V5 – peak detection

Detecting peaks in observed data is a common task in many fields.

Program for today:

- Principles of peak detection
- Peak detection in biomedical 1D-data
    - ChIP-seq data
    - MS data
- Peak detection in biomedical 2D-data
    - breathomics

Peak detection - basics

Computer scientists
(-> Cormen book)
are mostly interested in devising methods to determine peaks most efficiently
-> Divide & Conquer strategy

Noise is often irrelevant to computer scientists.

Instead, **bioinformaticians** must detect peaks in noisy data most precisely.

This an algorithm from the idealized world of CS …

1D Peak Finding

- Given an array $A[0..n-1]$:

- $A[i]$ is a **peak** if it is not smaller than its neighbor(s):

$$A[i-1] \leq A[i] \geq A[i+1]$$

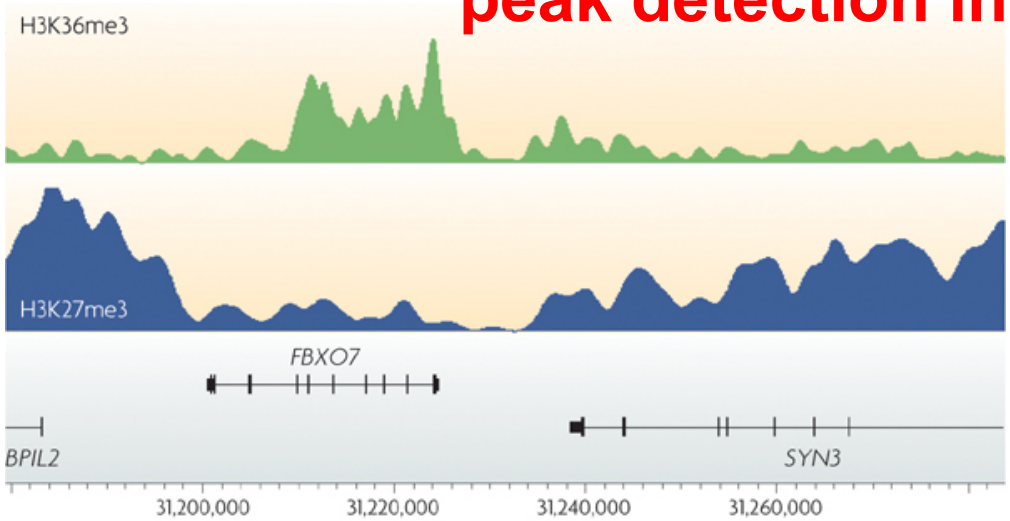
where we imagine

$$A[-1] = A[n] = -\infty$$

- Goal: Find *any* peak

https://courses.csail.mit.edu/6.006/spring11/lectures/lec02.pdf

peak detection in ChIP-seq data



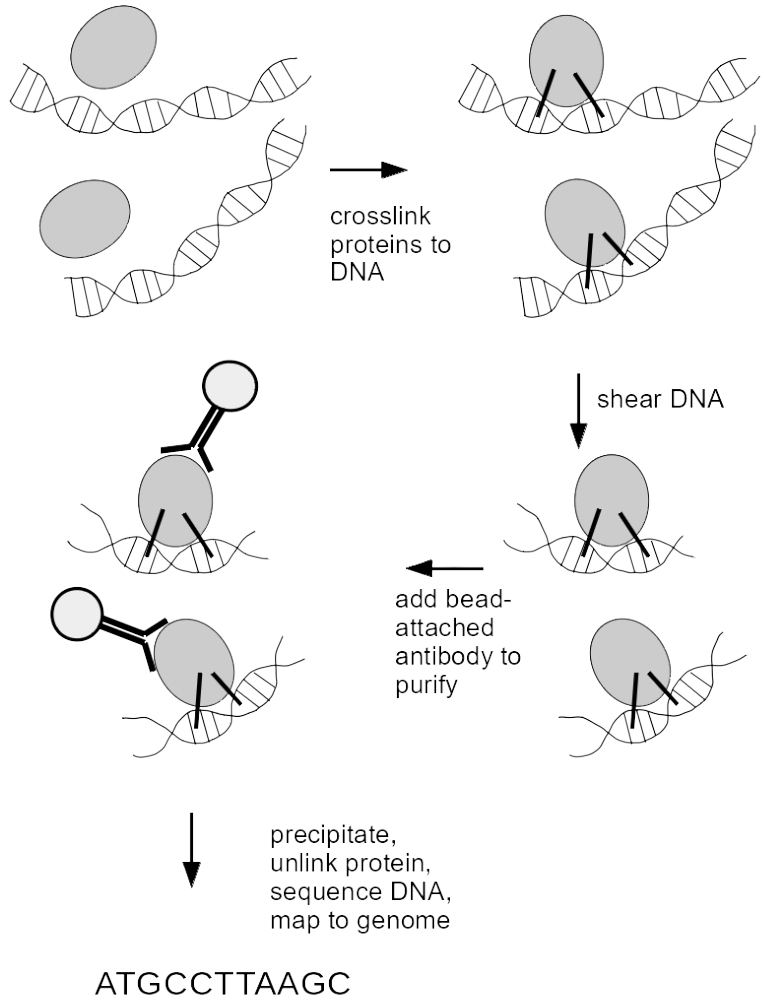
Regulation of gene expression is one of the fundamental means by which cells adapt to internal and external environments.

Many **regulatory mechanisms** rely on **modifying the DNA** either through covalent modification or by intermolecular interactions.

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) data provide readouts of these modifications, such as the location and frequency of binding of a transcription factor (TF) or the distribution of histone modifications that are used by the cell to establish or maintain specialized chromatin domains.

H3K36me3: Lys36 of histone 3 is tri-methylated

Main experimental steps of the **ChIP-seq** protocol.



Park J, Nature Reviews Genetics, 10, 669 (2009)
Thomas et al. Brief Bioinform. 18: 441–450 (2017) .

peak detection in ChIP-seq data

Data for ChIP-seq peak calling: stacks of **aligned reads** across a genome.

Some of these stacks correspond to the **signal of interest**.

Many other stacks are regarded as experimental noise.

Typically, there are 3 – 5 data sets of replicates.

Regions are scored by the number of tags in a window of a given size.

Then they are assessed by **enrichment** over control.

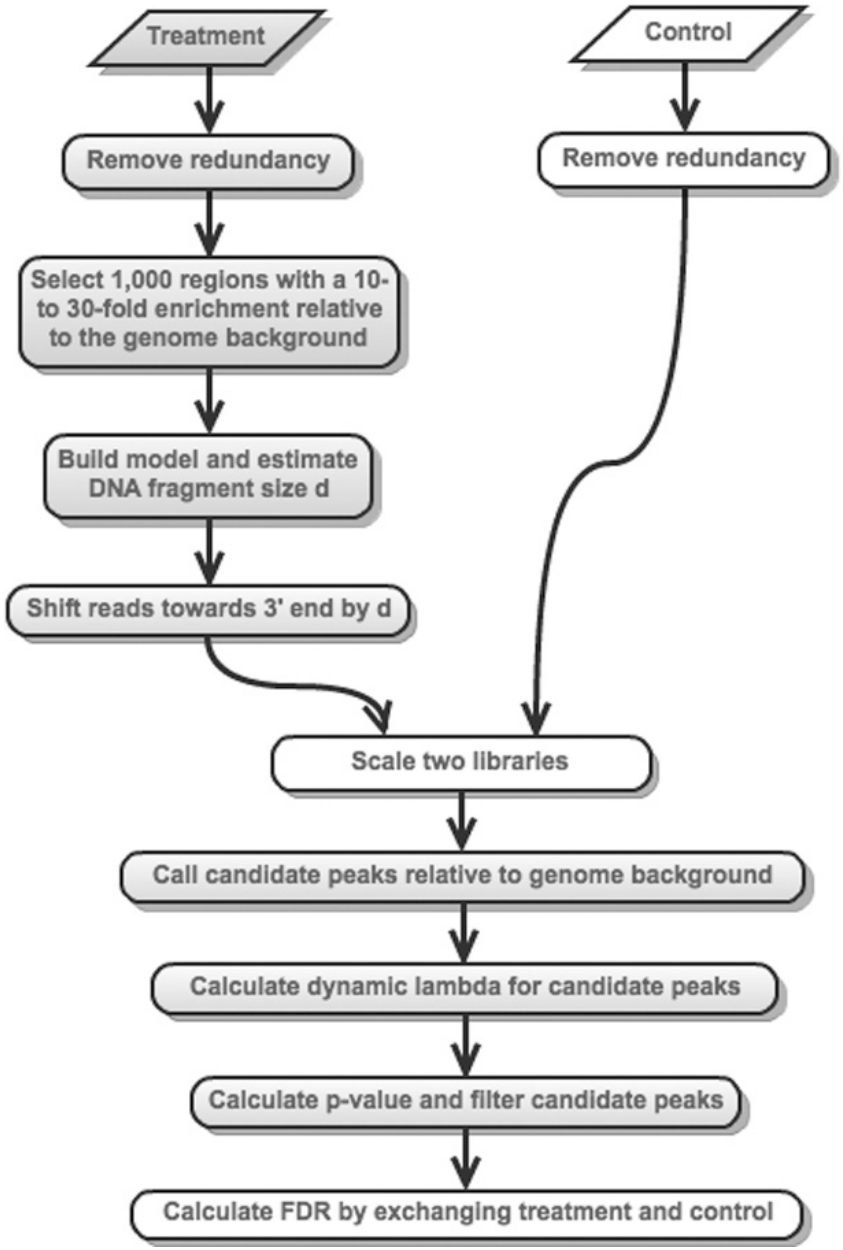
Different applications of ChIP-seq produce different types of peaks.

Most current tools are designed to detect **sharp peaks** (TF binding, histone modifications at regulatory elements)

Alternative tools exist to detect **broader peaks** (expressed/repressed domains).

Park J, Nature Reviews Genetics, 10, 669 (2009)
Thomas et al. Brief Bioinform. 18: 441–450 (2017) .

MACS: popular for detecting peaks in ChIP-seq data



MACS slides a window of 2× sonication size across the genome to identify regions that are significantly enriched relative to the genome background.

MACS models the number of reads from a genomic region as a **Poisson distribution** with dynamic parameter λ_{local}.

$$f(n, \lambda) = (\lambda^n e^{-\lambda})/n!$$

Based on λ_{local}, MACS assigns every candidate region an enrichment p-value. Those regions passing a user-defined threshold (default 10^{-5}) are reported as the final **peaks**.

Zhang et al. Genome Biol. (2008) 9, R137
Feng et al. Nature Prot 7, 1728 (2012)

Features of ChIP-seq peak detection methods

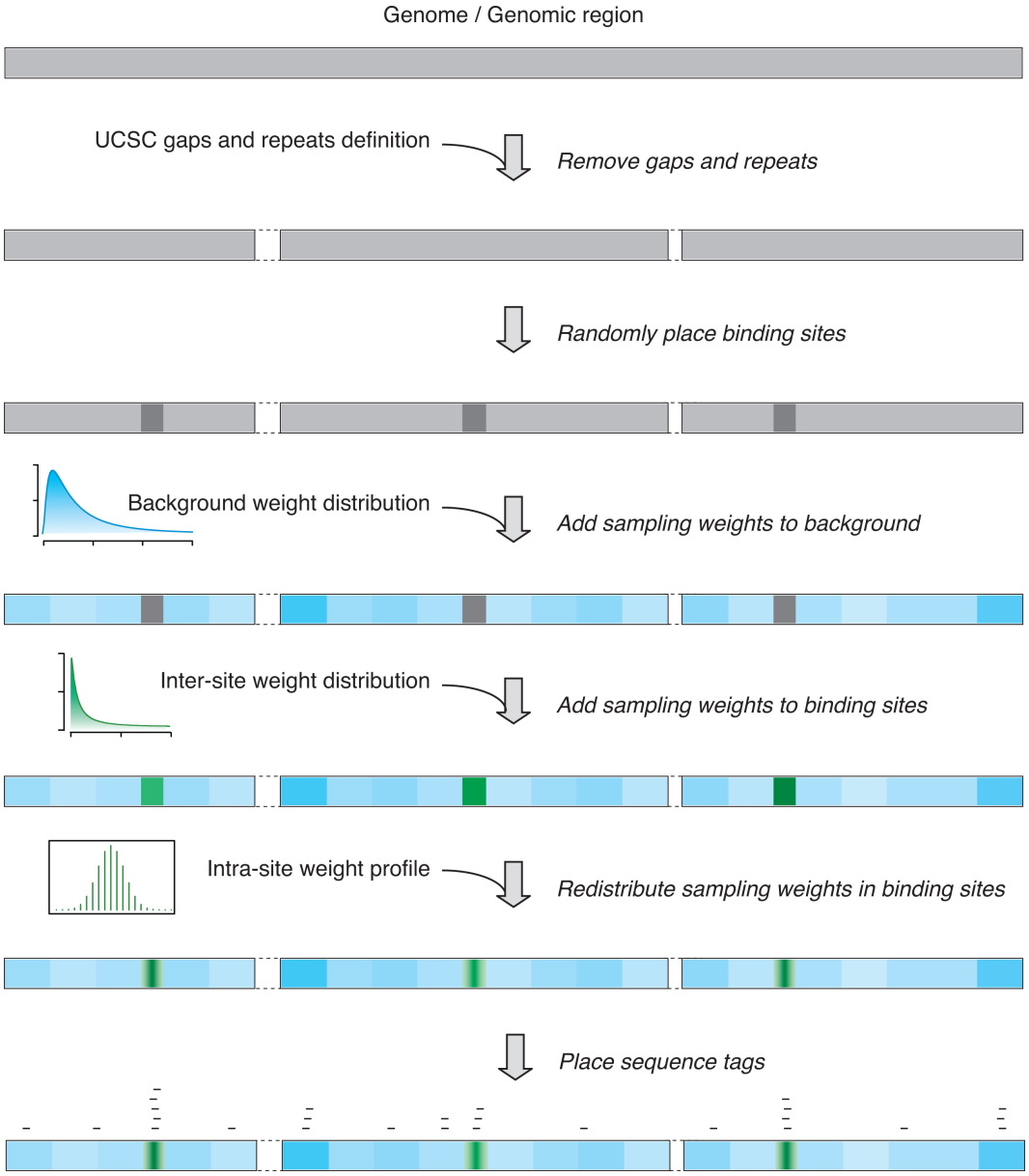
Table 1. Features of peak calling methods

	GEM	BCP (TF)	BCP (Histone)	MUSIC	MACS2	ZINBA	TM
Locating the potential peaks							
High resolution	Yes	Yes	No	Yes	Yes	No	Yes
ChIP and input sample signals combined	No	No	No	No	No	Yes	Yes
Multiple alternate window sizes	Yes	Yes	Yes	Yes	No	No	No
Use of variability of local signal	Yes	Yes	Yes	No	Yes	Yes	No
Ranking of peaks							
Binomial test	Yes	No	No	Yes	No	No	No
Poisson test	No	Yes	No	No	Yes	No	No
Normalized difference score	No	No	No	No	No	No	Yes
Use of underlying genome sequence	Yes	No	No	No	No	No	No
Posterior probability of enrichment	No	No	Yes	No	No	Yes	No

Representative selection from over 30 existing tools.

Park J, Nature Reviews Genetics, 10, 669 (2009)
Thomas et al. Brief Bioinform. 18: 441–450 (2017) .

Generate synthetic ChIP-seq data



The process of the chromosomal immunoprecipitation and the subsequent unique mapping and extension of sequence reads can be simulated by randomly placing uniquely mapped sequence tags onto the chromosome, according to certain sampling weight at each nucleotide position.

Zhang et al. PLoS Comput Biol 4, e10000158 (2008)

Comparison of actual and synthetic ChIP-seq data

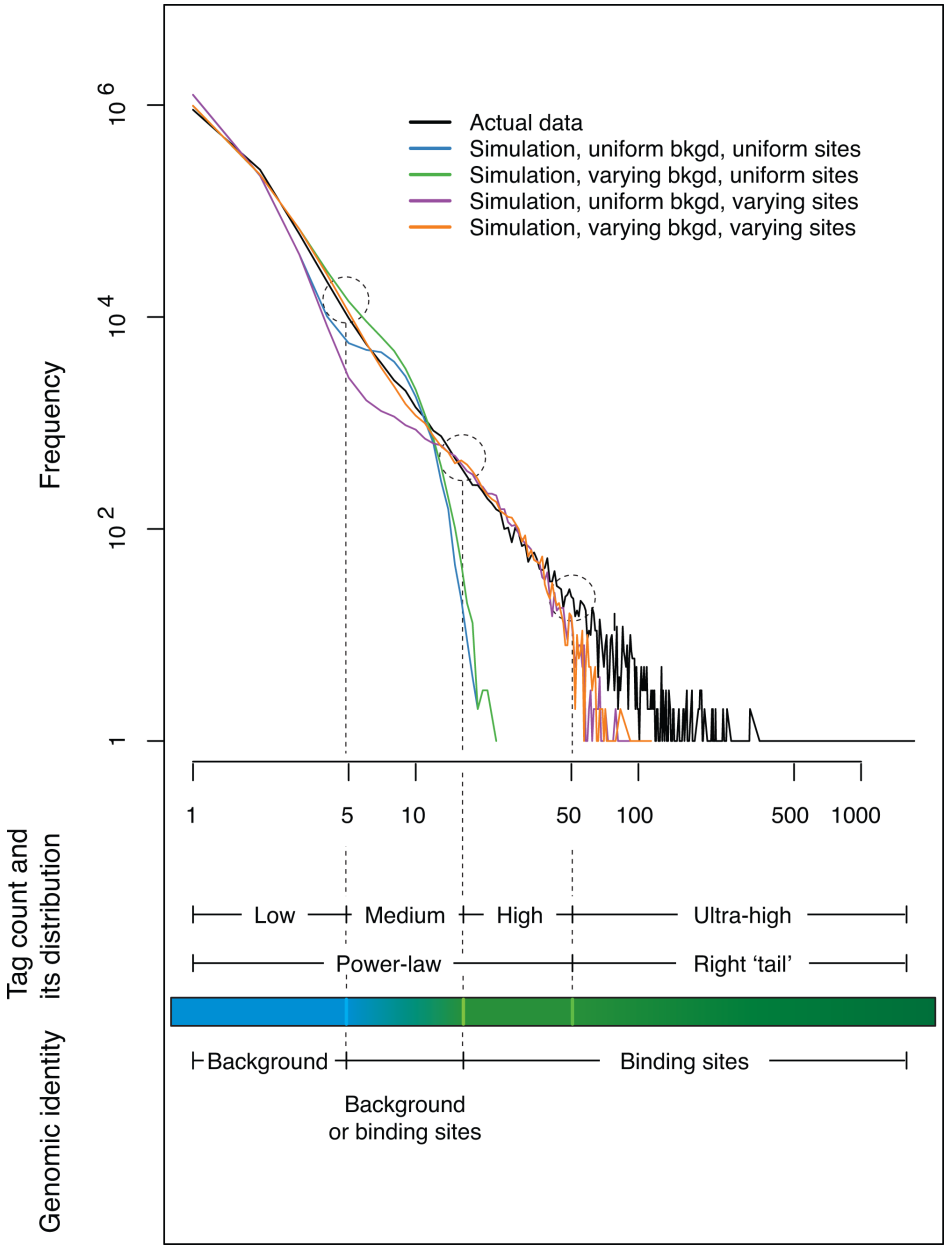


Figure shows the actual tag count distribution and the simulated ones generated under different background and site models.

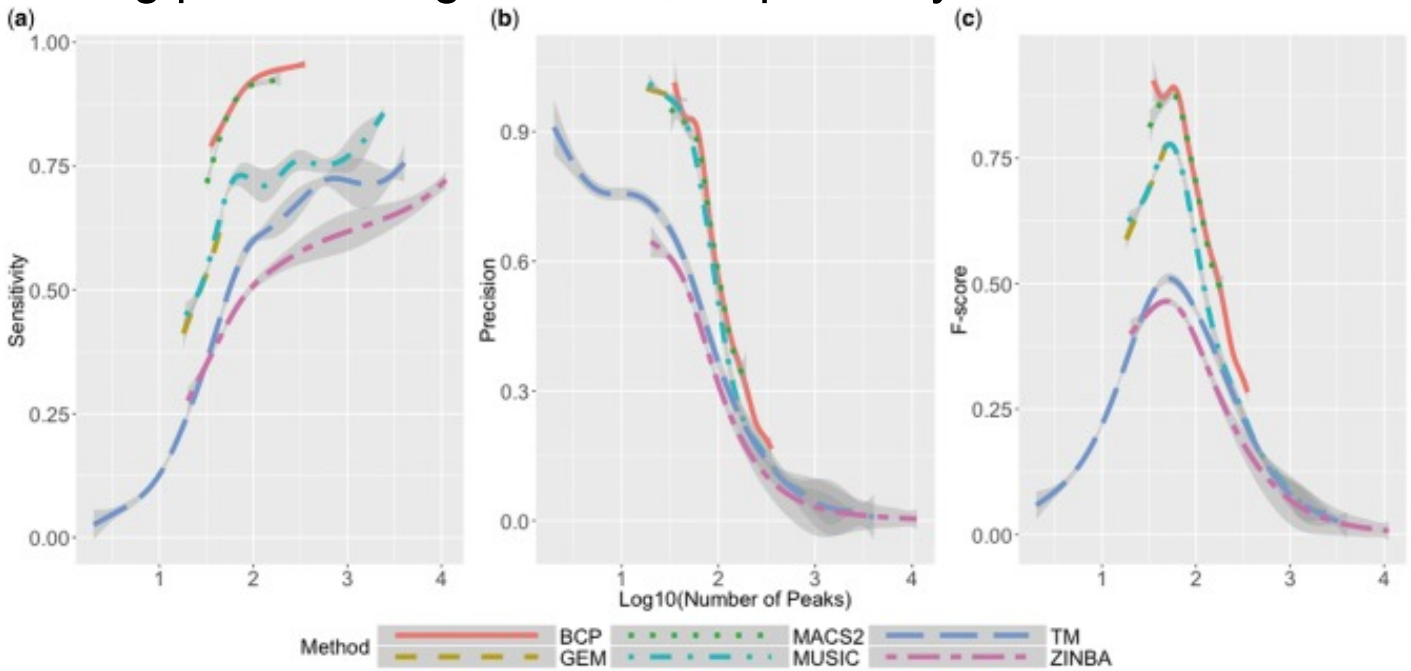
Actual data contains more sites with ultra-high tag counts (right trail of distribution).

Zhang et al. PLoS Comput Biol 4, e1000158 (2008)

Benchmarking of ChIP-seq peak calling

Abstract the peak calling problem into two sub-problems:

- identifying peaks and
- testing peaks for significance, respectively.

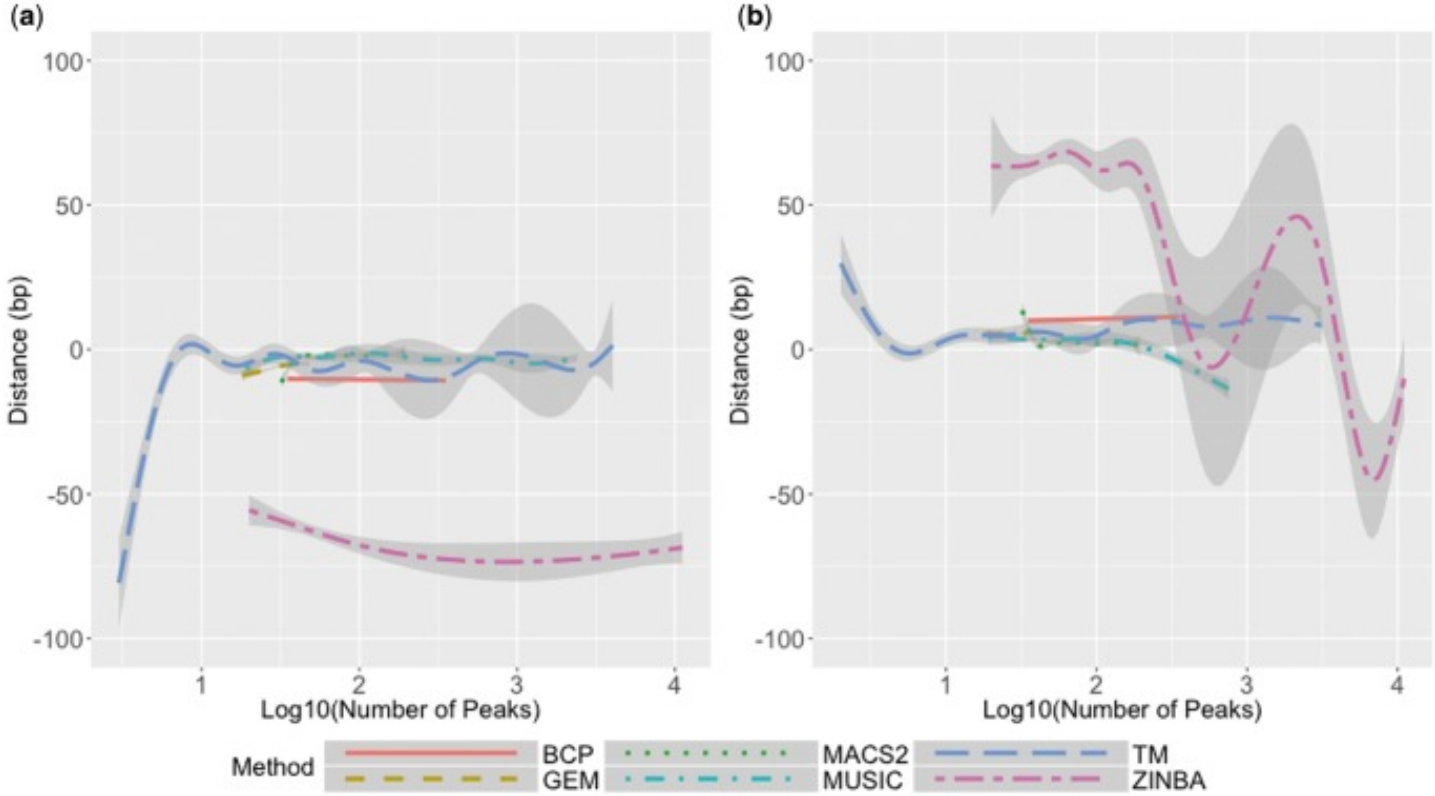


BCP and MACS2 perform best.

Sensitivity (**a**), Precision (**b**) and F-score (**c**) as a function of the $\log_{10}$ of the number of called peaks for 6 peak calling methods on 100 simulated transcription factor ChIP-seq data sets.

Thomas et al. Brief Bioinform. 18: 441–450 (2017) .

Identifcation of the peak position

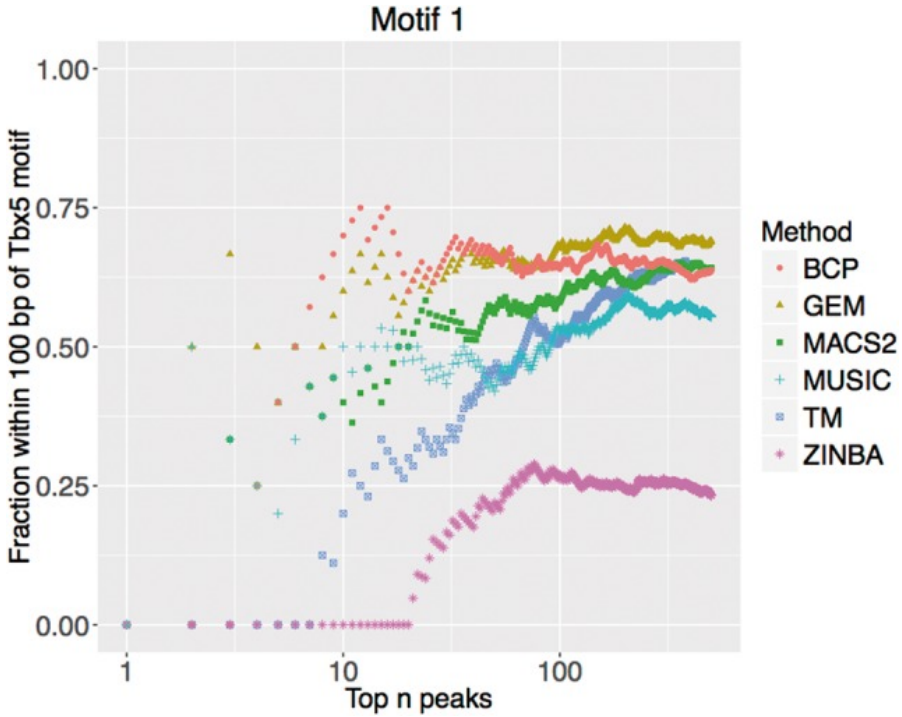


Median distance-binding (**a**) and *Median distance-peak* (**b**) as a function of the $\log_{10}$ of the number of called peaks for the 6 peak calling methods on 100 simulated data sets.

The **peak position** is typically identified quite precisely, except for ZINBA.

Thomas et al. Brief Bioinform. 18: 441–450 (2017) .

Performance on real data from Tbx5 ChIP-seq experiment

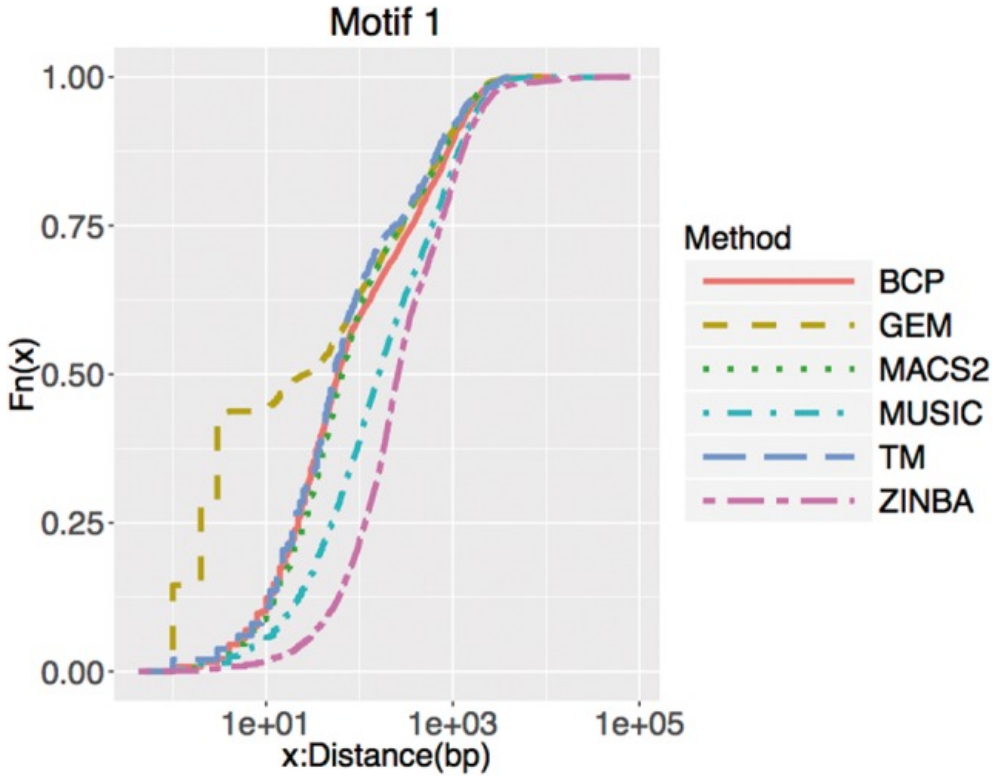


Fraction of top *n* peaks within 100 bp of the Tbx5 motif 1 for the 6 methods.

BCP and GEM perform particularly well = high fraction with 100 bp.

Thomas et al. Brief Bioinform. 18: 441–450 (2017) .

Performance on real data from Tbx5 ChIP-seq experiment



Empirical distribution of the shortest distance to the Tbx5 motif 1 of the significant peaks called by the 6 methods.

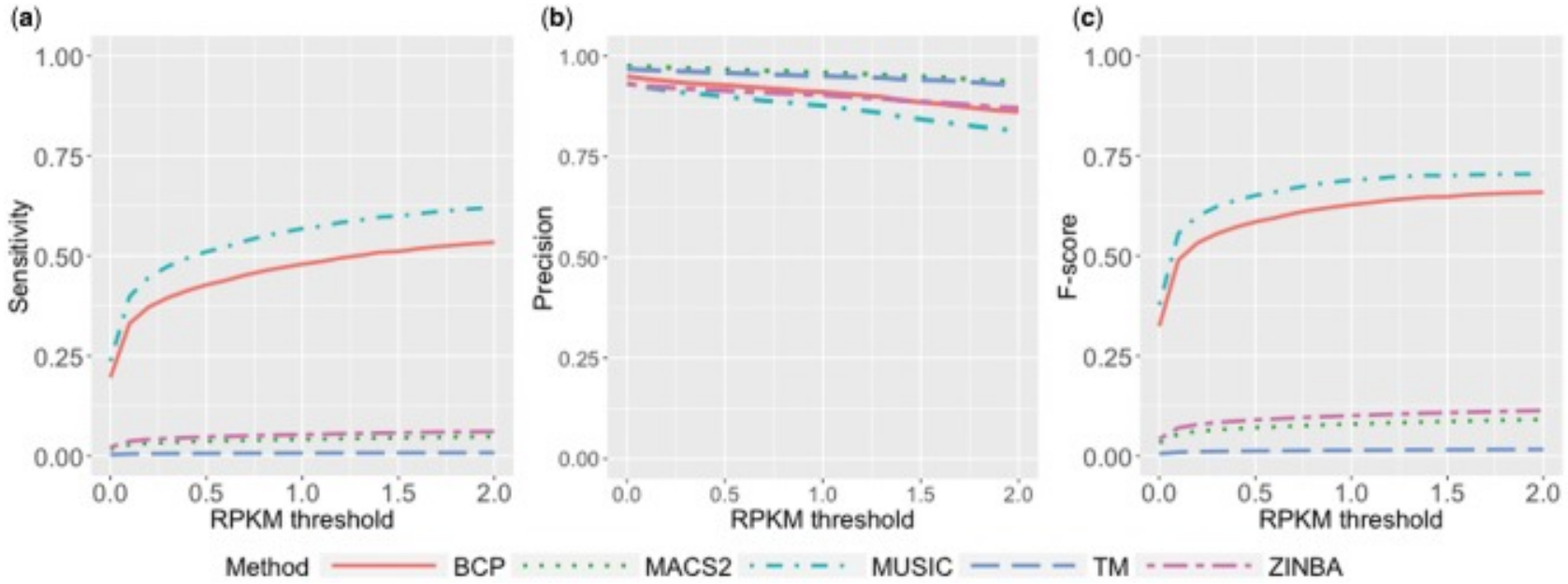
GEM peaks are closer to a Tbx5 motif than any other method.

Thomas et al. Brief Bioinform. 18: 441–450 (2017) .

Benchmarking of ChIP-seq peak calling

Histones typically have wider peaks than TFs.

Test how well H3K36me3 peaks overlap genes that are actively transcribed.



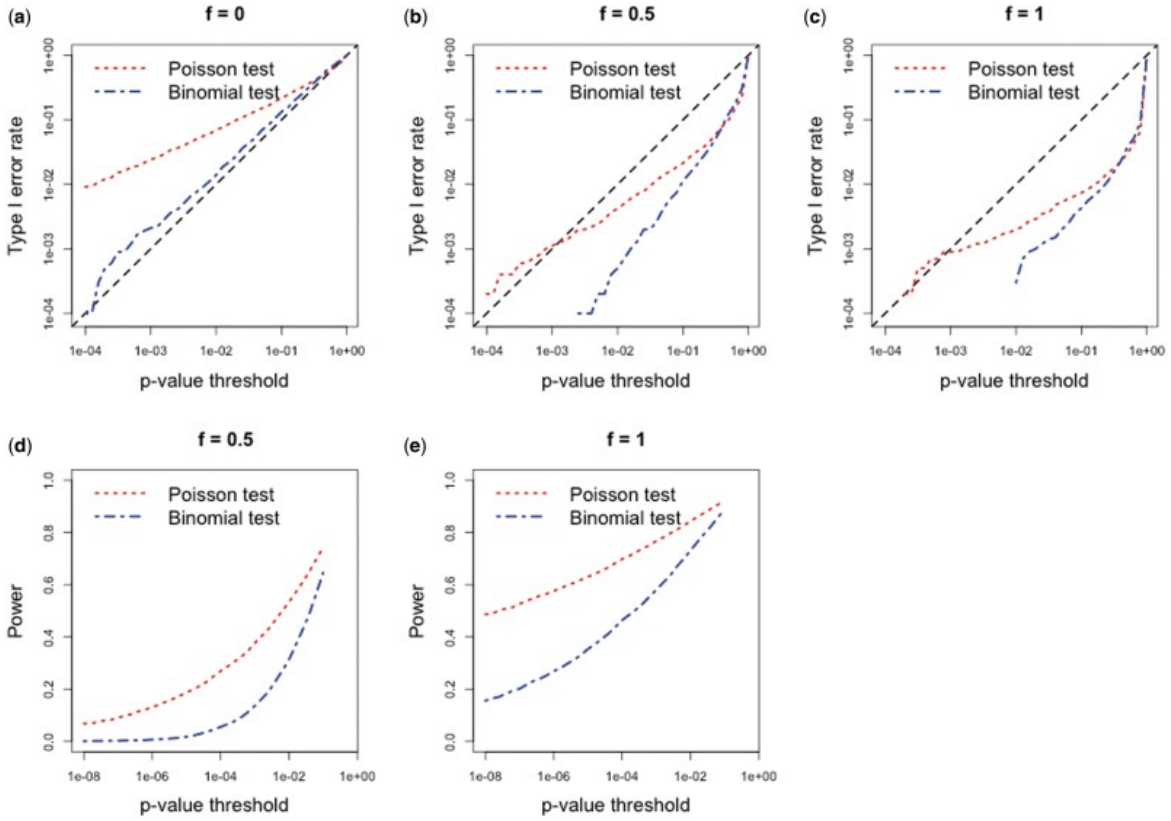
Sensitivity (**A**), precision (**B**) and F-score (**C**) of the overlap of the called significant peak regions with active gene bodies for H3K36me3 data.

The threshold for defining active genes was varied from 0 to 2 RPKM.

MUSIC and BCP perform best.

Thomas et al. Brief Bioinform. 18: 441–450 (2017) .

Benchmarking of ChIP-seq peak calling



BCP and MACS2 give the best performance on simulated data.

Both use the Poisson test, whereas MUSIC and GEM use the Binomial test.

The Poisson test is more powerful in detecting enriched regions!.

Type I error rate and statistical power comparison between Poisson and Binomial tests is given.

f is a parameter that controls the increase in the proportion of DNA from a given region in the input relative to the ChIP sample for the Type I error evaluations, (**A**), (**B**) and (**C**), and increase in this proportion for the ChIP relative to the input sample for the power evaluations ((**D**) and (**E**)).

Thomas et al. Brief Bioinform. 18: 441–450 (2017) .

Benchmarking of ChIP-seq peak calling: key points

Peak calling using Chip-seq data consists of 2 sub-problems: identifying candidate peaks and testing candidate peaks for statistical significance.

Methods that explicitly combine the signals from ChIP and input samples to define candidate peaks are less powerful than methods that do not.

Methods that use windows of different sizes to scan the genome for potential peaks are more powerful than ones that do not.

Methods that use a Poisson test to rank their candidate peaks are more powerful than those that use a Binomial test.

Thomas et al. Brief Bioinform. 18: 441–450 (2017) .

Basics of mass spectroscopy

3 key stages of a basic mass spectrometer (no high-end instrument):

1. Ionization.

Molecules in a sample may be vaporized by heating. Then, an electron beam bombards the vapors, which converts the vapors to ions.

Because mass spectroscopy measures the mass of charged particles, only ions will be detected. Neutral molecules will not be seen.

Ions are created by either adding electrons to a molecule (yields negatively charged ion) or abstracting electrons from a molecule (yields positively charged ion).

2. Acceleration and Deflection.

Next, the ions are sorted according to their mass in 2 stages.

Acceleration is simple Coulombic attraction. The positive ions created in the ionization stage accelerate towards negative plates at a speed dependent on their mass. Lighter molecules move quicker than heavier ones.

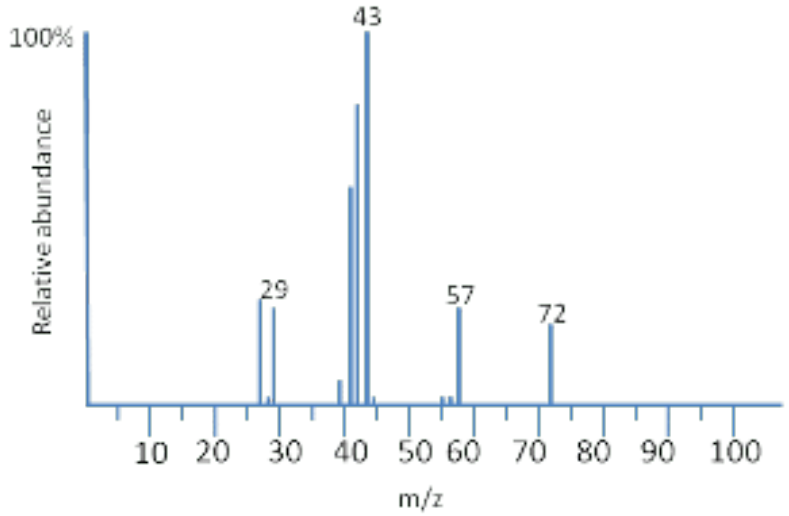
Deflection: the ions are then deflected by a magnetic field. The extent of deflection is again dependent on mass.

https://bitesizebio.com/6016/how-does-mass-spec-work/

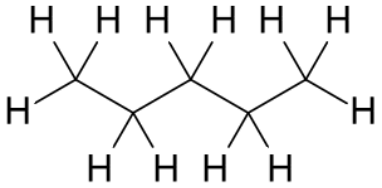
Basics of mass spectroscopy

3. Detection.

Ions of increasing mass eventually reach the detector one after another. This yields a spectrum as shown in the figure.

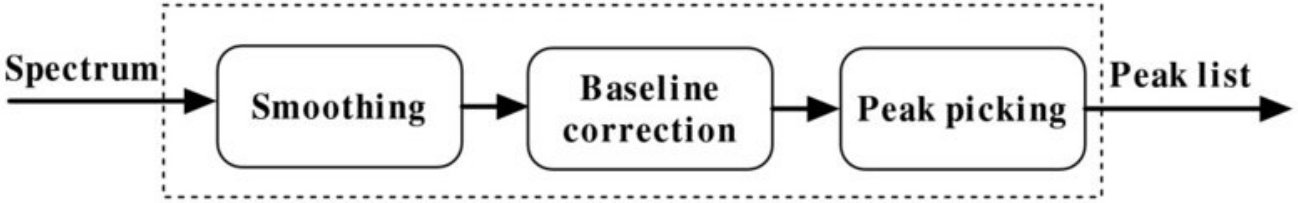


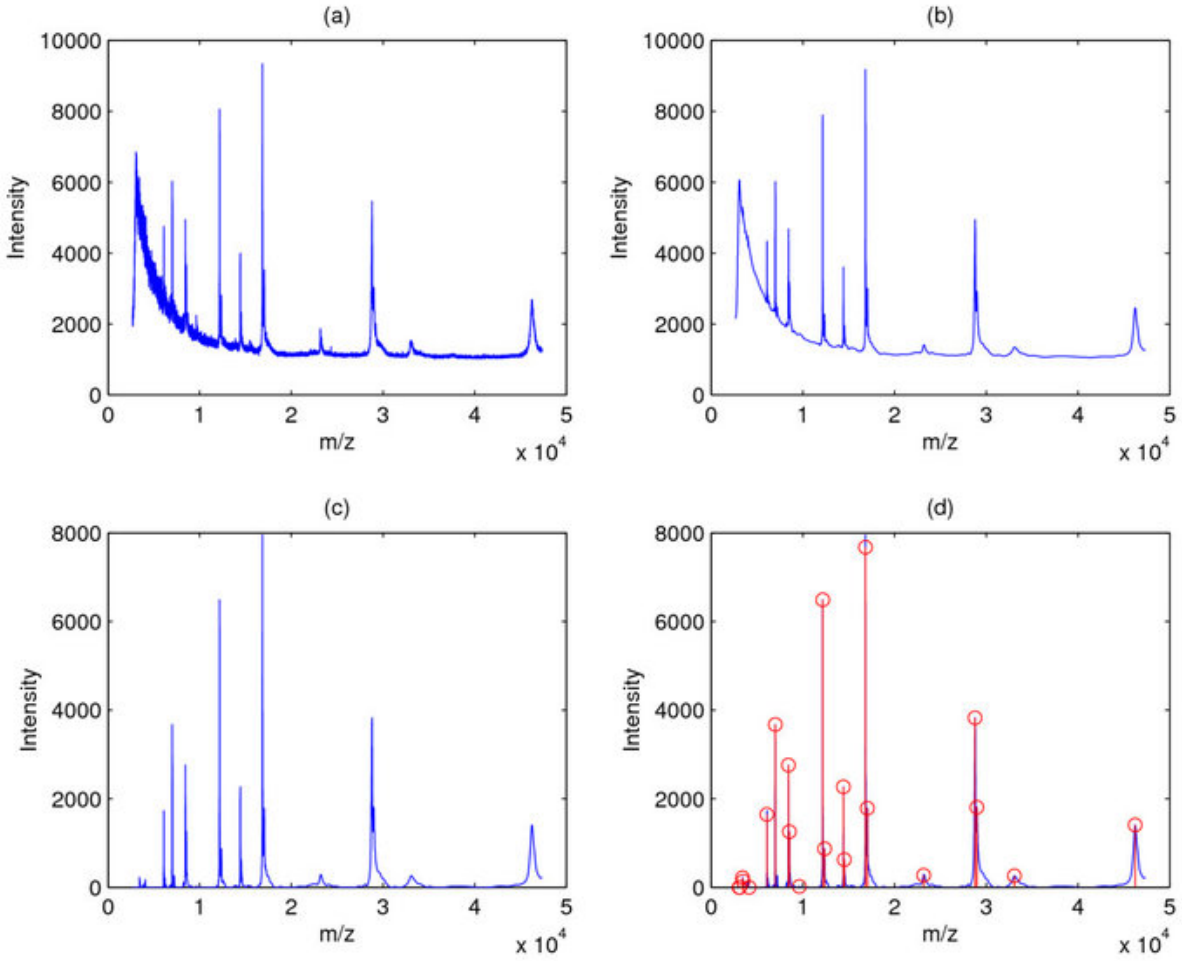
Simplified mass spectrum of **pentane** produced by a mass spectrometer.



https://bitesizebio.com/6016/how-does-mass-spec-work/

Peak detection in MS data: workflow





An example of the peak detection process.

(a) A raw spectrum,

(b) the spectrum after **smoothing**,

(c) the spectrum after smoothing and **baseline correction** and

(d) final peak detection result where **peaks** are marked as circles.

Yang et al. BMC Bioinformatics (2009) **10**:4

S: smoothing strategy
B: baseline correction strategy
P: peak finding strategy

Peak detection in MS data

Table 1: Open source software packages for MS data analysis

Program	S	B	P
Cromwell [12]	S7	B1	P1, P4
LCMS-2D [20]	-	B5	P1, P2
LIMPIC [21]	S4	B2	P1, P3
LMS [22]	S3	B2	P1, P4
MapQuant [16]	S1,S2,S3	-	P7
CWT [10]	S5	B4	P1, P6
msInspect [23]	S6	B2	P5
mzMine [24]	S1, S2	-	P1, P2, P8
OpenMS [15]	S5	B4	P7
PROcess [13]	S1	B2, B3	P1, P2, P5
PreMS [25]	S7	B1	P1, P4
XCMS [8]	S3	-	P1, P4

- Smoothing

S1: Moving average filter

S2: Savitzky-Golay filter

S3: Gaussian filter

S4: Kaiser window

S5: Continuous Wavelet Transform

S6: Discrete Wavelet Transform

S7: Undecimated Discrete Wavelet Transform

- Baseline Correction

B1: Monotone minimum

B2: Linear interpolation

B3: Loess

B4: Continuous Wavelet Transform

B5: Moving average of minima

- Peak Finding Criterion

P1: SNR

P2: Detection/Intensity threshold

P3: Slopes of peaks

P4: Local maximum

P5: Shape ratio

P6: Ridge lines

P7: Model-based criterion

P8: Peak width

Yang et al. BMC Bioinformatics (2009) **10**:4

Peak detection in MS data: smoothing

Aim: remove high-frequency (likely unimportant) variations from the data

Approach: replace current value *x(n)* by an average taken over its neighbor points.

Moving average filter

$$y[n] = x[n] * w[n] = \frac{1}{2k+1} \sum_{i=-k}^{k} x[n-i]$$

2k +1 is the **filter width**

$$w[n] = \frac{1}{2k+1}\ ,\ -k \le n \le k$$

* stands for “convolution”

Gaussian filter

$$y(t) = x(t) * w(t) = \int_{-\infty}^{+\infty} x(\tau) w(t-\tau) d\tau$$

$$w(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}}$$

Yang et al. BMC Bioinformatics (2009) **10**:4

Peak detection in MS data: continuous wavelet transform

CWT

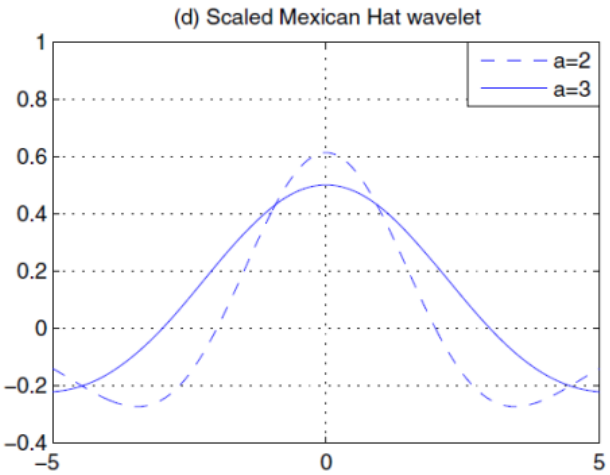
$$y(t) = x(t) * w(t) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} x(\tau)\psi\left(\frac{t-\tau}{a}\right)d\tau$$

$$w(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t}{a}\right)$$

$\psi(t)$ is a **wavelet function**,

e.g. a **Mexican-hat wavelet**
(an inverted parabola, that is squeezed (in the middle) and flattened (at the sides) by multiplication with an exponential function)

$$\psi(t) = \frac{2}{\sqrt{3}\pi^{1/4}}(1-t^2)e^{-t^2/2}$$



Yang et al. BMC Bioinformatics (2009) **10**:4

Peak detection in MS data: peak identification

Signal-to-noise ratio (SNR)

Different methods define noise differently. E.g. noise may be estimated as:

• 95-percentage quantile of absolute continuous wavelet transform (CWT) coefficients of scale one within a local window.

• the median of the absolute deviation (MAD) of points within a window.

Slopes of peaks

This criterion uses the shape of peaks to remove false peak candidates.

- A peak candidate is discarded if both **left slope** and **right slope** are smaller than a threshold.
- This threshold may e.g. taken as half of the local noise level

Yang et al. BMC Bioinformatics (2009) **10**:4

Peak detection in MS data: peak identification

Local maximum

A peak is a local maximum of *N* neighboring points.

Shape ratio

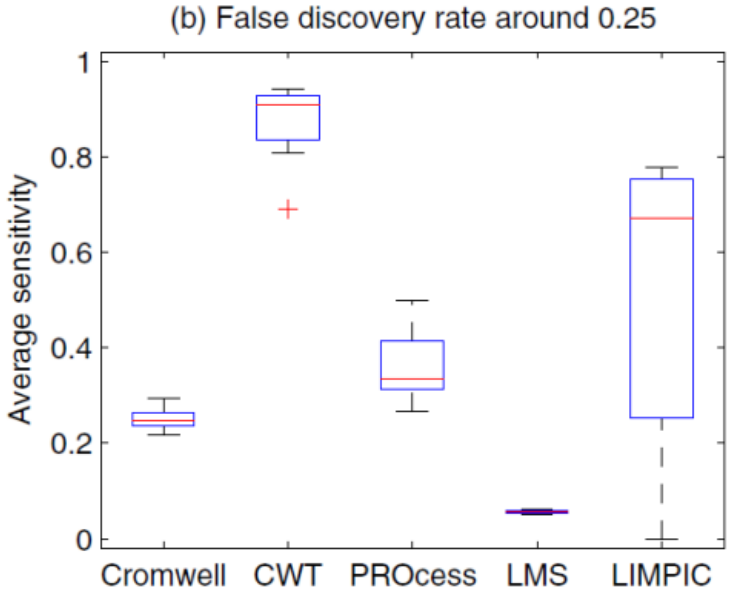
A “peak area” is computed as the area under the curve within a small distance of a peak candidate.

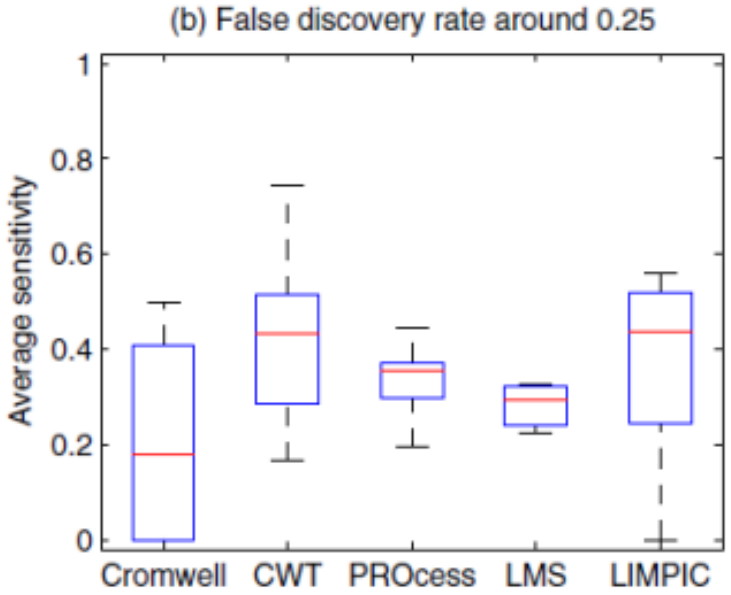
A “shape ratio” is then computed as the peak area divided by the maximum of all peak areas.

The shape ratio of a **peak** must be larger than a threshold.

Yang et al. BMC Bioinformatics (2009) **10**:4

Peak detection in MS data: continuous wavelet transform





Performance on simulated data that was generated using a model that incorporates some characteristics of real MALDI-TOF mass spectrometers.

CWT performed best in this comparison.

The reason is likely that its **shape** matches best the shape of experimental MS peaks.

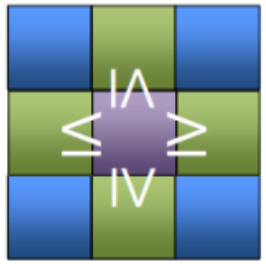
Aurum Dataset is a high resolution data set, which contains spectra from 246 known, individually purified and trypsin-digested protein samples taken with an ABI 4700 MALDI TOF/TOF mass spectrometer.

Yang et al. BMC Bioinformatics (2009) **10**:4

Case study: peak detection in breathomics

2D Peak Finding

- Given $n \times n$ matrix of numbers
- Want an entry not smaller than its (up to) 4 neighbors:



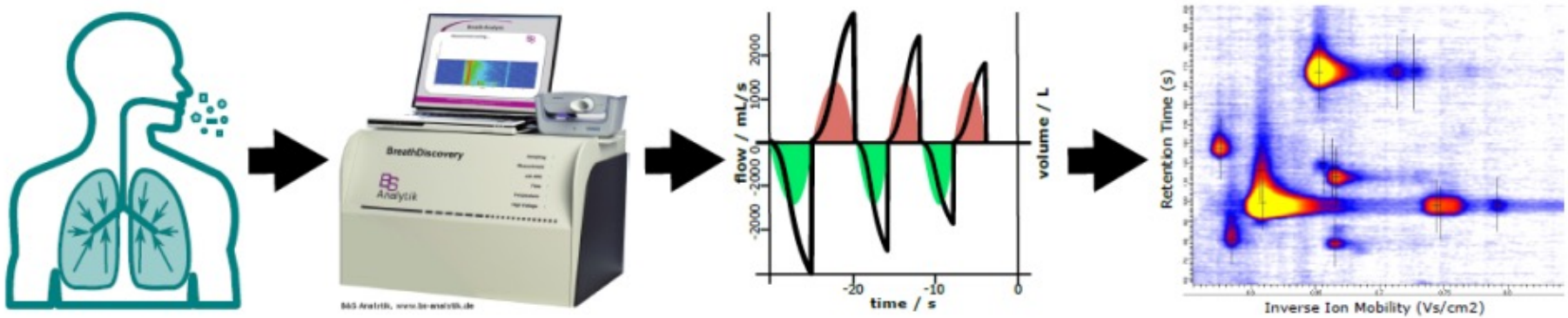
9	3	5	2	4	9	8
7	2	5	1	4	0	3
9	8	9	3	2	4	8
7	6	3	1	3	2	3
9	0	6	0	4	6	4
8	9	8	0	5	3	0
2	1	2	1	1	1	1

https://courses.csail.mit.edu/6.006/spring11/lectures/lec02.pdf

breathomics

MCC/IMS: Ion mobility spectrometry (IMS), coupled with multi-capillary columns (MCCs) is gaining importance for biotechnological and medical applications.

With MCC/IMS, one can e.g. measure the presence and concentration of volatile organic compounds in the air or in **exhaled breath** with high sensitivity.



Kopczynski, Rahmann,
Algorithms for Molecular Biology
(2015) **10**:17
PhD thesis Ann-Christin Hauschild,
Saarland University (2016)

MCC/IMS experiments: output

In an MCC/IMS experiment, a mixture of several unknown volatile organic compounds is separated in two dimensions:

(1) By the **retention time** *r* in the capillary column (the time required for a particular compound to pass through the column). The retention time is proportional to the substance's **affinity** for the stationary phase.

(2) By the **drift time** *d* through the ion mobility spectrometer.

Instead of the drift time itself, one uses a quantity that is normalized for pressure and temperature called the **inverse reduced mobility** (IRM) *t*.

This allows comparing spectra taken under different or changing conditions.

Kopczynski, Rahmann,
Algorithms for Molecular Biology
(2015) **10**:17

MCC/IMS experiments: inversed reduced mobility

the reduced mobility of an ion drifting through a buffer gas in an electric field is given by

$$K = (3q/16N)(2\pi/\mu kT)^{1/2}(1/\Omega_D) \qquad (1)$$

where q is the charge of the ion and m its mass, N is the density of the neutral molecules and M their mass, μ is the reduced mass $\mu = mM/(m + M)$, k is the Boltzmann constant, T is the effective temperature, and Ω_D is the collision cross section.

From K, one derives the reduced (normalized) ion mobility:

$$K_0 = K(273/T)(P/760)$$

and the **inversed reduced ion mobility** (after some rearrangement)

$$K_0^{-1} = 1.697 \times 10^{-4}(\mu T)^{1/2}\Omega_D$$

Karpas et al. JACS 111, 6015 (1989)

IM spectrum-chromatogram

r : set of (equidistant) **retention time** points

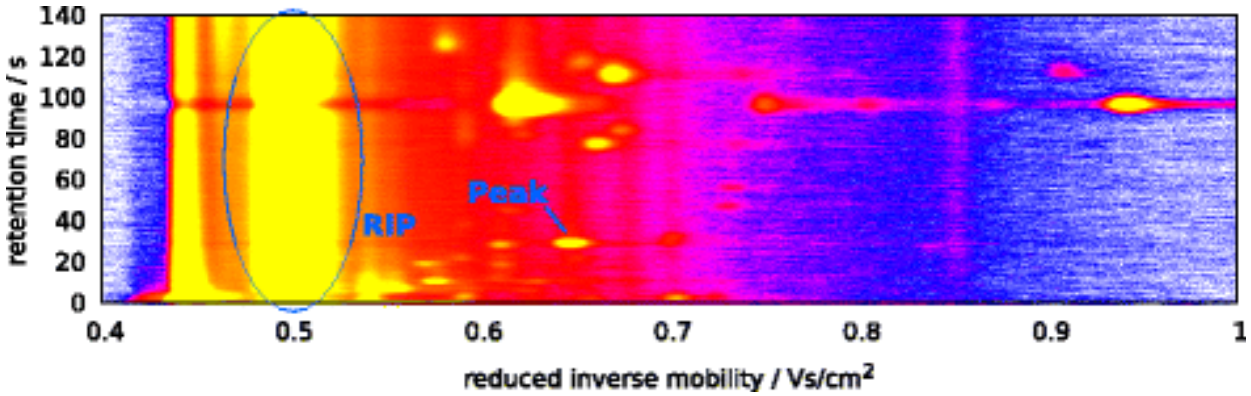
t : set of (equidistant) **IRMs** where a measurement is made,

e.g. 12500 time points taken every 4 x 10^{-6} s -> 50 ms in total)

Then the data is an $|r| \times |t|$ matrix of measured ion intensities,

which we call an *IM spectrum-chromatogram* (IMSC).

The matrix can be visualized as a **heat map**.

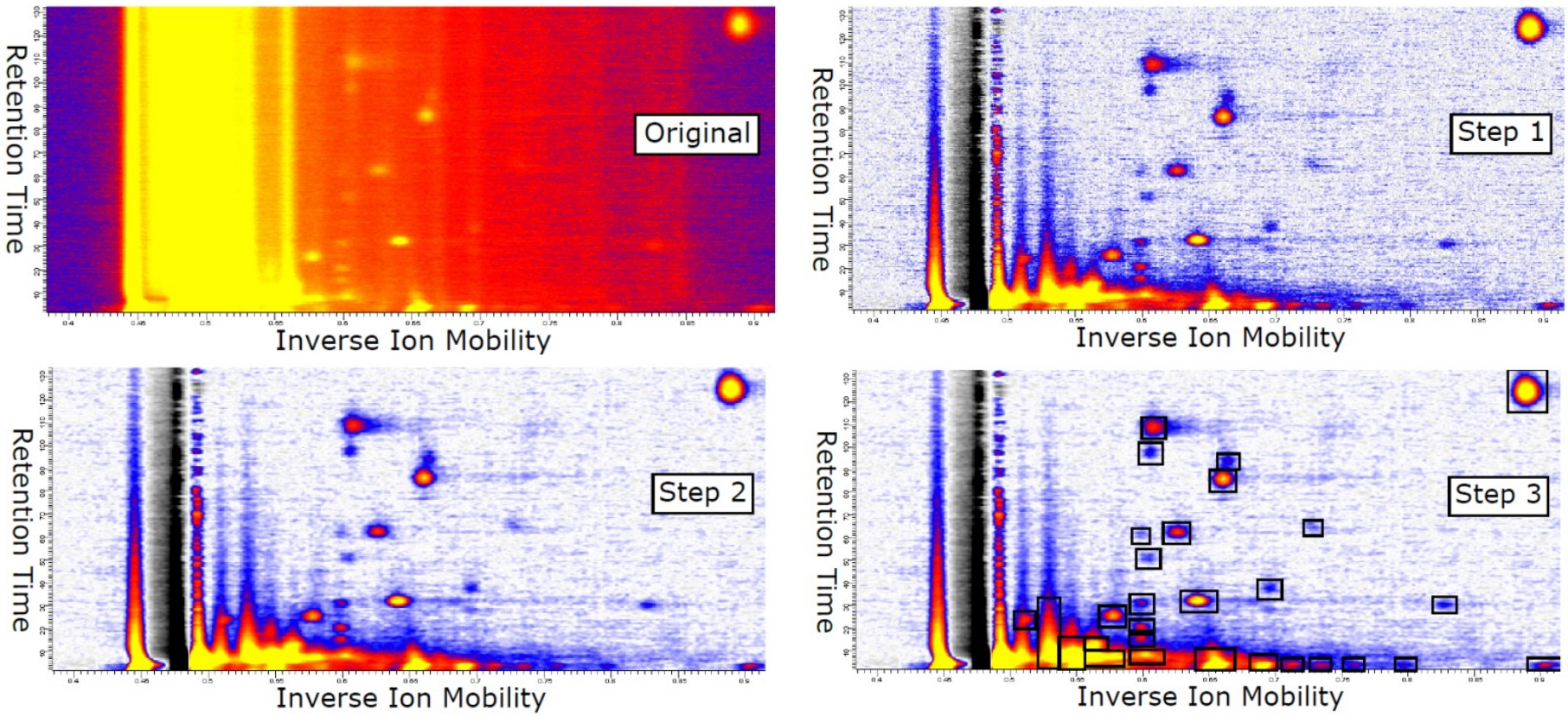


An IM spectrometer uses an ionized **carrier gas**. These ions are present in every spectrum in addition to the analyte ions, and they create the **reactant ion peak (RIP)**.

The reduced inverse ion mobility (x-axis) is proportional to the drift time.
The colors reflect the signal height:
[white (low) < blue < purple < red < yellow (high signal)].

Kopczynski, Rahmann,
Algorithms for Molecular Biology
(2015) **10**:17

breathomics



Example of a processing strategy of MCC/IMS data involving

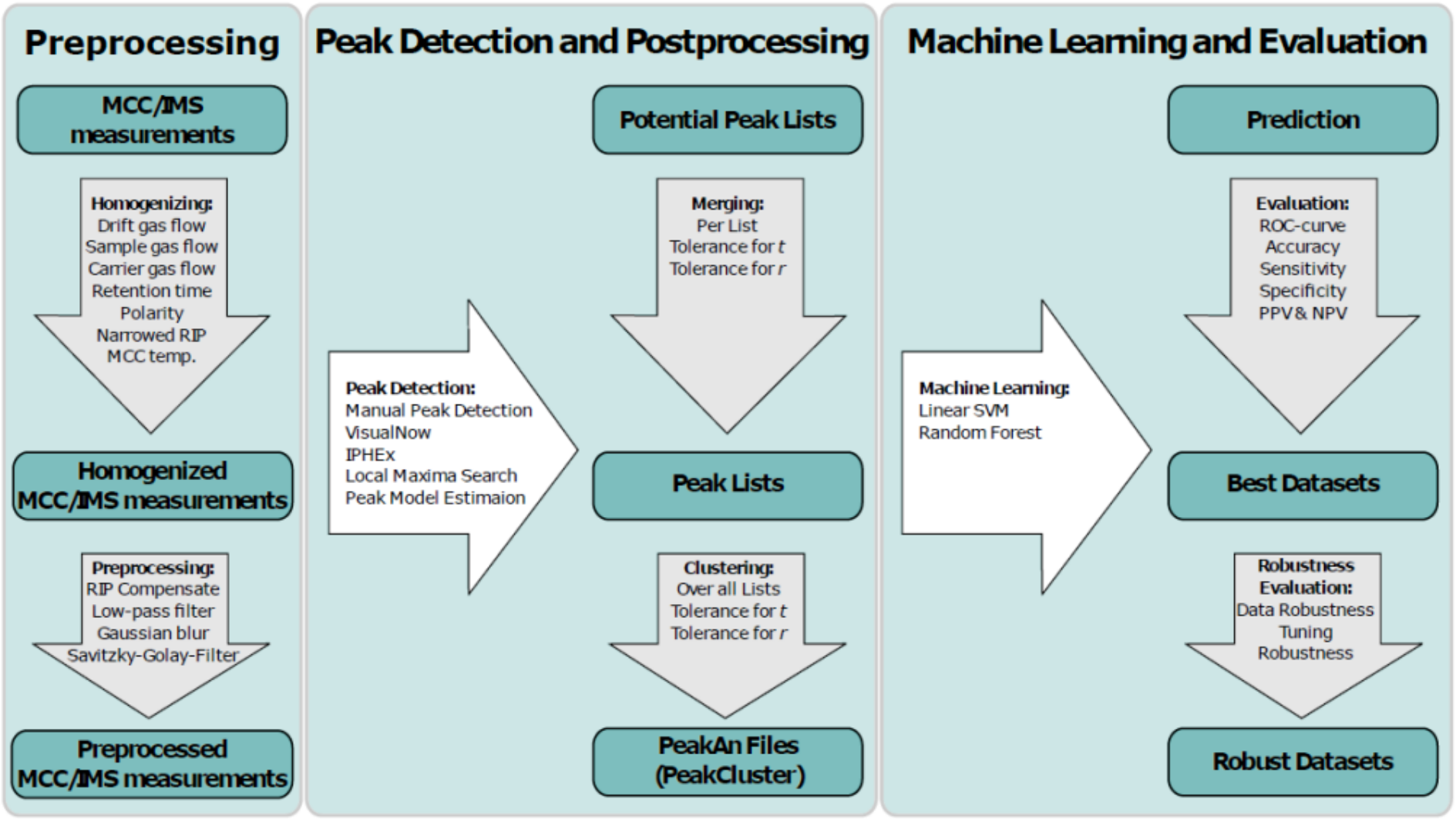
(Step 1) RIP-detailing (removal of RIP peak)

(Step 2) denoising and baseline correction

(Step 3) peak picking.

PhD thesis Ann-Christin Hauschild, Saarland University (2016)

Breathomics Work flow



PhD thesis Ann-Christin Hauschild, Saarland University (2016)

Manual Peak detection

The easiest and most intuitive way of peak detection is **manual evaluation** of a visualization of the measurement.

The human eye and visual cortex is optimized for pattern recognition in 3D.

Therefore one can immediately spot most of the peaks in the measurement.

There are several **drawbacks** of the manual approach:

- it is **time consuming** and therefore inappropriate in a high-throughput context,

- the results depend on a **subjective** assessment, and are therefore hardly reproducible.

Nevertheless, manual evaluation is still the state of the art for the evaluation of smaller MCC/IMS data sets.

Manually created peak lists are used as “**gold standard**" in MCC/IMS studies.

PhD thesis Ann-Christin Hauschild, Saarland University (2016)

Local maxima search

According to this criterion, a point is a **local maximum** if all 8 neighbors in the matrix have a lower intensity than the intensity at the central point.

We call the neighborhood of a point “significant" if

- its own intensity,
- the intensity of its 8 neighbors, and
- that of A additional adjacent points (e.g. A = 2),

lie above a given intensity threshold I.

PhD thesis Ann-Christin Hauschild, Saarland University (2016)

Merged peak cluster localization (MPCL)

The MPCL consists of two phases: (1) clustering and (2) merging.

(1) each data point in the chromatogram is assigned to one of 2 classes, either **peak** or **non-peak**.

For this, one uses a clustering method that is based e.g. on the Euclidean distance metric of the intensity values.

(2) neighboring data points that are both labeled as **peak** can be assumed to belong to the same peak and are **merged together**.

(3) each peak of the analyzed measurement is characterized by its **centroid point**, i.e. the data point, which has the smallest mean distance to all other points in this peak region.

PhD thesis Ann-Christin Hauschild, Saarland University (2016)

Watershed algorithm

Here, the IMS chromatogram is treated like a **landscape** including hills and valleys.

The algorithm starts with a water level above the highest intensity followed by a continuous lowering of the level while uncovering more and more of the local maxima.

In each step, the new uncovered data points are annotated by the label of adjacent labeled neighbors. Those data points that remain unlabeled are identified as a new peak and receive a new label.

The highest data point among a set of new labeled positions denotes the **peak** coordinate.

The algorithm stops if all data points are labeled or the level drops below a given threshold.

PhD thesis Ann-Christin Hauschild, Saarland University (2016)

Watershed algorithm: implementation

The watershed algorithm can be implemented as a **priority queue** to sort all data points.

(1) The largest data point is extracted and labeled first.

(2 - n) This is followed by the next largest point in the queue and so on.

- Each point drawn out of the queue is compared with its neighbors.
- If the neighbors are of equal or larger value, the extracted point is given the same label as its largest neighbor.

(comment: if of equal value, neighbor has not necessarily been labeled …)

- In contrast, if the data point is larger than its neighbors (i.e. the neighbors have not been labelled sofar), the data point is given a new label to indicate that it is part of another peak.

(n + 1) This procedure is repeated until the queue is empty.

Latha et al. Journal of Chromatography A, 1218 (2011) 6792– 6798

Peak model estimation

In the PME method, the expectation maximization (EM) algorithm is used to optimize the parameters of a mixture model from a given set of starting values.

The algorithm requires a given set of “seed" coordinates for each peak to be modeled.

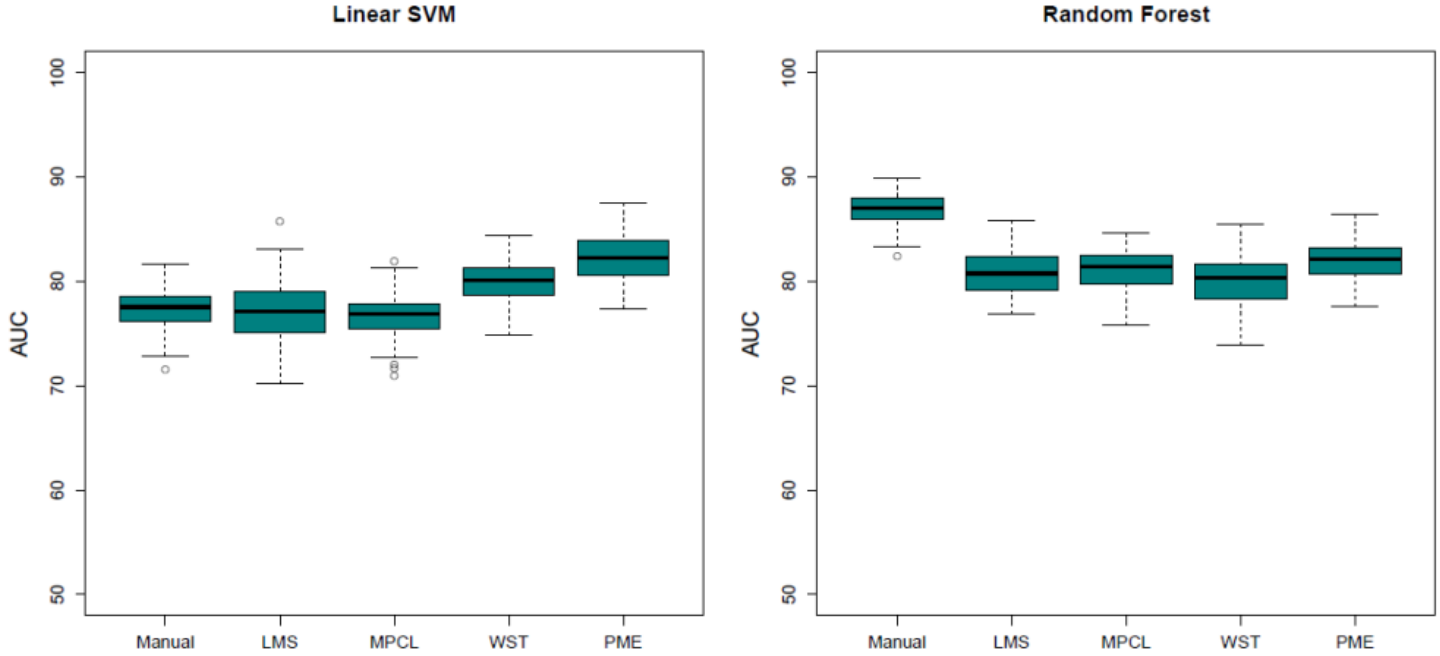
In general, any peak detection method is suitable to provide these initial " seeds". However, the quality of the results strongly depends on the chosen seed-ing approach.

Utilizing the EM algorithm, each peak is described by a model function consisting of two shifted Gaussian distributions and an additional peak volume parameter.

Finally, the set of model functions plus a noise component describe the whole MCC/IMS measurement.

PhD thesis Ann-Christin Hauschild, Saarland University (2016)

breathomics



Boxplots of 100 runs of the ten-fold CV for the linear SVM and the random forest method.

LMS : Automated **local maxima search**

WST : Automated peak detection via **water shed transformation** implemented in IPHEx,

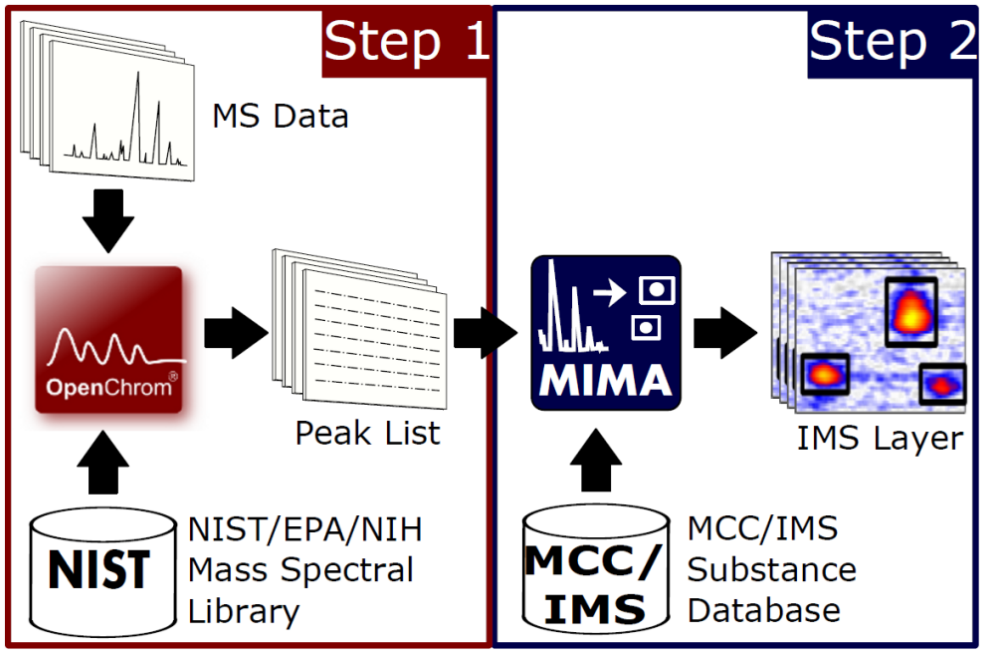
MPCL : Automated peak detection via **merged peak cluster localization** supported by VisualNow

PME : **Peak model estimation** approach by the PeaX tool.

PhD thesis Ann-Christin Hauschild, Saarland University (2016)

Automated metabolite detection

Aim: annotate peaks to chemicals (not only detecting peaks)



Collect **reference IMS data** for compound library

Run IMS experiment on sample of interest - compare against reference data

PhD thesis Ann-Christin Hauschild, Saarland University (2016)

Proof of principle

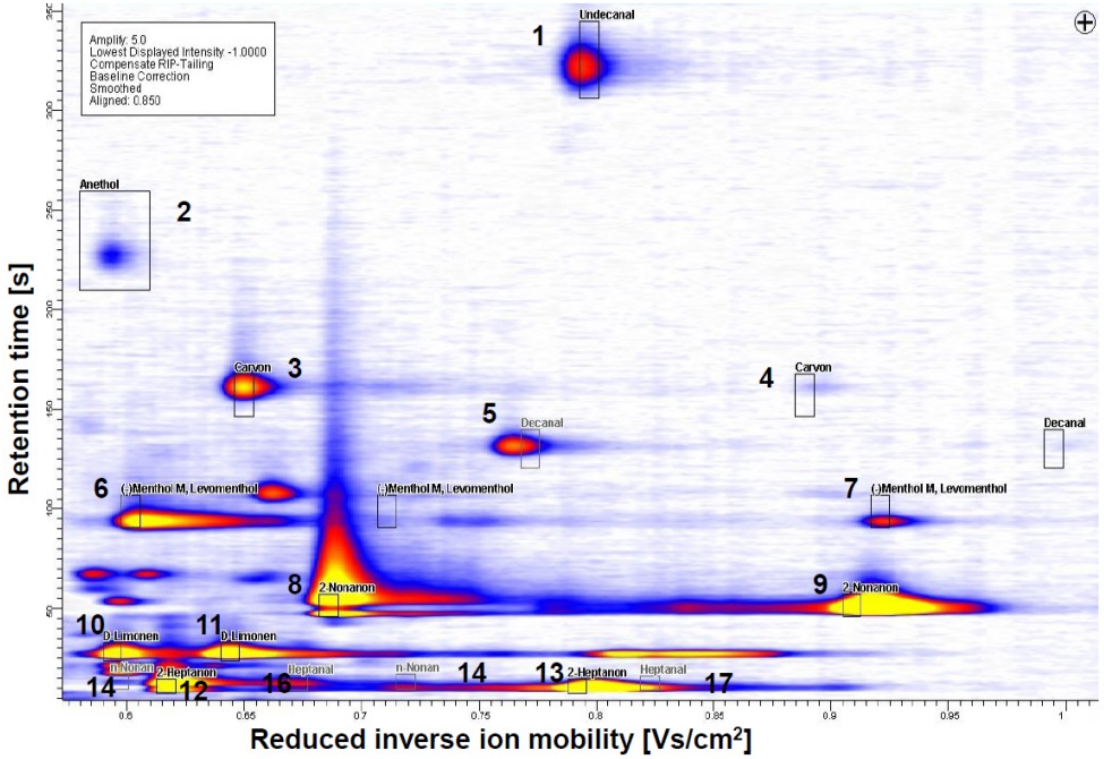


Table 7.1: Automatically identified signals

No.	CAS	compound
1	112-44-7	undecanal
2	104-46-1	anethol (trans-anethol)
3	6485-40-1	carvon (monomer)
4	6485-40-1	carvon (dimer)
5	112-31-2	decanal
6	2216-51-5	(-)menthol (monomer)
7	2216-51-5	(-)menthol (trimer)
8	821-55-6	2-nonanon (monomer)
9	821-55-6	2-nonanon (dimer)
10	5989-27-5	D-limonen (monomer)
11	5989-27-5	D-limonen (dimer)
12	110-43-0	2-heptanon (monomer)
13	110-43-0	2-heptanon (dimer)
14	111-84-2	n-nonan (monomer)
15	111-84-2	n-nonan (dimer)
16	111-71-7	heptanal (monomer)
17	111-71-7	heptanal (dimer)

Test on a mixture of 7 reference compounds

17 signals in the measurement could be matched

12 of the 17 signals originate from the reference compounds

(including dimers and trimers)

PhD thesis Ann-Christin Hauschild, Saarland University (2016)

Application: can one detect COPD in exhaled breath?

Chronic obstructive pulmonary disease (COPD) is an umbrella term used to describe chronic lung diseases that cause a permanent blockage of airflow from the lungs, which is not fully reversible (WHO).

The most prominent symptoms are

- breathlessness,
- a chronic cough, and
- excessive sputum production.

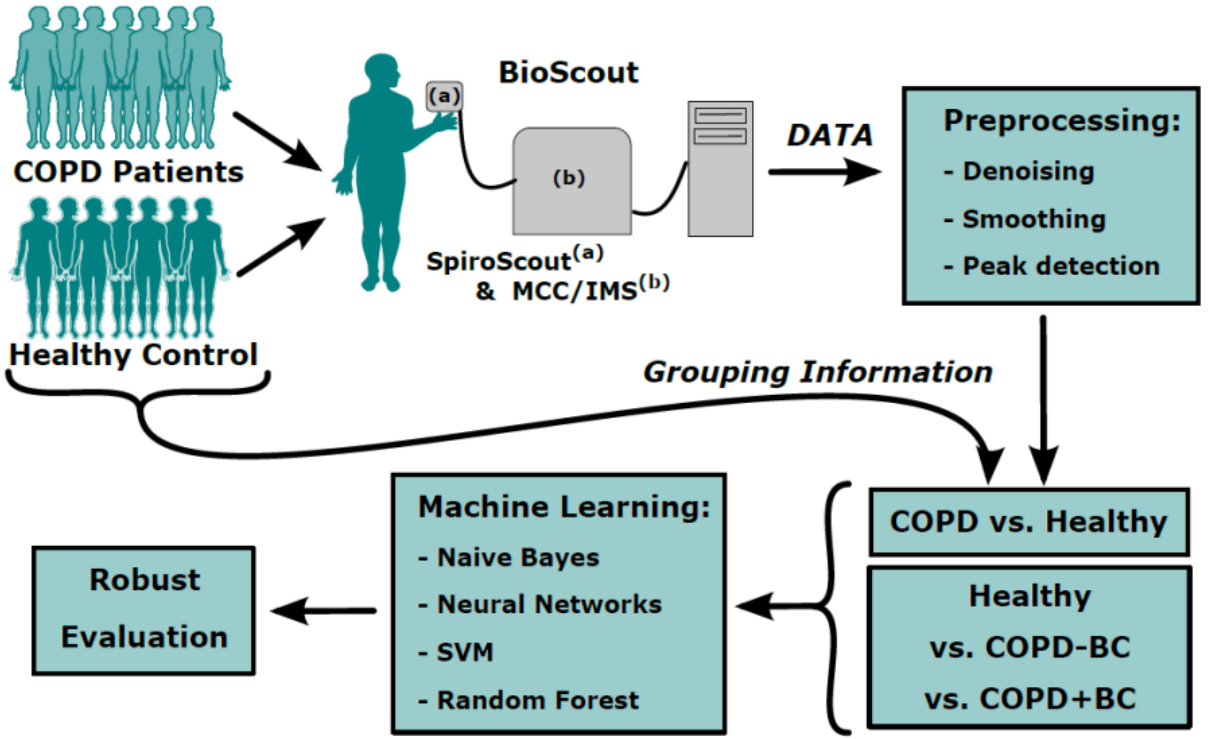
Airways and lungs react to noxious particles or gases, like smoke from cigarettes or fuel, with an increased inflammatory response.

The World Health Organization (WHO) reported COPD as one of the four most frequent causes of death.

PhD thesis Ann-Christin Hauschild, Saarland University (2016)

Application: can one detect COPD in exhaled breath?

Westhoff et al. (2011) took MCC/IMS breath probes of 42 COPD patients and of 35 healthy volunteers (HC).

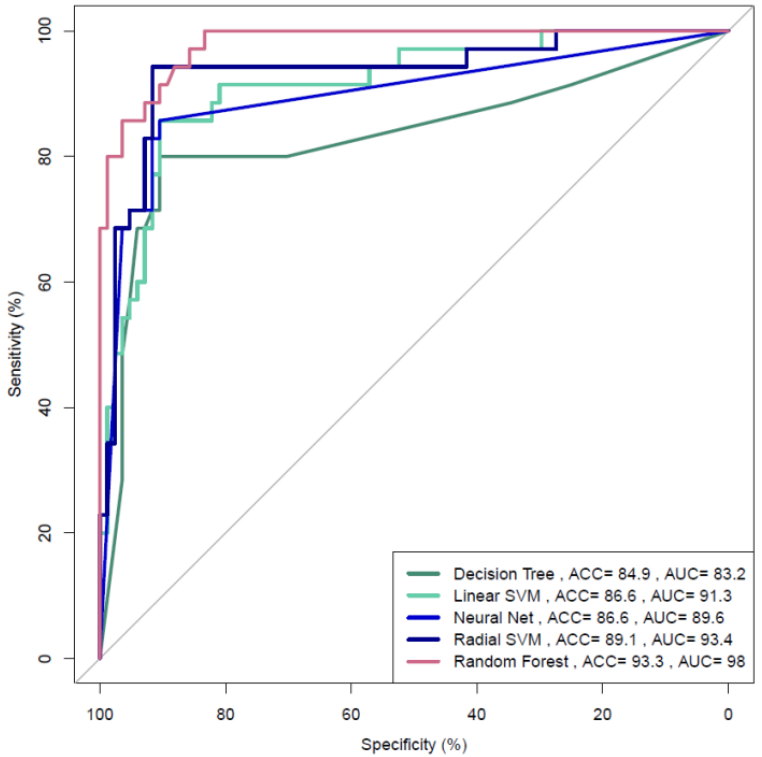


PhD thesis Ann-Christin Hauschild, Saarland University (2016)

Application: can one detect COPD in exhaled breath?

Table 5.1: Results of the two-class-classification problem, evaluating the differences between COPD and the HC.

Method	AUC	Accuracy	Sensitivity	Specificity
Decision Tree	81	85	91	71
Linear SVM	83	87	92	74
Naive Bayes	79	82	87	71
Neural Net	86	89	93	80
Radial SVM	87	89	92	83
Random Forest	92	94	98	86



Distinguishing COPD patients from healthy controls based on IMS spectra of exhaled air works really well!

Distinguishing COPD patients from patients that also have breast cancer did not work equally well.

PhD thesis Ann-Christin Hauschild, Saarland University (2016)

Summary

Peak detection is a frequent task in diverse areas of biology.

The challenge is posed by the noisy nature of biological data and the irregular shape of peaks.

Testing and benchmarking of methods is typically done with synthetic (artificially generated) data.

Peak detection and judging their significance are equally important tasks.