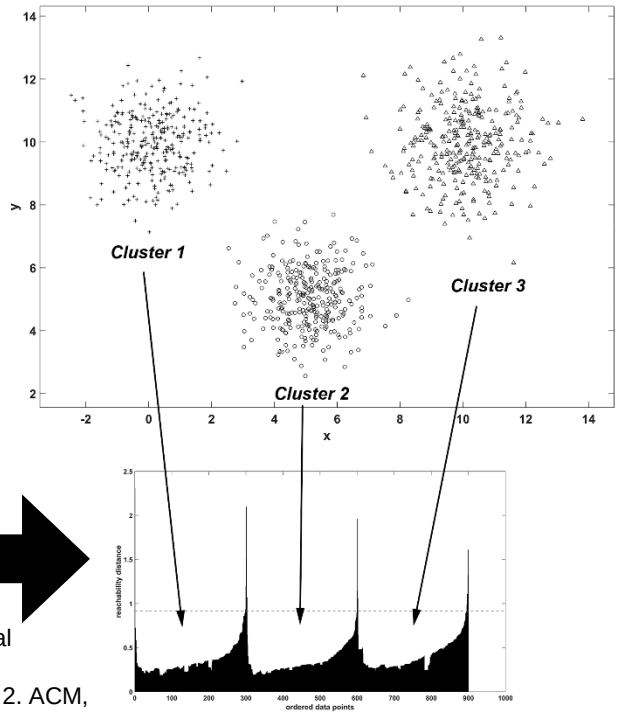
Tumour spectra



Unsupervised dimensionality reduction (parametric t-SNE)*



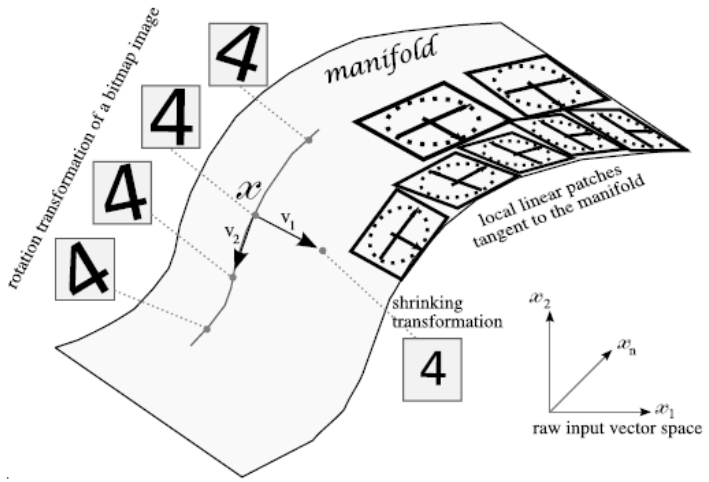
OPTICS: density-based clustering**



*Maaten, L. (2009). *Learning a parametric embedding by preserving local structure*. International Conference on Artificial Intelligence and Statistics.

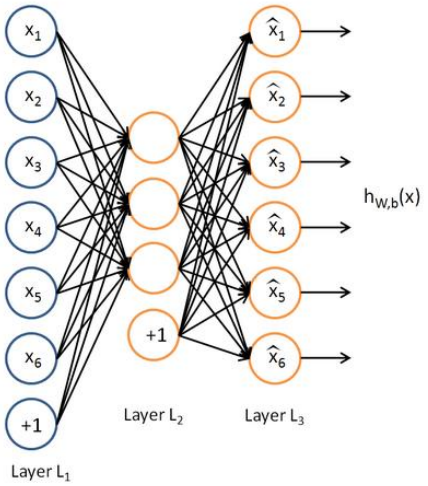
**Ankerst, Mihael, et al. "OPTICS: ordering points to identify the clustering structure." *ACM Sigmod record*. Vol. 28. No. 2. ACM, 1999.

Using autoencoders to learn the ‘shape’ of the data – compress it into a lower dimensionality



Manifold learning determines the global low-dimensional embedding of the data space by looking at the local characteristics of the high-dimensional space.

Autoencoders can be seen as simple parametric manifold learning models.



Autoencoders are simple feed-forward ANN (Artificial Neural Network) trained to reconstruct the input.
The hidden layer can be used to ”compress” the data into a lower-dimensional space (latent).
They are **deterministic** (input X always returns the same output f(X)).

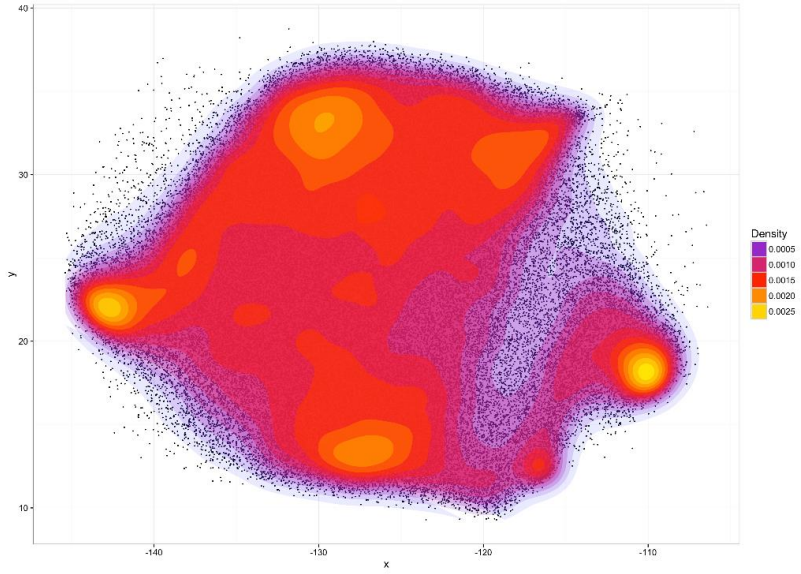
Bengio, Yoshua. "Learning deep architectures for AI." *Foundations and trends® in Machine Learning* 2.1 (2009): 1-127.
*http://ufldl.stanford.edu/wiki/index.php/Autoencoders_and_Sparsity

A comparison with PCA shows the limits of linear methods. Parametric t-SNE is capable of finding a more complex mapping of the similarities between spectra.

parametric t-SNE (391-250-250-1000-2)

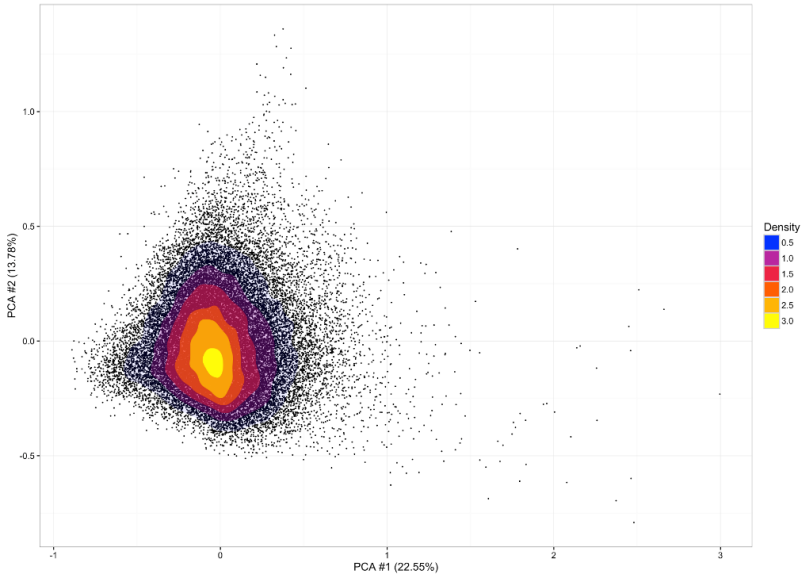
Perplexity = 30

Train: non-lin. GD (Polak-Ribiere) 500 epochs



PCA

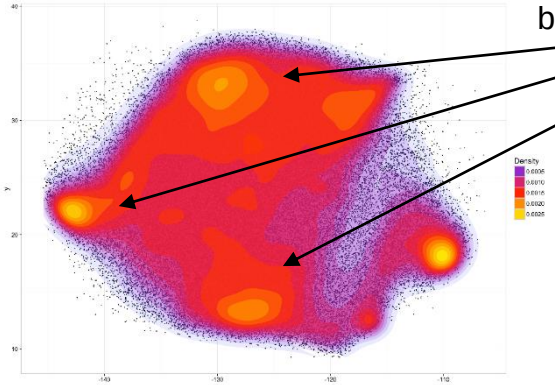
(tested: autoscaling, pareto, vast, range, level scaling)

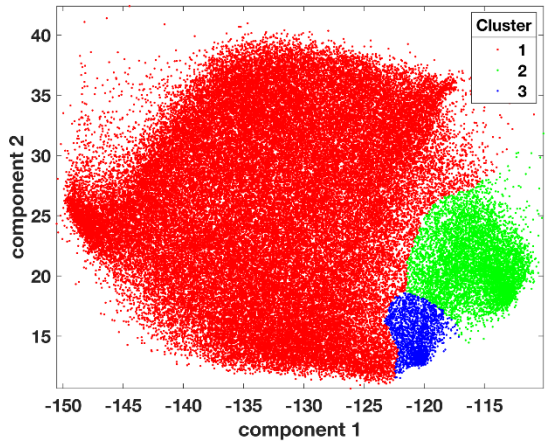


Simple linear techniques (widely used) are not capable of capturing the complexity of the data structure.

The low-dimensional data (from parametric t-SNE) can be clustered using a density-based clustering algorithm: OPTICS (similar to DBSCAN). The optimal number of clusters can be identified through the reachability plot.

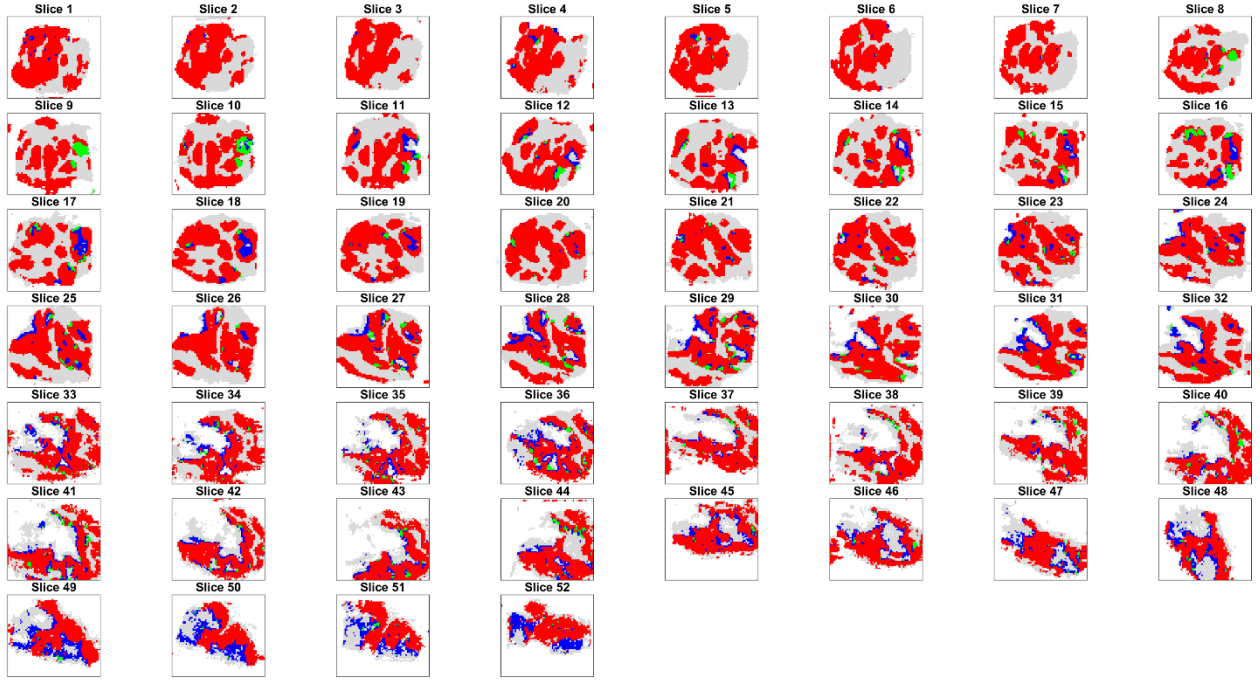
Some high dense regions were not separated enough to be considered individual clusters.





Clustering results are compatible with a sequence of contiguous tissue sections.

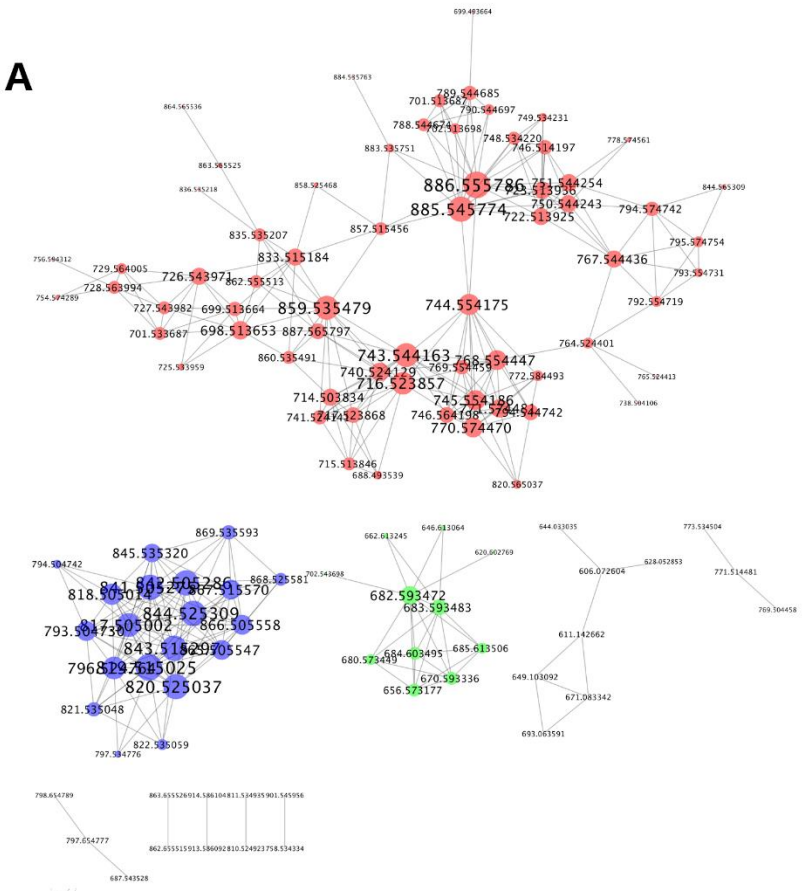
Projection of the cluster labels on the tissue



Comparing the results of OPTICS with those of a co-expression network: from spectra clustering to ion clustering. Identifies the key differentiating ions in the clusters.

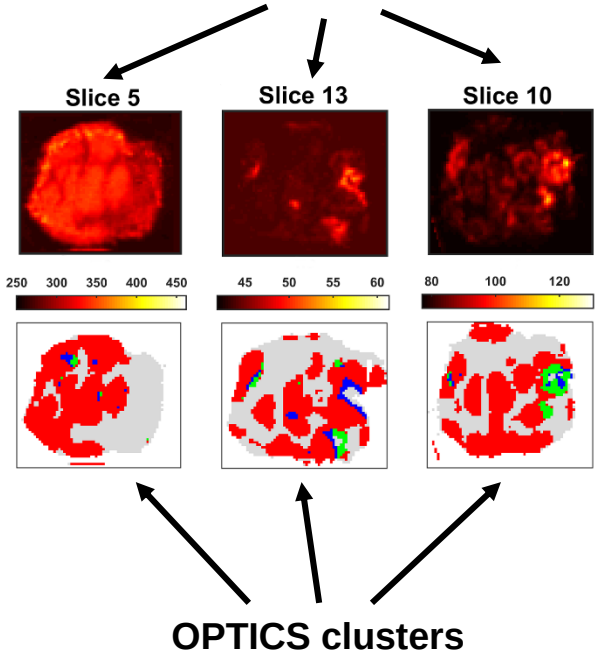
Co-expression network:
adjacency matrix = pairwise ion correlations

A



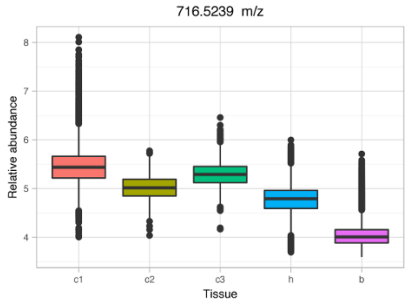
Look for concordance between the sub-networks and the OPTICS clusters.

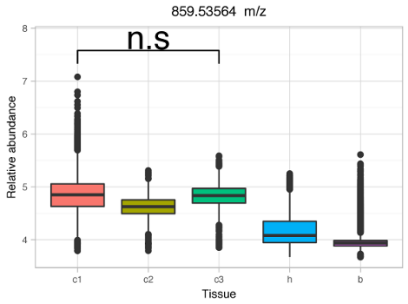
Sum of ion intensities in the 3 largest sub-networks

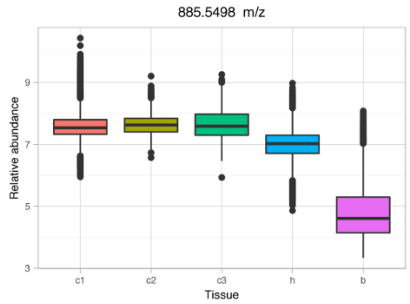


Matching between the 3 largest sub-networks and the 3 OPTICS clusters can be used to identify the characteristic ions of the 3 tissue sub-types. Here the intensities of the 3 ions most connected in each sub-networks are compared in the 3 clusters (c1, c2, c3), the healthy (h) and the background (b) classes.
PI – phosphoinositols, PE – phosphotidylethanolamines, PG – phosphtidylglycerols, Ceramides

PIs and PEs

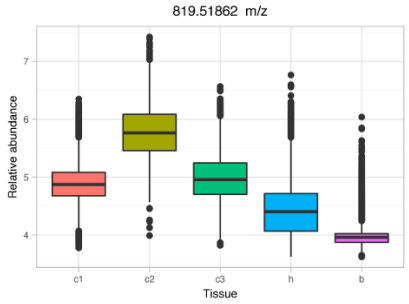


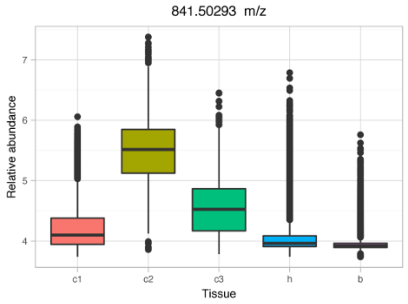


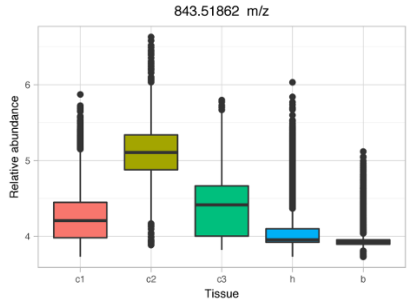


Main tumour tissue

PGs

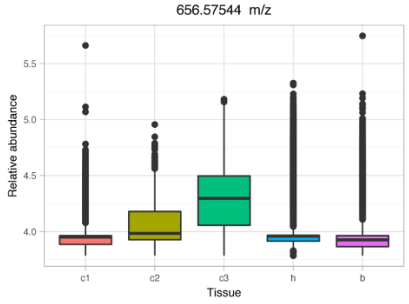


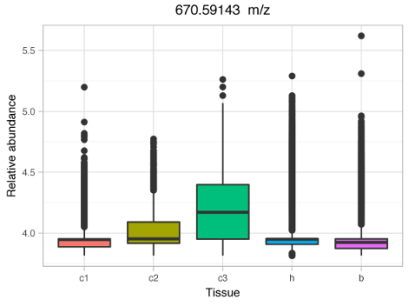


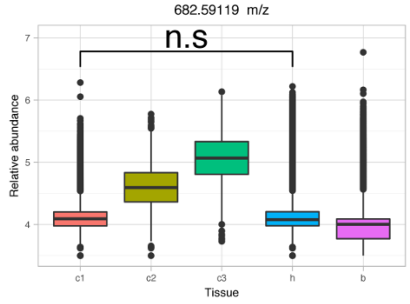


Apoptotic surrounding Necrotic tissue

Ceramides





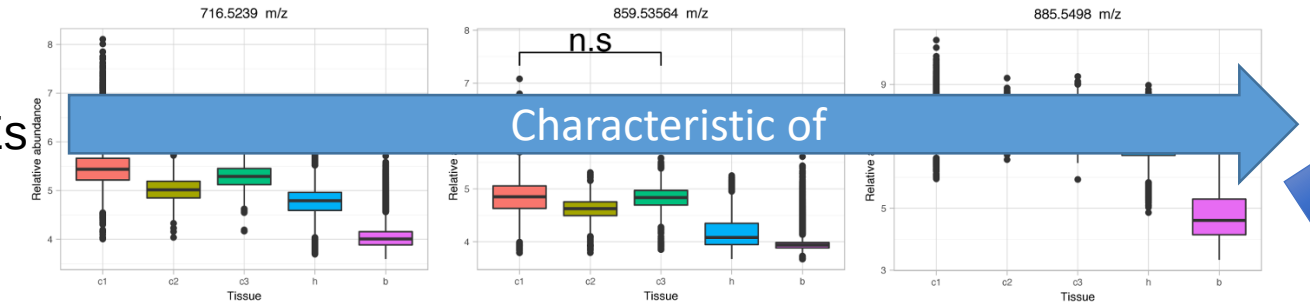


Necrotic tissue

We have now a deeper insight of the molecular distributions inside the tumour.

Matching between the 3 largest sub-networks and the 3 OPTICS clusters can be used to identify the characteristic ions of the 3 tissue sub-types. Here the intensities of the 3 ions most connected in each sub-networks are compared in the 3 clusters (c1, c2, c3), the healthy (h) and the background (b) classes.
PI – phosphoinositols, PE – phosphotidylethanolamines, PG – phosphtidylglycerols, Ceramides

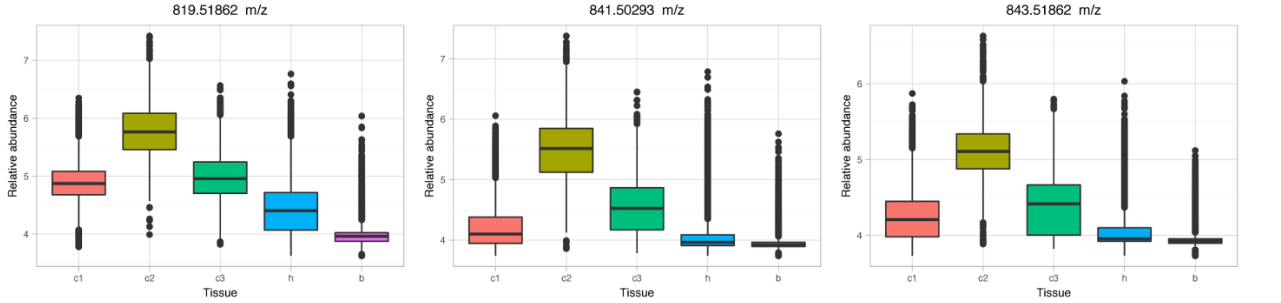
PIs and PEs



Characteristic of

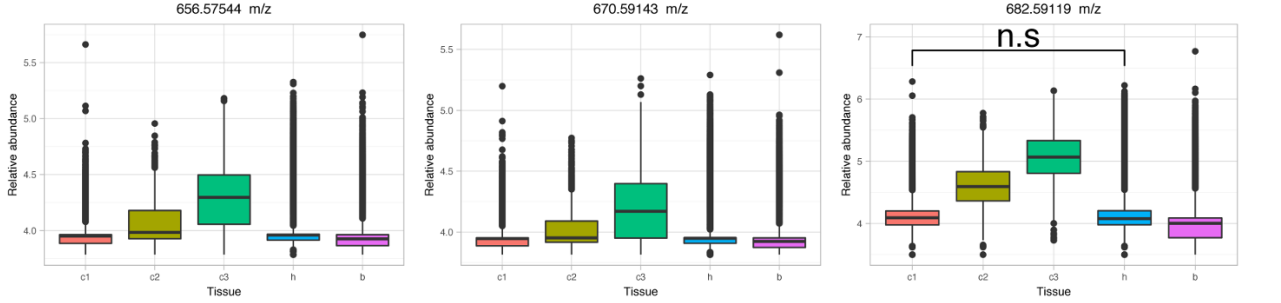
Main tumour tissue

PGs



Apoptotic surrounding Necrotic tissue

Ceramides

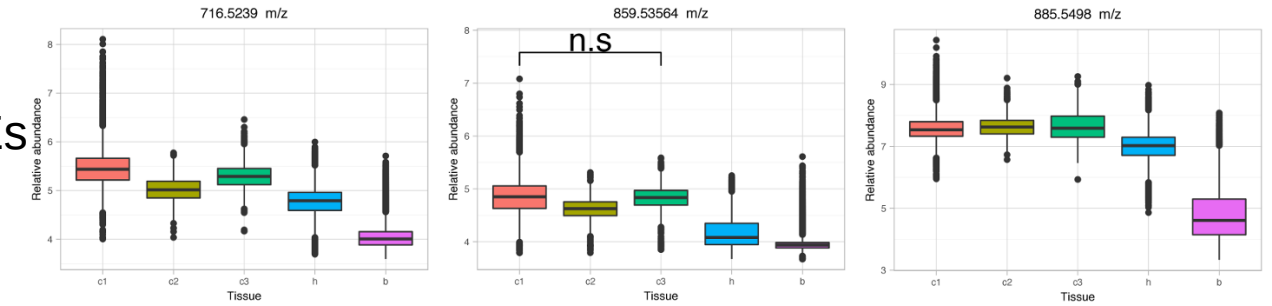


Necrotic tissue

We have now a deeper insight of the molecular distributions inside the tumour.

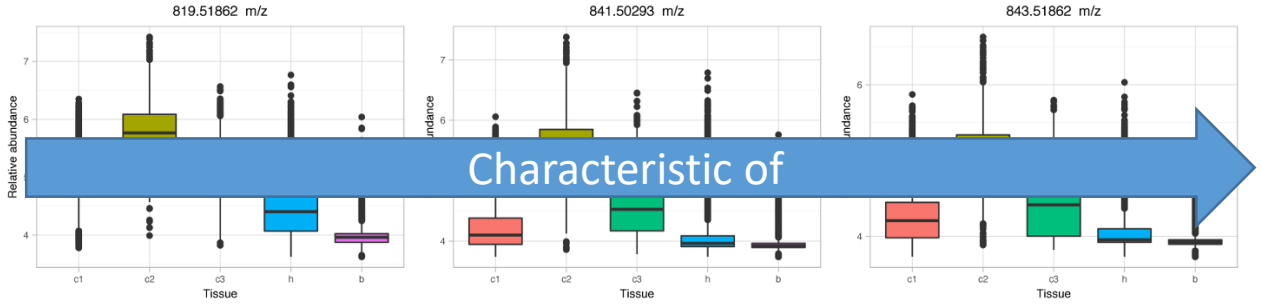
Matching between the 3 largest sub-networks and the 3 OPTICS clusters can be used to identify the characteristic ions of the 3 tissue sub-types. Here the intensities of the 3 ions most connected in each sub-networks are compared in the 3 clusters (c1, c2, c3), the healthy (h) and the background (b) classes.
PI – phosphoinositols, PE – phosphotidylethanolamines, PG – phosphtidylglycerols, Ceramides

PIs and PEs



Main tumour tissue

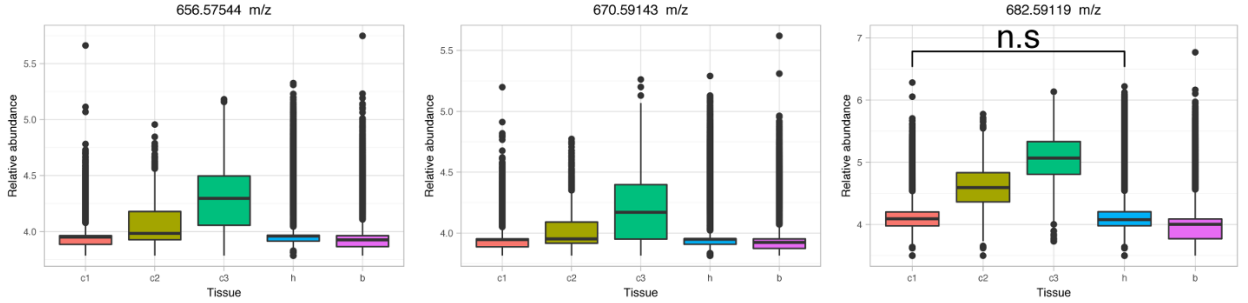
PGs



Characteristic of

Apoptotic surrounding Necrotic tissue

Ceramides

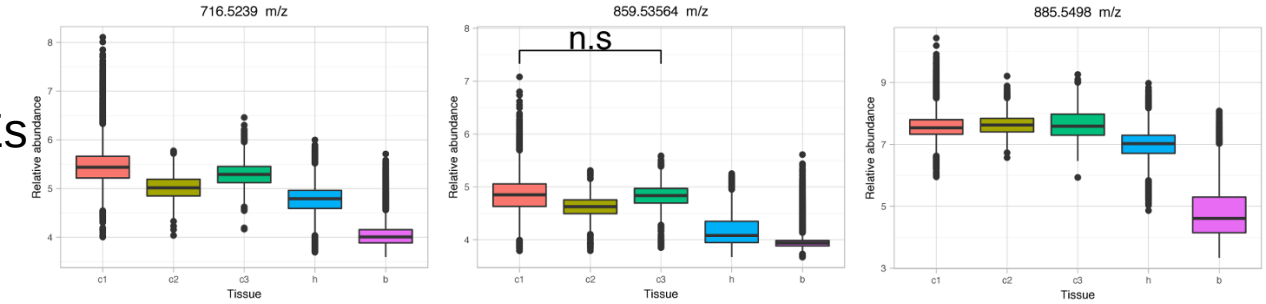


Necrotic tissue

We have now a deeper insight of the molecular distributions inside the tumour.

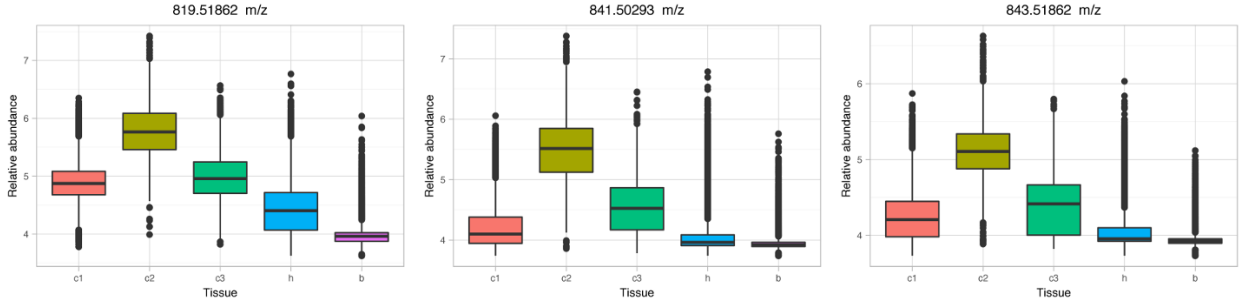
Matching between the 3 largest sub-networks and the 3 OPTICS clusters can be used to identify the characteristic ions of the 3 tissue sub-types. Here the intensities of the 3 ions most connected in each sub-networks are compared in the 3 clusters (c1, c2, c3), the healthy (h) and the background (b) classes.
PI – phosphoinositols, PE – phosphotidylethanolamines, PG – phosphtidylglycerols, Ceramides

PIs and PEs



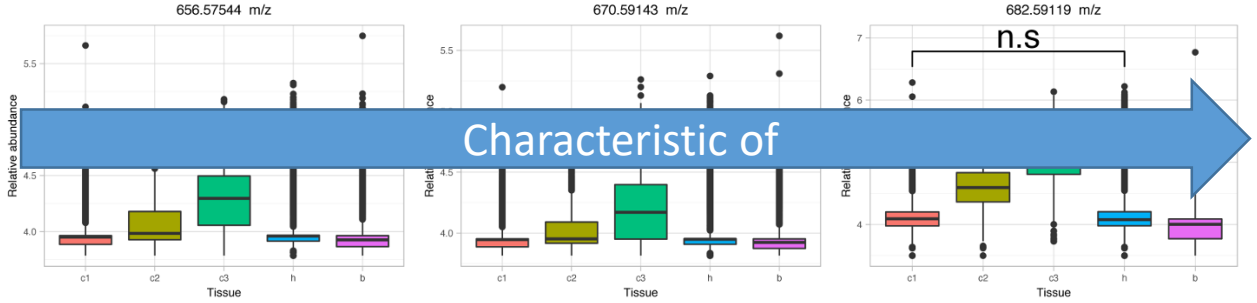
Main tumour tissue

PGs



Apoptotic surrounding Necrotic tissue

Ceramides



Characteristic of

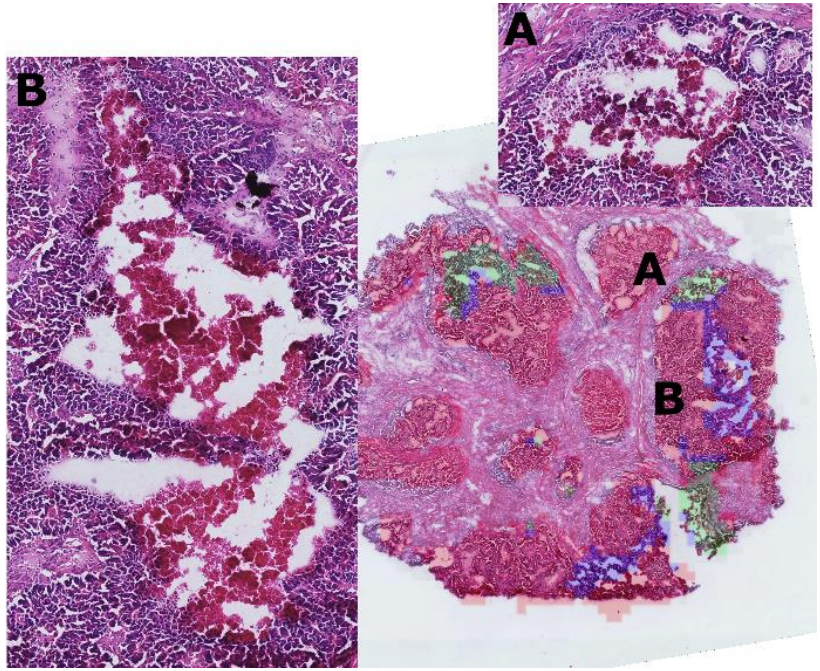
Necrotic tissue

We have now a deeper insight of the molecular distributions inside the tumour.

Clusters and Chemistry

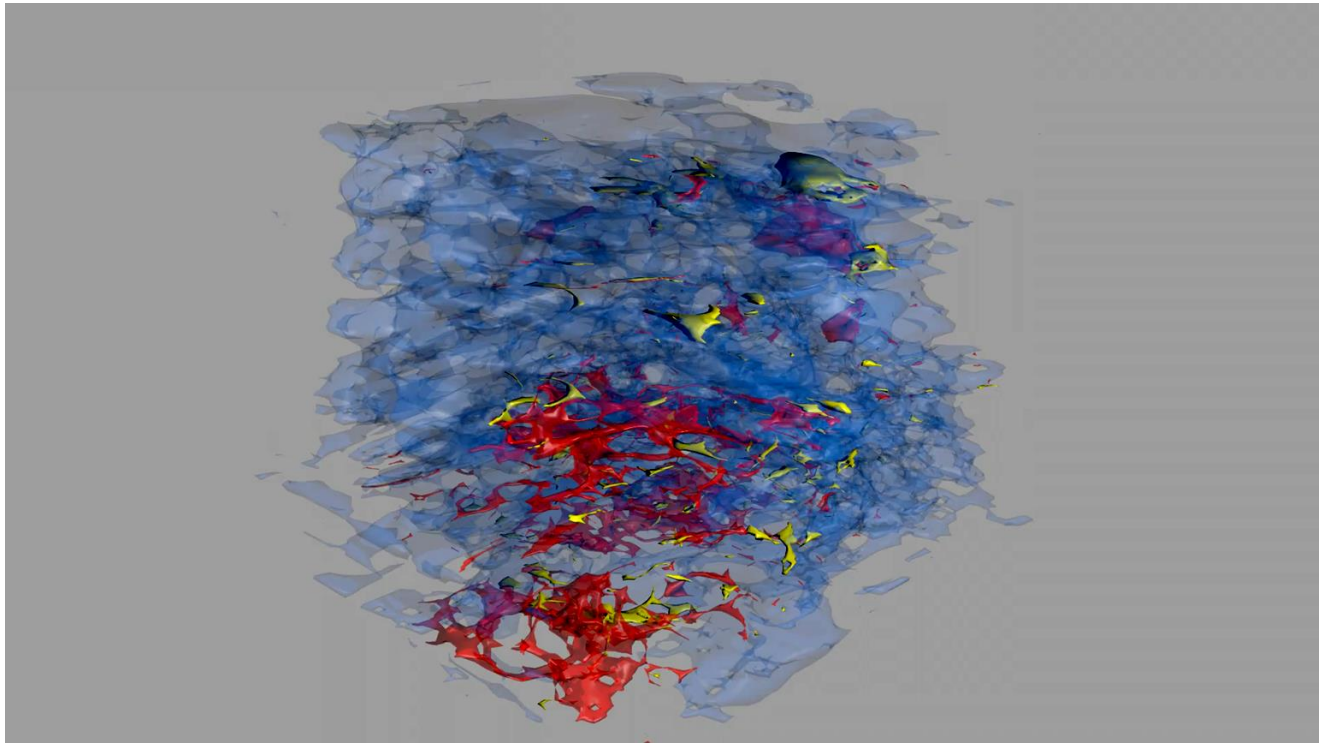
- Cluster 1. Was associated with ions expressed more extensively in the entire tumour, was characterised by an abundance of phophatidylethanolamines (PE), and these high levels have been associated with rapidly proliferating human colorectal cancer in previous work. Additionally, the abundance of phospatidylinositols (PI) was found only in cluster one, which are also hallmarks of viable cancer tissue.
- Cluster 2. Phosphatidylglycerols (PG) were found in cluster two, indicating the presence of mucus in mucinous subtype colorectal malignant tissue as PGs generally serve as surfactants in the human body. The presence of very long acyl chains (n>18) excludes a bacterial origin and indicates peroxisomal dysfunction in this segment. An abundance of phosphatidylserine (PS) was found only in cluster two, which has previously been associated with apoptosis of colon cancer cells.
- Cluster 3. Characterised by an abundance of ceramides, which indicates the presence of a process of necrosis/apoptosis, in agreement with the gross histological appearance in this sub-region. The increased concentration of ceramides is clearly associated with the degradation of sphingolipids in the necrotic cell debris.

Reconstruction of the 3D volume of the tumour



Example of comparison between clusters and histological characteristics of the tissue. Micro-environments are identifiable from the analysis and aid the histologist

Rendering of the smoothed tumour clusters and identify cancer subtypes based on metabolism.



Where to find out about Deep Learning

THEANO – python library and tutorials (from MILA lab at University of Montreal) http://deeplearning.net/tutorial/

List of Deep Learning software tools
http://deeplearning.net/software_links/

Some of the more popular packages

Theano – CPU/GPU symbolic expression compiler in python (from MILA lab at University of Montreal)
Torch – provides a Matlab-like environment for machine learning algorithms in lua
Pylearn2 – Pylearn2 is a library designed to make machine learning research easy.
Blocks – A Theano framework for training neural networks
Tensorflow – TensorFlow is an open source software library for numerical computation using data flow graphs.
MXNet – MXNet is a deep learning framework designed for both efficiency and flexibility.
Caffe -Caffe is a deep learning framework made with expression, speed, and modularity in
Lasagne – Lasagne is a lightweight library to build and train neural networks in Theano.
Keras– A theano based deep learning library.
……. Lots more