Analysis of changes in microtubule network morphology and localization after docetaxel treatment of sensitive and resistant gastric cancer cells

Alexandre Matov

3/2014

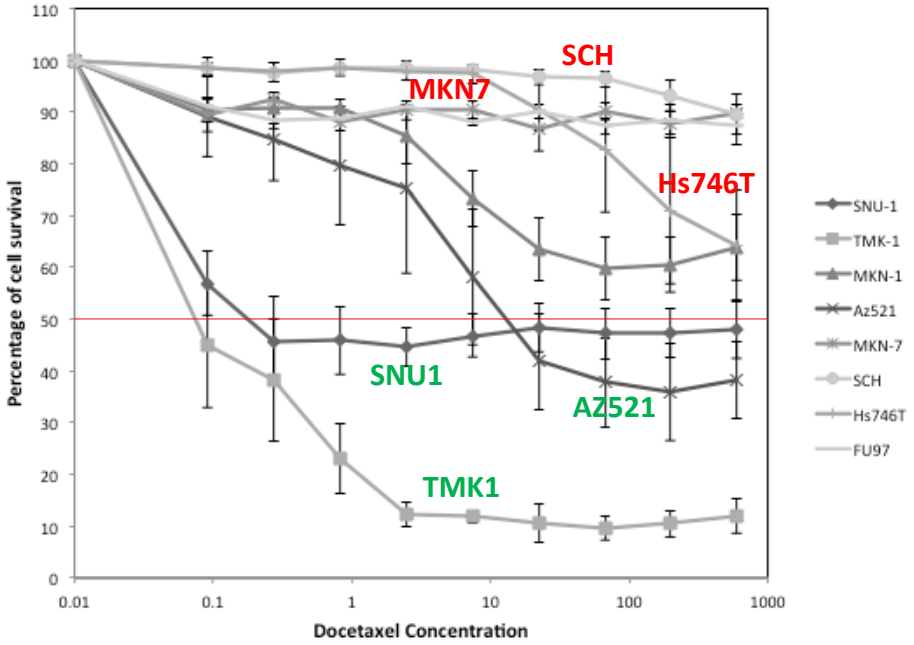
Cell Lines and Treatment

- **Cell Line : SCH – resistant (30 cells)**
- **Cell Line : HS746T – resistant (30 cells)**
- **Cell Line : MKN7 – resistant (30 cells)**
- **Cell Line : TMK1 – sensitive (50 cells)**
- **Cell Line : SNU1 – sensitive (30 cells)**
- **Cell Line : AZ521 – sensitive (30 cells)**

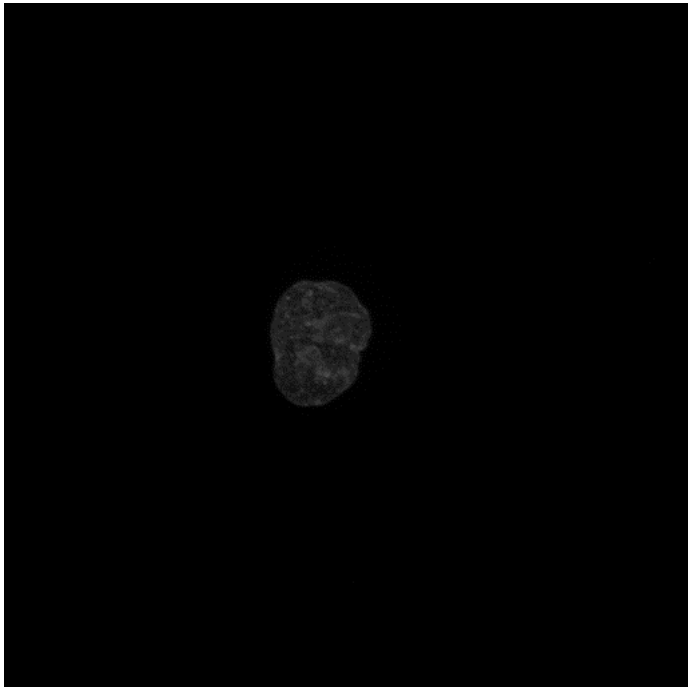
Imaging MTs before and after treatment with 100nM DTX for each cell line

920 high magnification (63x) images in total for MT analysis (together with DAPI)

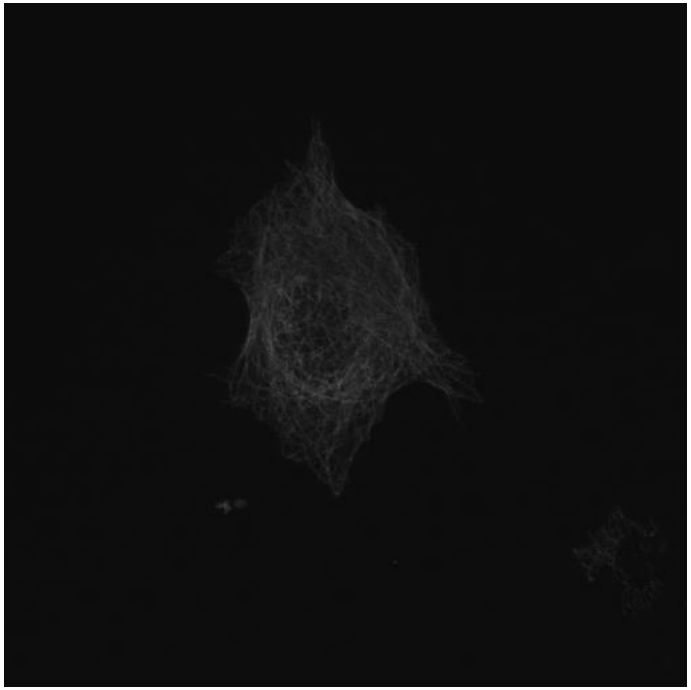
Diffuse Gastric Cancer cell lines (above red line are resistant, below red line are sensitive)



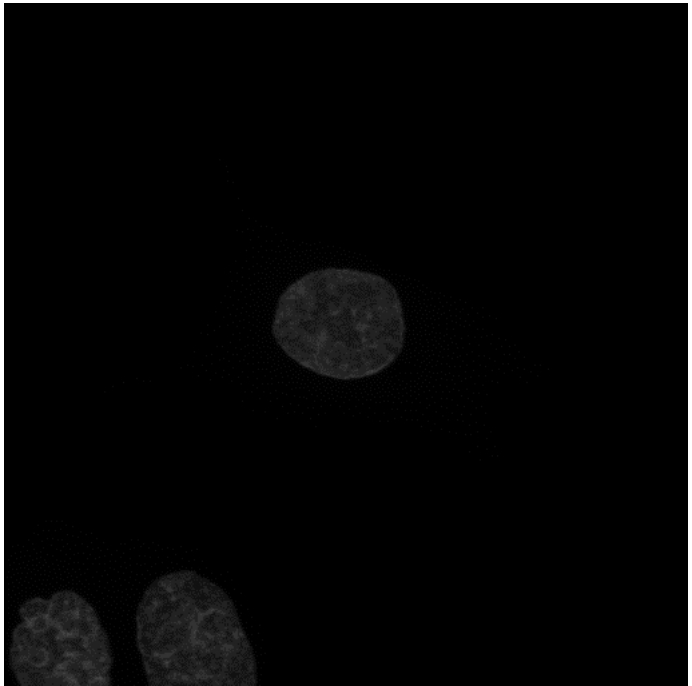
Sensitive TMK1 Non-Treated Cells

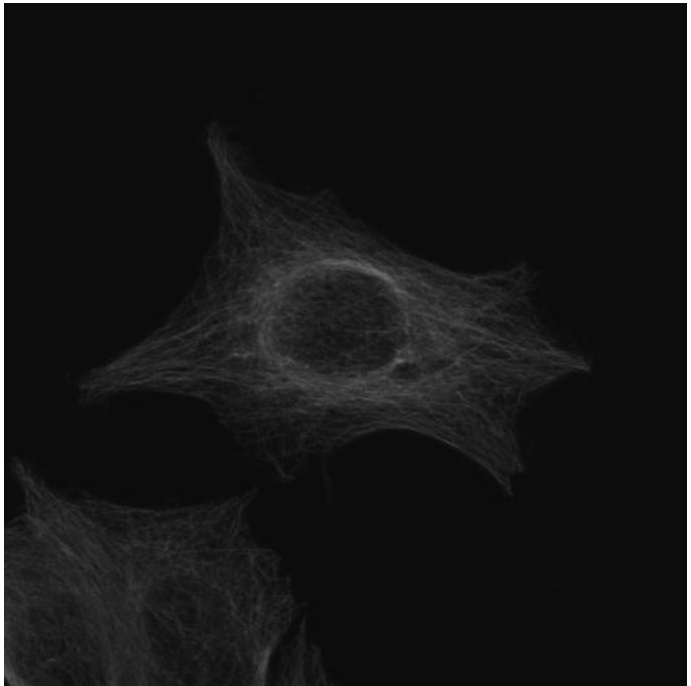


DAPI

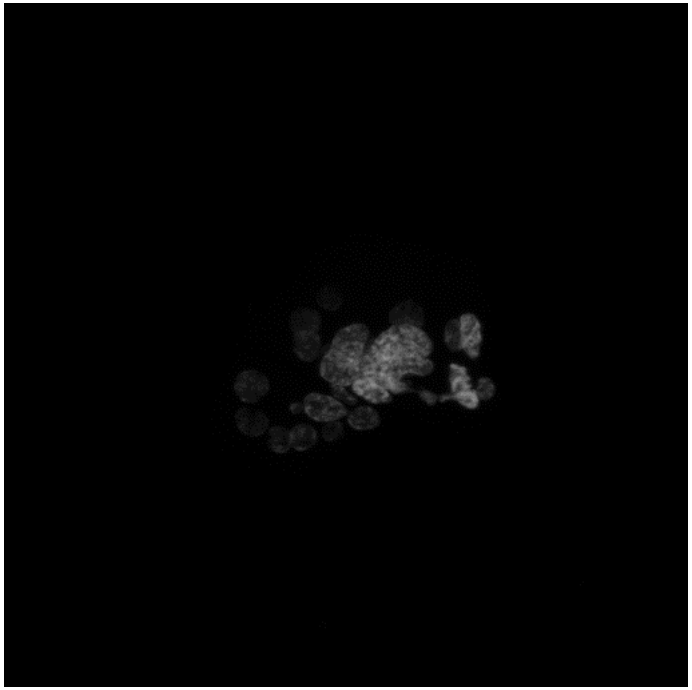


MTs

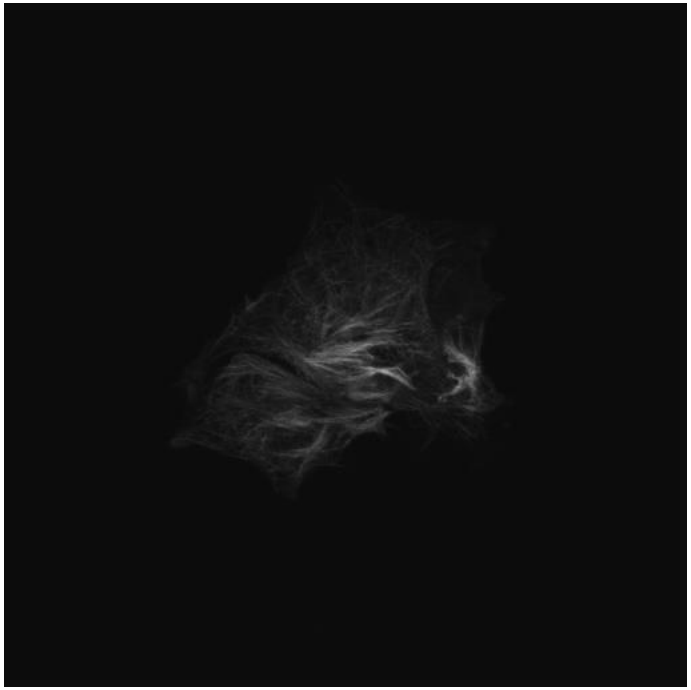




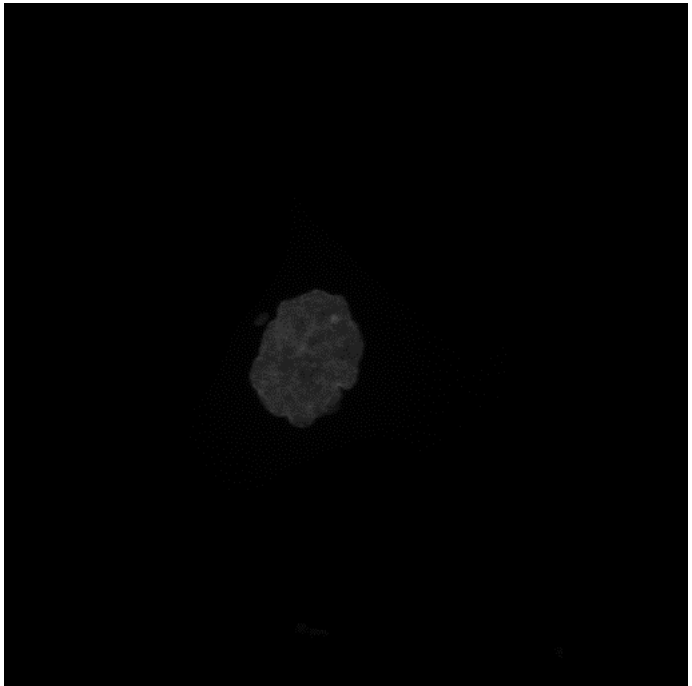
Sensitive TMK1 100nM DTX Treated Cells

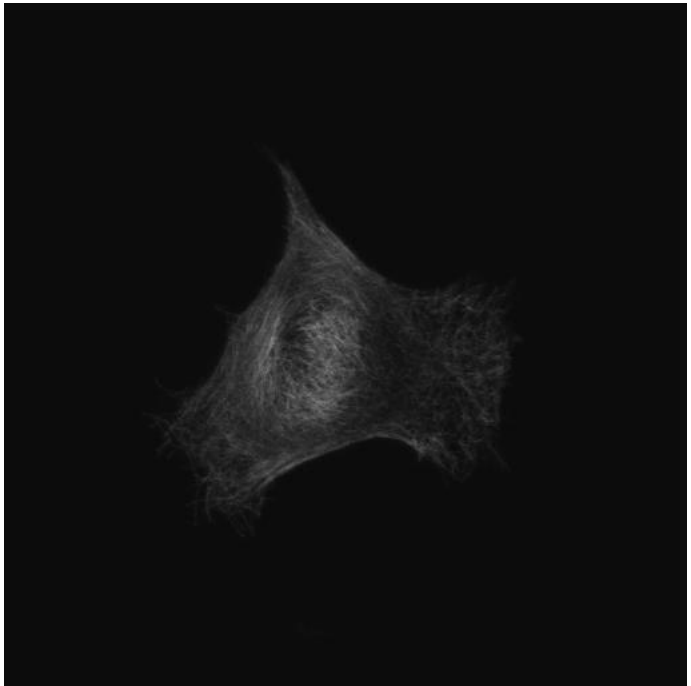


DAPI

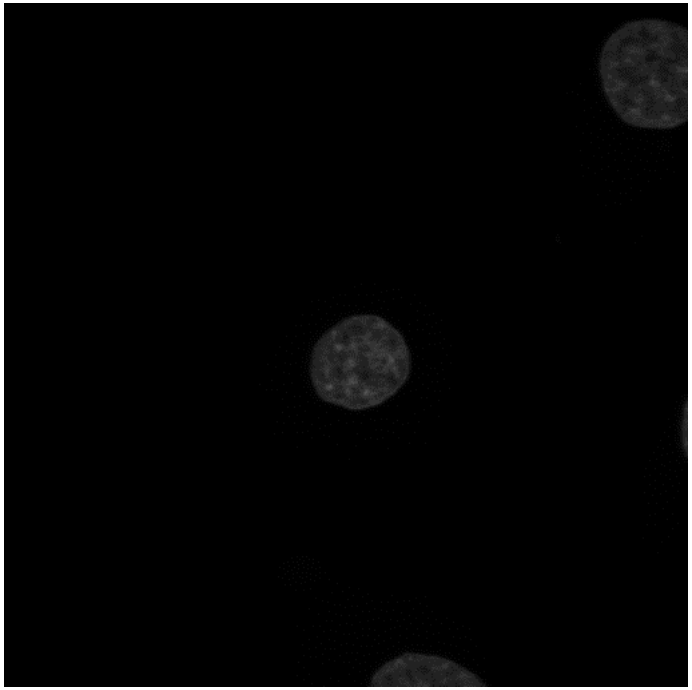


MTs

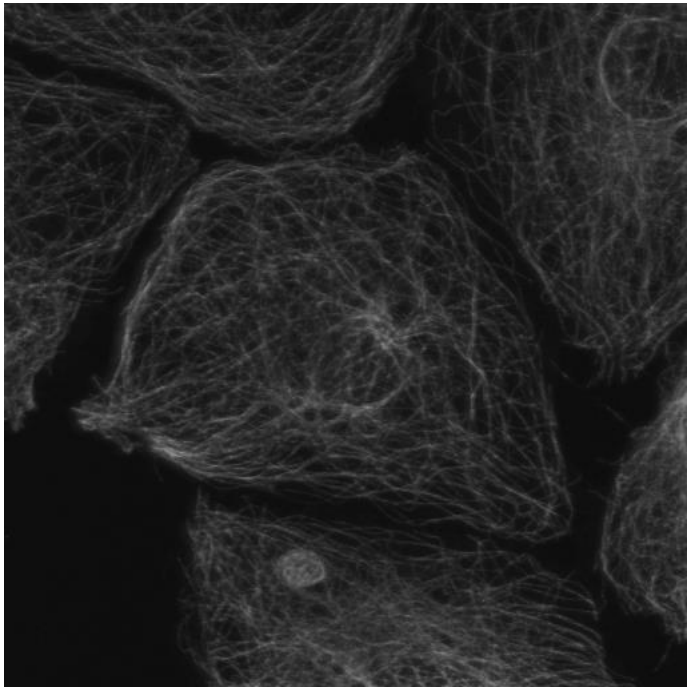




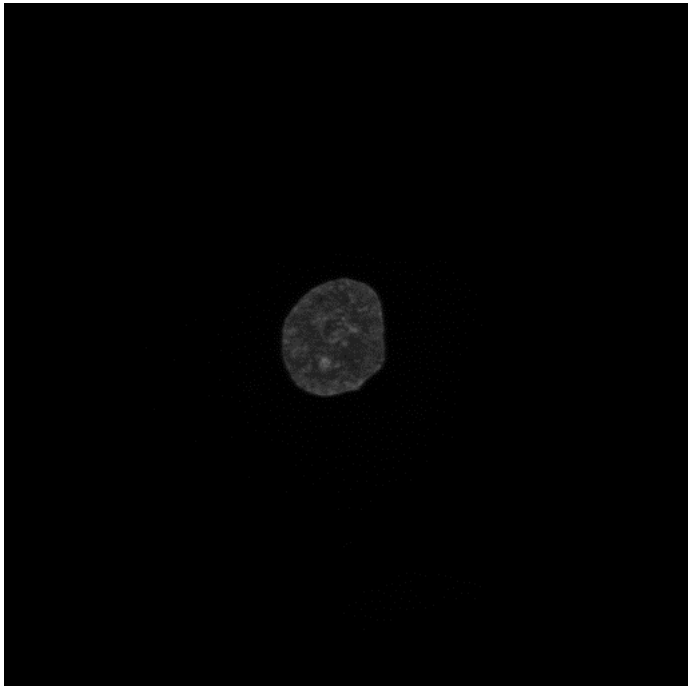
Resistant MKN7 Non-Treated Cells

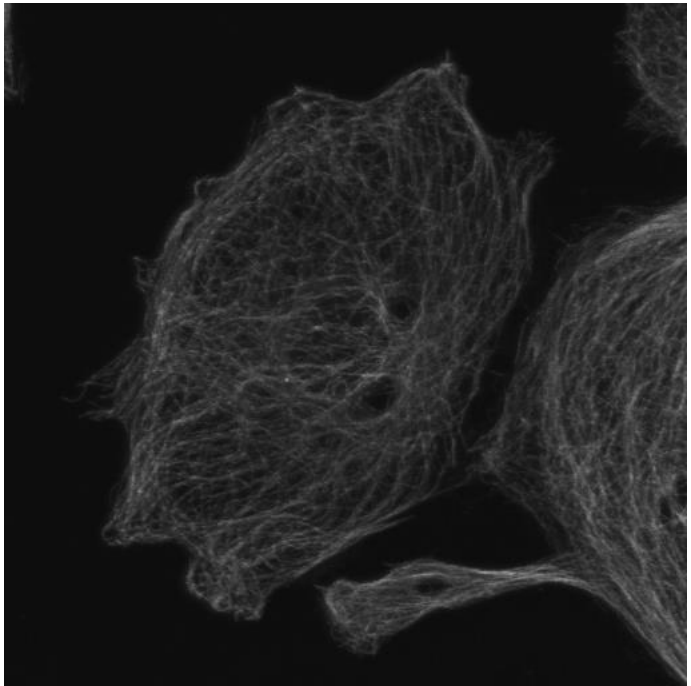


DAPI

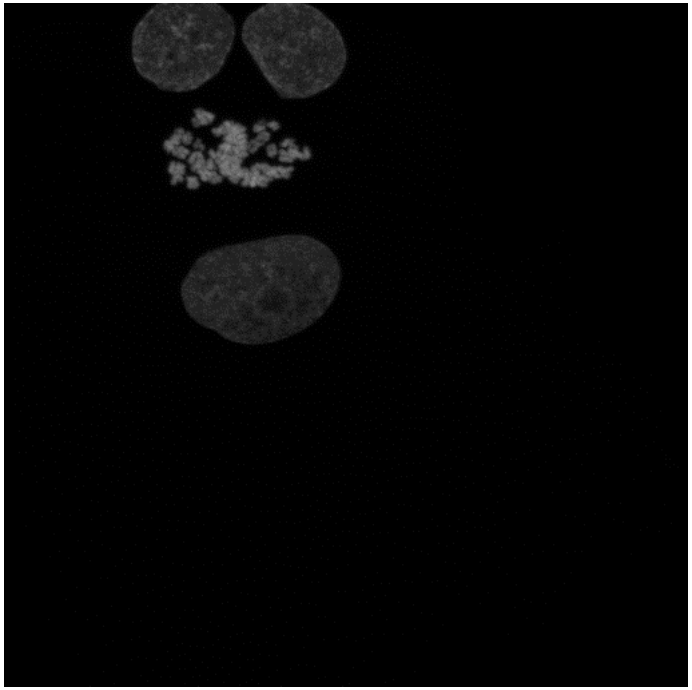


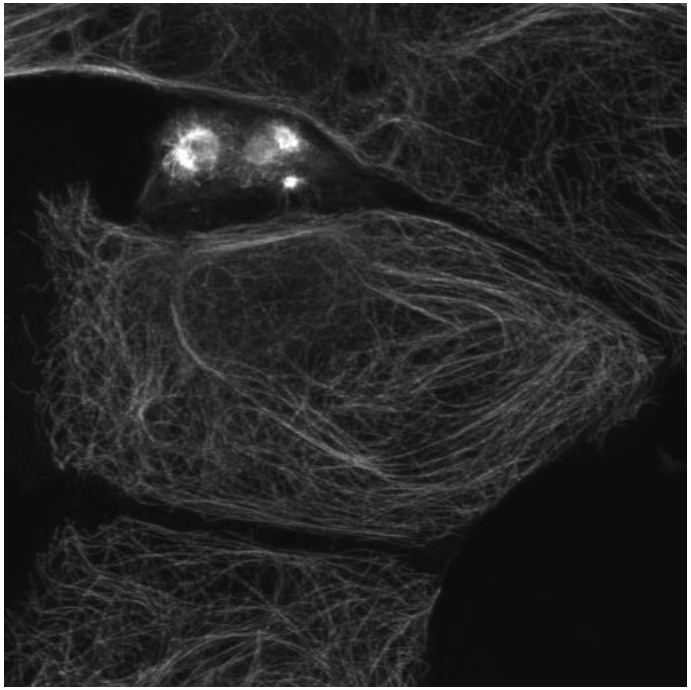
MTs





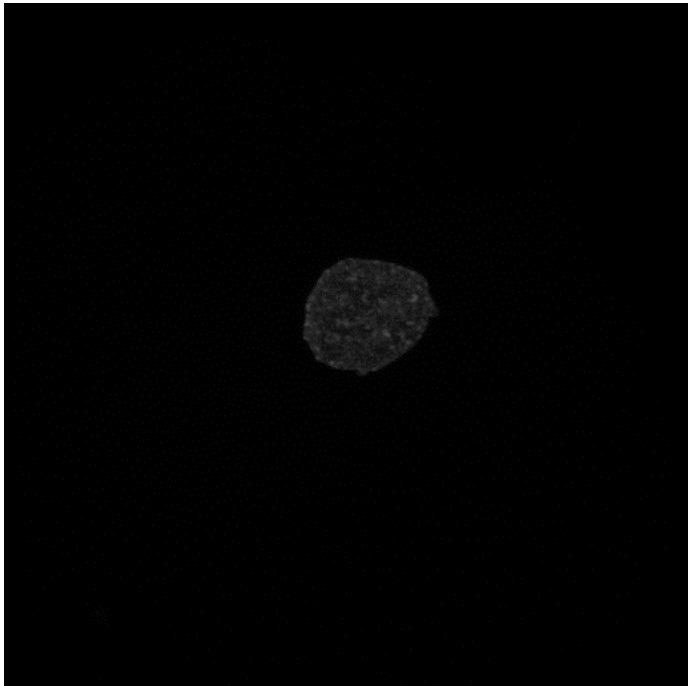
Resistant MKN7 100 nM DTX Treated Cells

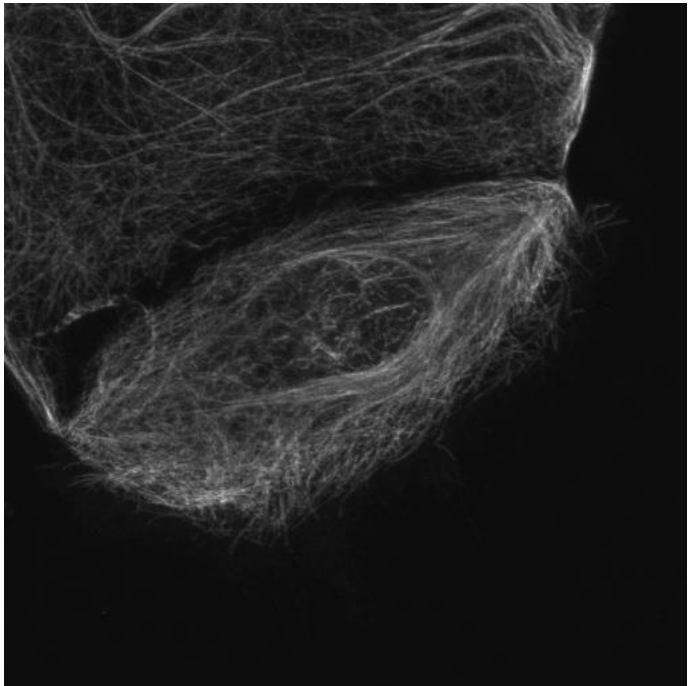


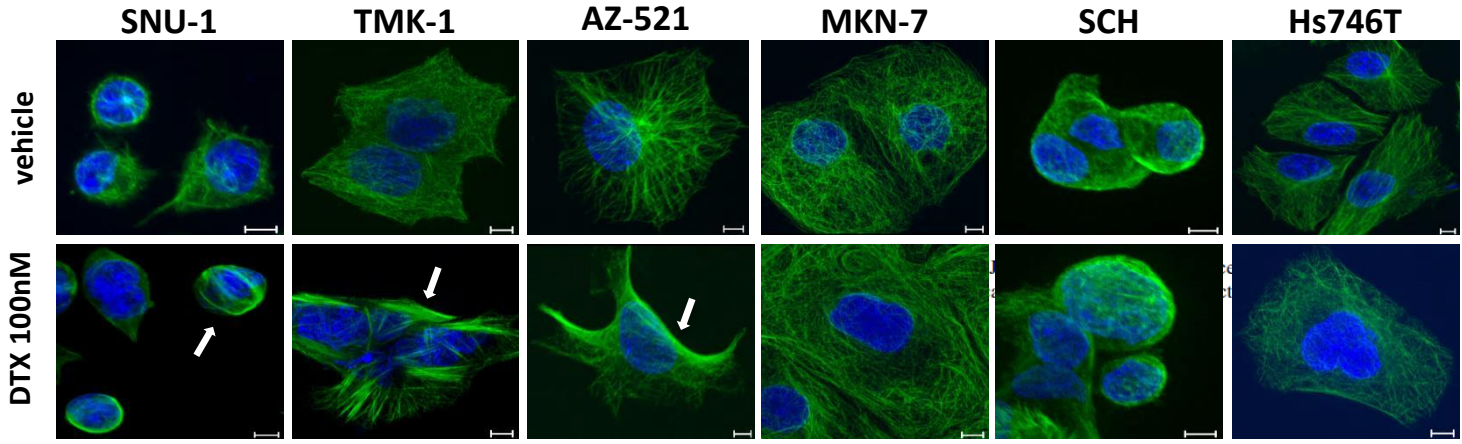


DAPI

MTs







ORMATICS

Vol. 17 no. 12 2001
Pages 1213–1223

A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells

*Michael V. Boland and Robert F. Murphy**

Center for Light Microscope Imaging and Biotechnology, Biomedical and Health Engineering Program, and Department of Biological Sciences, Carnegie Mellon University, 4400 Fifth Ave., Pittsburgh, PA 15213, USA

Robust Numerical Features for Description and Classification of Subcellular Location Patterns in Fluorescence Microscope Images*

ROBERT F. MURPHY, MEEL VELLISTE† AND GREGORY PORRECA**
Departments of Biological Sciences and Biomedical Engineering, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, PA 15213, USA

Received December 12, 2002; Revised February 17, 2003

Abstract. The ongoing biotechnology revolution promises a complete understanding of the mechanisms by which cells and tissues carry out their functions. Central to that goal is the determination of the function of each protein that is present in a given cell type, and determining a protein's location within cells is critical to understanding its function. As large amounts of data become available from genome-wide determination of protein subcellular location, automated approaches to categorizing and comparing location patterns are urgently needed. Since subcellular location is most often determined using fluorescence microscopy, we have developed automated systems for interpreting the resulting images. We report here improved numeric features for describing such images that are fairly robust to image intensity binning and spatial resolution. We validate these features by using them to train neural networks that accurately recognize all major subcellular patterns with an accuracy higher than any previously reported. Having validated the features by using them for classification, we also demonstrate using them to create Subcellular Location Trees that group similar proteins and provide a systematic framework for describing subcellular location.

Keywords: protein localization, subcellular location features, fluorescence microscopy, pattern recognition, location proteomics

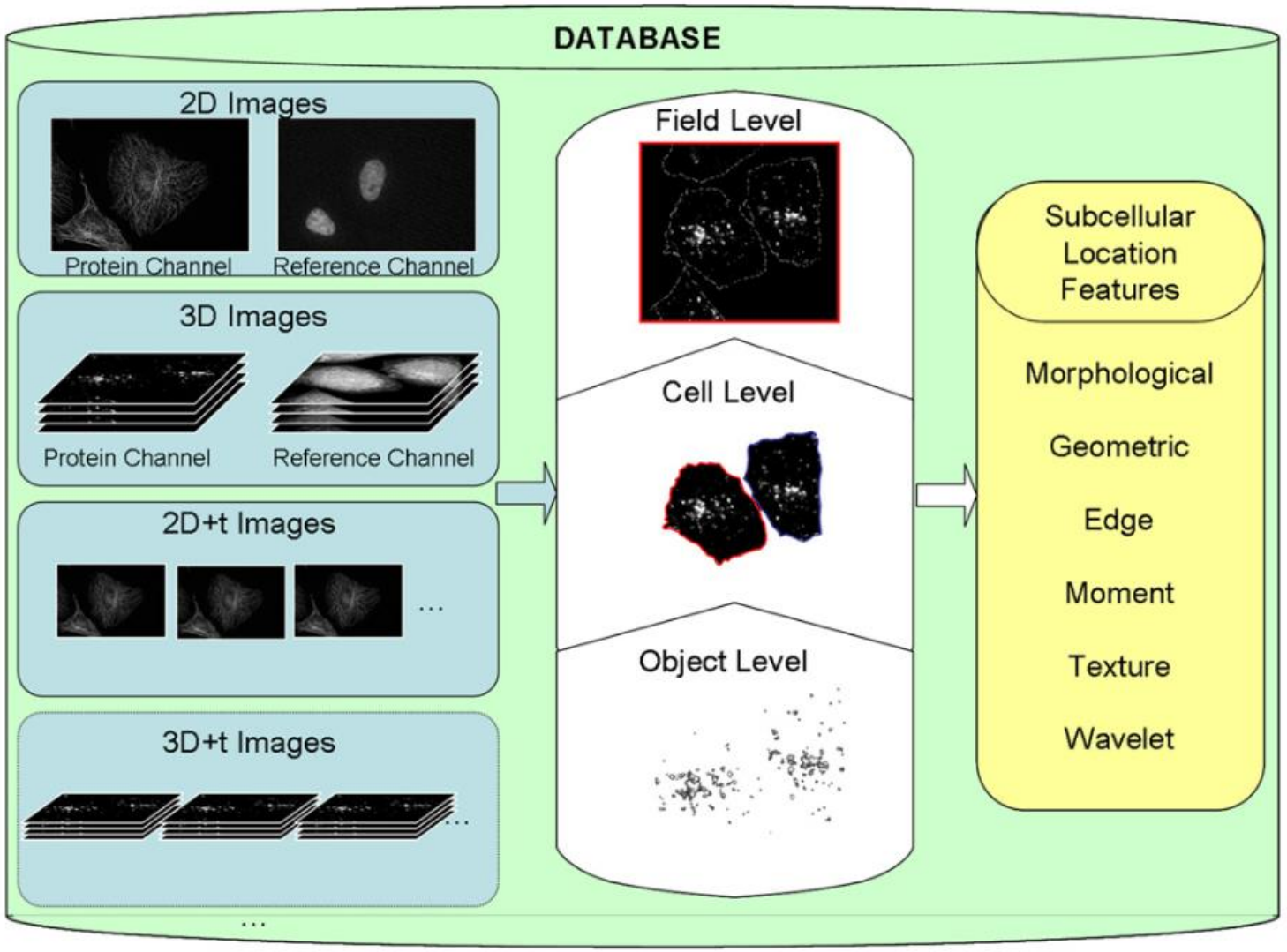


Figure 1. The Image Database Depicted Contains Images with Related Biological Protocol, Acquisition Parameters, and Subcellular Location Features

These numerical descriptors are computed at different semantic levels of the image content. The field-level features are calculated on the whole image, while cell- and object-level features require segmentation. The subcellular location features characterize the number, shape, gray-level distribution (texture, moments, and frequency), and relative size and position of the objects, in some cases relative to a reference channel. Some specific features are added for describing 3D and 2D+t stacks of images to improve the description by taking into account higher dimensions. As the number of dimensions increases by sampling in 3D or over time, more complex and informative features can be calculated.

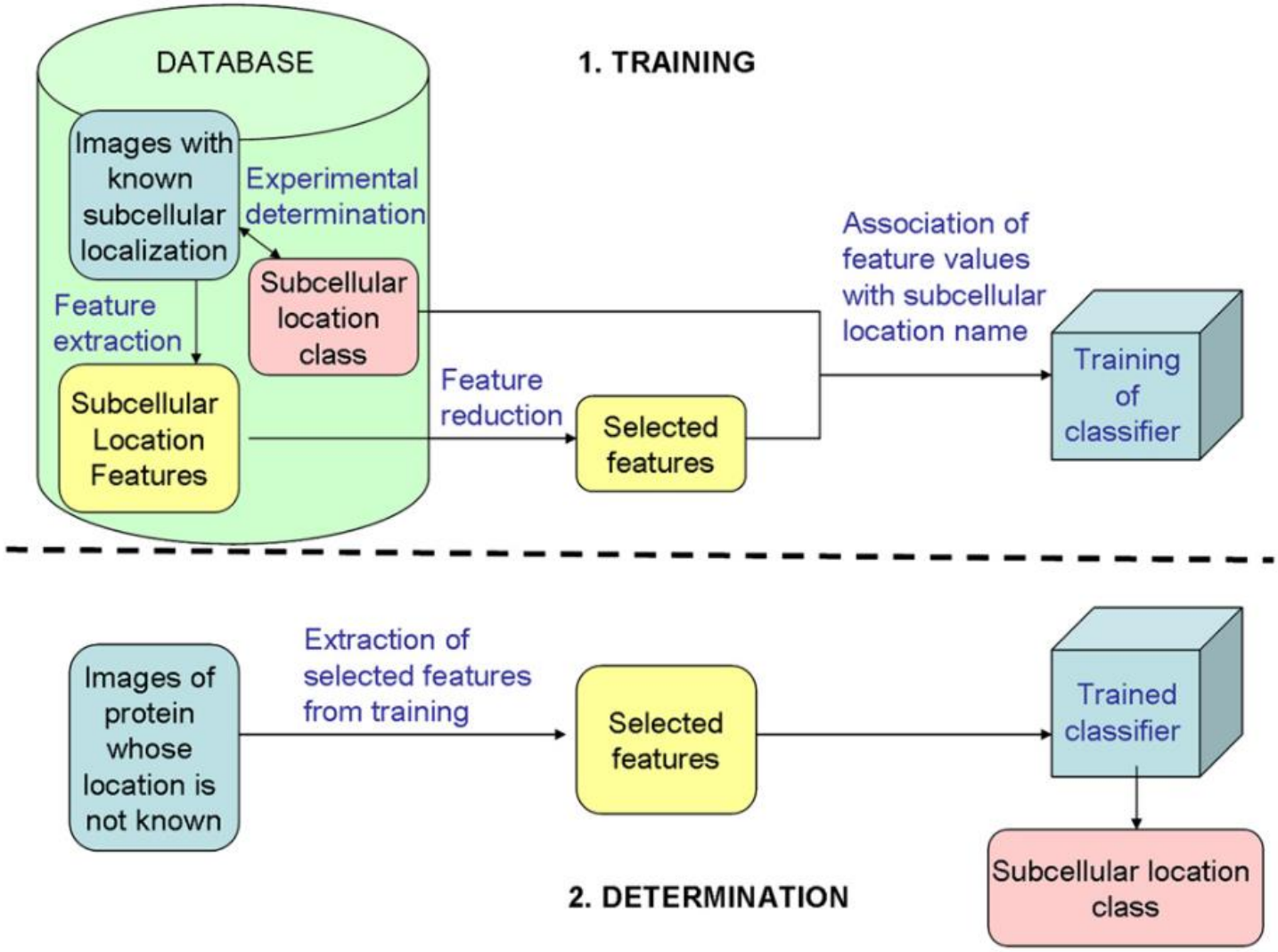


Figure 2. This Schematic Describes the Two Steps of the Supervised Classification for Determining the Subcellular Location Pattern of a Protein

The first step requires a set of images which represents the different classes of subcellular location patterns to be recognized. The feature extraction provides a numerical description for each image. The classifier is trained to distinguish the subcellular location patterns given the values of a selected set of the most discriminative features. The second step determines the subcellular location class of a target protein from its fluorescent microscope images. The selected features are computed and used as inputs of the trained classifier. The classifier assigns one of the known classes to the protein in each image. The accuracy is improved when 3D stacks, 2D time sequences, or several images are used.

Table 1. Examples of Questions about Subcellular Location Successfully Addressed by Computational Methods Incorporated in Publicly Available Fluorescence Microscope Image Databases

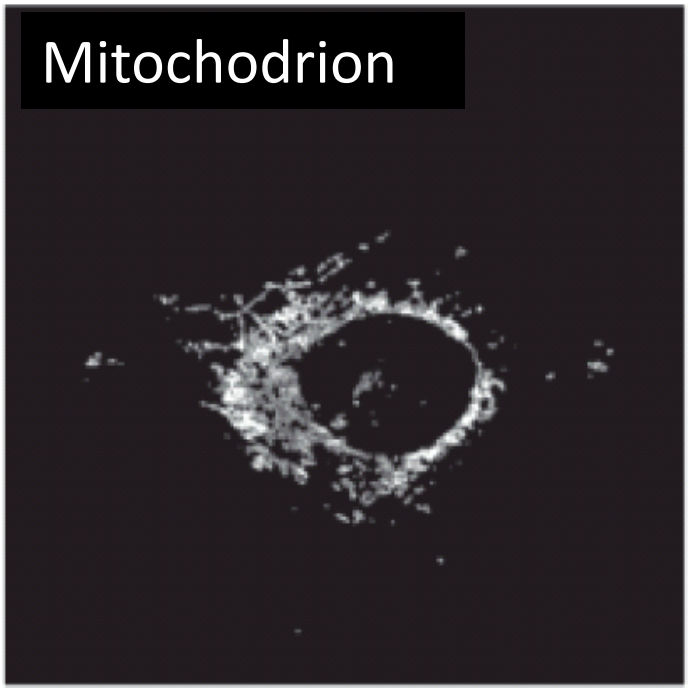
Questions	Method
Can I find images of a particular protein in this database?	Context-based image retrieval
How can images that look like a specific image be retrieved?	Content-based image retrieval
Do these two proteins have the same location pattern?	Statistical tests
Does the modification of the biological protocol change the location pattern of the target protein?	Statistical tests
What is the most representative image of this experiment/set of images?	Measuring distance in feature space from population mean
In what subcellular compartment is this protein?	Supervised classification
How can proteins that have the same location pattern be grouped together into families?	Clustering (unsupervised classification) and tree generation

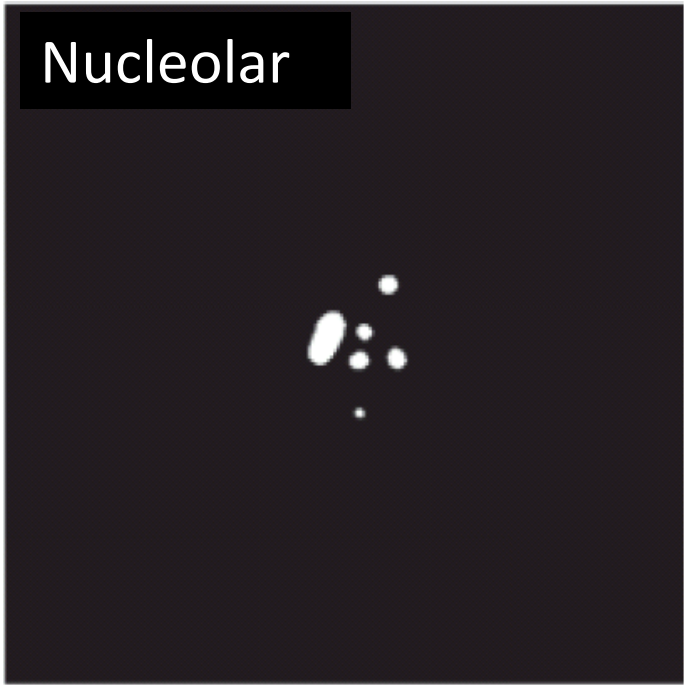
Automated Subcellular Location Determination and High-Throughput Microscopy

Estelle Glory & Robert F. Murphy

Developmental Cell – Review; 2007

Example of subcellular location features





(Boland and Murphy, 2001)

108	Number of objects	6
83	Average size of objects	232

morphological features, texture features and DNA features

Approach – compute 85 numerical features based on fluorescent intensity patterns in IF images

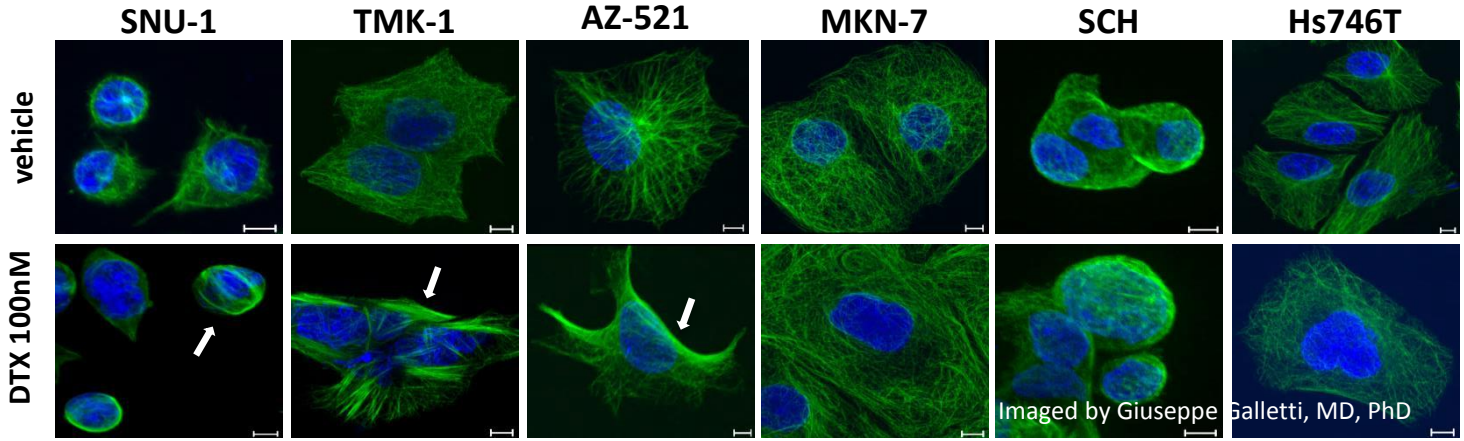
Comparison of subcellular location feature sets. All features that measure length or area are calculated in pixels that are 0.23 μm square in the sample plane.

Feature description	SLF3	SLF7
Morphological features: Number of fluorescent objects in image, Euler number of image, average object size, variance of object size, ratio of largest to smallest object size, average object distance to cell center of fluorescence, variance of object distance to cell center, ratio of largest to smallest object distance to cell center	SLF1.1 through SLF1.8	SLF1.1 through SLF1.8
Edge-related features: Fraction of above-threshold pixels along edge, measure of edge gradient intensity homogeneity, measure of edge direction homogeneity 1, measure of edge direction homogeneity 2, measure of edge direction difference	SLF1.9 through SLF1.13	SLF7.9 through SLF7.13 (minor error corrections)
Convex hull features: Fraction of convex hull occupied by above-threshold pixels, roundess of convex hull, eccentricity of convex hull	SLF1.14 through SLF1.16	SLF1.14 through SLF1.16
Zernike moment features through order 12, calculated for a unit circle with radius equal to the average radius of the cell type being analyzed (150 pixels or 34.5 μm for HeLa)	SLF3.17 through SLF3.65	SLF3.17 through SLF3.65
Haralick texture features: angular second moment, contrast, correlation, sum of squares variation, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, info. measure of correlation 1, info. measure of correlation 2	SLF3.66 through SLF3.78	SLF7.66 through SLF7.78 (after downsampling to 1.15 μm/pixel and 256 gray levels)
Fraction of non-object fluorescence	–	SLF7.79
Skeleton features (see text)	–	SLF7.80 through SLF7.84

Table 1 Feature sets defined for 2-D fluorescence microscope images.

Set	SLF number	Feature description
SLF1	SLF1.1	The number of fluorescence objects in the image
	SLF1.2	The Euler number of the image (no. of holes minus no. of objects)
	SLF1.3	The average number of above-threshold pixels per object
	SLF1.4	The variance of the number of above-threshold pixels per object
	SLF1.5	The ratio of the size of the largest object to the smallest
	SLF1.6	The average object distance to the cellular center of fluorescence (COF)
	SLF1.7	The variance of object distances from the COF
	SLF1.8	The ratio of the largest to the smallest object to COF distance
	SLF1.9	The fraction of the nonzero pixels that are along an edge
	SLF1.10	Measure of edge gradient intensity homogeneity
	SLF1.11	Measure of edge direction homogeneity 1
	SLF1.12	Measure of edge direction homogeneity 2
	SLF1.13	Measure of edge direction difference
	SLF1.14	The fraction of the convex hull area occupied by protein fluorescence
	SLF1.15	The roundness of the convex hll
	SLF1.16	The eccentricity of the convex hull
SLF2	SLF2.1 to 2.16	SLF1.1 to SLF1.16
	SLF2.17	The average object distance from the COF of the DNA image
	SLF2.18	The variance of object distances from the DNA COF
	SLF2.19	The ratio of the largest to the smallest object to DNA COF distance
	SLF2.20	The distance between the protein COF and the DNA COF
	SLF2.21	The ratio of the area occupied by protein to that occupied by DNA
	SLF2.22	The fraction of the protein fluorescence that co-localizes with DNA
SLF3	SLF3.1 to 3.16	SLF1.1 to SLF1.16
	SLF3.17 to 3.65	Zernike moment features
	SLF3.66 to 3.78	Haralick texture features
SLF4	SLF4.1 to 4.22	SLF2.1 to 2.22
	SLF4.23 to 4.84	SLF3.17 to 3.78
SLF5	SLF5.1 to SLF5.37	37 features selected from SLF4 using stepwise discriminant analysis
SLF6	SLF6.1 to 6.65	SLF3.1 to SLF3.65
SLF7	SLF7.1 to 7.9	SLF3.1 to 3.9
	SLF7.10 to 7.13	Minor corrections to SLF3.10 to SLF3.13
	SLF7.14 to 7.65	SLF3.14 to SLF3.65
	SLF7.66 to 7.78	Haralick texture features calculated on fixed size and intensity scales
	SLF7.79	The fraction of cellular fluorescence not included in objects
	SLF7.80	The average length of the morphological skeleton of objects
	SLF7.81	The average ratio of object skeleton length to the area of the convex hull of the skeleton
	SLF7.82	The average fraction of object pixels contained within its skeleton
	SLF7.83	The average fraction of object fluorescence contained within its skeleton
	SLF7.84	The average ratio of the number of branch points in skeleton to length of skeleton
SLF8	SLF8.1 to 8.32	32 features selected from SLF7 using stepwise discriminant analysis
SLF12	SLF12.1 to 12.8	SLF8.1 to 8.8, the smallest feature set able to achieve 80% accuracy
SLF13	SLF13.1 to 13.31	31 features selected from SLF7 and SLF2.17-2.22 using stepwise discriminant analysis

Chart with statistically significant changes after treatment



P<0.01
Red 1 - increase after treatment
Green 1 –decrease after treatment

Featname\# feat	4	10	5	1	3	2
2 obj:EulerNmbr		1				
3 obj_size:aver		1				
4 obj_size:var		1				
6 obj_dist:aver				1		
7 obj_dist:var		1				
9 Edge area fract	1		1			1
10 Edge homog		1				
11 E dir maxmin	1	1			1	1
12 E dir maxnex	1					
13 E dir differen	1					
14 Skel length		1	1			
15 S hul area rat			1			
19 Hull overlap		1	1			
20 Hull shape		1	1		1	
21 Hull eccentric		1			1	

Morphology changes of the microtubule network after docetaxel treatment measured by subcellular localization features in high resolution images

References

- *Murphy, Robert F, "Robust Numerical Features for Description and Classification of Subcellular Location Patterns in Fluorescence Microscope Images", VLSI Signal Processing, 2003*
- *Boland, MV et al, "A Neural Network Classifier Capable of Recognizing the Patterns of All Major Subcellular Structures in Fluorescence Microscope Images of HeLa Cells", Bioinformatics, 2001*

21 MT CHANNEL FEATURES

SIZE AND DISTRIBUTION OF MTs

1) object: number – number of fluorescent objects in image

2) object: EurlerNumber - number of objects in the image minus the total number of holes in those objects – **distinguishes reticular or mesh-like patterns vs more uniformly distributed patterns**

3) object_size:average - The average number of above-threshold pixels per object – captures information about the **size of objects in cell MT area**

4) object_size:variance - The variance of the number of above-threshold pixels per object – quantifies the homogeneity of fluorescent objects in cells

5) object_size:ratio - The ratio of the size of the largest object to the smallest within the cell – assessing the distribution of florescent object sizes

6) object_distance: average – average object distance to cell center of fluorescence (COF) – provides information about how individual objects are distributed throughout the cell

7) object_distance: variance – variance of object distance to cell center of fl (COF) – captures information about the distribution of objects around a central point

8) Object_distance: ratio – ratio of largest to smallest distance to cell center of fl (COF – center of fluorescence)

OBJECT EDGE FEATURES

9) edges:area_fraction - fraction of the non-zero pixels in a cell that are along an edge – **distinguishes protein that localizes along the edges**

10) edges:homogeneity - Measure of edge intensity homogeneity - captures homogeneity of edge gradients, or '**are the edges primarily steep or more gradually sloping**?'

11) edges:direction_maxmin_ratio - Measure of edge direction homogeneity 1 – captures homogeneity of edge direction, or are the edges primarily in one direction or are they more evenly distributed? images with **patterns containing edges oriented predominantly along a particular direction** result in edge gradient histograms

12) edges:direction_maxnextmax_ratio - Measure of edge direction homogeneity 2 – ratio of the largest to the next largest value in the histogram from above feature

13) edges:direction_difference - Measure of edge direction difference - this feature **distinguish MT patterns in which there are parallel edges**

SHAPE OF THE MT NEWTORK

14) obj_skel_len - The average length of the morphological skeleton of objects

15) obj_skel_hull_area_ratio - The ratio of object skeleton length to the area of the convex hull of the skeleton, averaged over all objects

16) obj_skel obj area ratio - The fraction of object pixels contained within the skeleton

17) obj_skel obj fluor ratio - The fraction of object fluorescence contained within the skeleton

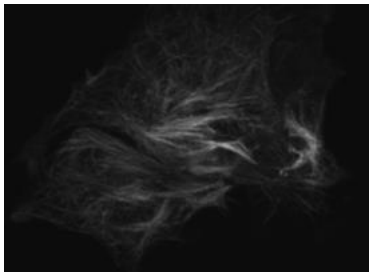
18) obj_skel_branch_per_len - The ratio of the number of branch points in skeleton to length of skeleton

19) convex_hull: fraction of overlap - fraction of convex hull occupied by protein fluorescence (above-threshold pixels)

20) convex_hull: shape_factor- roundness of convex hull

21) convex_hull: eccentricity - eccentricity (elongation) of convex hull

Sensitive TMK1, SNU1 and AZ521 cells change significantly the Edge Direction Homogeneity (feature 6) of the MT area (bundles) after DTX treatment, while resistant cells do not



SLF1.11—measure of edge direction homogeneity 1. The edge direction gradient at each point in the image **G** was then calculated from the convolved images, $\mathbf{G_N}$ and $\mathbf{G_W}$, used in SLF1.10:

$$G(x, y) = \tan^{-1}\left(\frac{G_N(x, y)}{G_W(x, y)}\right).$$

The value of each pixel in the image **G** is therefore the direction (from $-\pi$ to π) of the intensity gradient at that point in the image, **I**. An eight-bin histogram was calculated using all of the values in the gradient image **G**. The final feature was calculated as the ratio of the largest to smallest value in the histogram. This feature was designed to capture the homogeneity of edge direction, i.e. are the edges primarily in one direction or are they more evenly distributed? Images with patterns containing edges oriented predominantly along a particular direction (some patterns of actin filaments, for example) result in edge gradient histograms in which a few bins will dominate. Histograms of edge direction are not completely

Sensitive TMK1 and SNU1 cells change significantly the Edge Intensity Homogeneity (feature 5) of the MT area (bundles) after DTX treatment, while resistant cells do not

TMK1, AZ 521 sensitive cells change significantly the Ratio of the Size of the Largest to the Smallest MT area (feature 3) after DTX treatment

SLF1.5—the ratio of the size of the largest object to the smallest. This was defined as the number of pixels in the largest object divided by the number of pixels in the smallest object. Like SLF1.4, this feature was included as a means of assessing the distribution of fluorescent object sizes.

MKN7 resistant cells increase significantly the Ratio of the Number of Branch Points in Skeleton to Length of Skeleton (feature 13) after DTX treatment

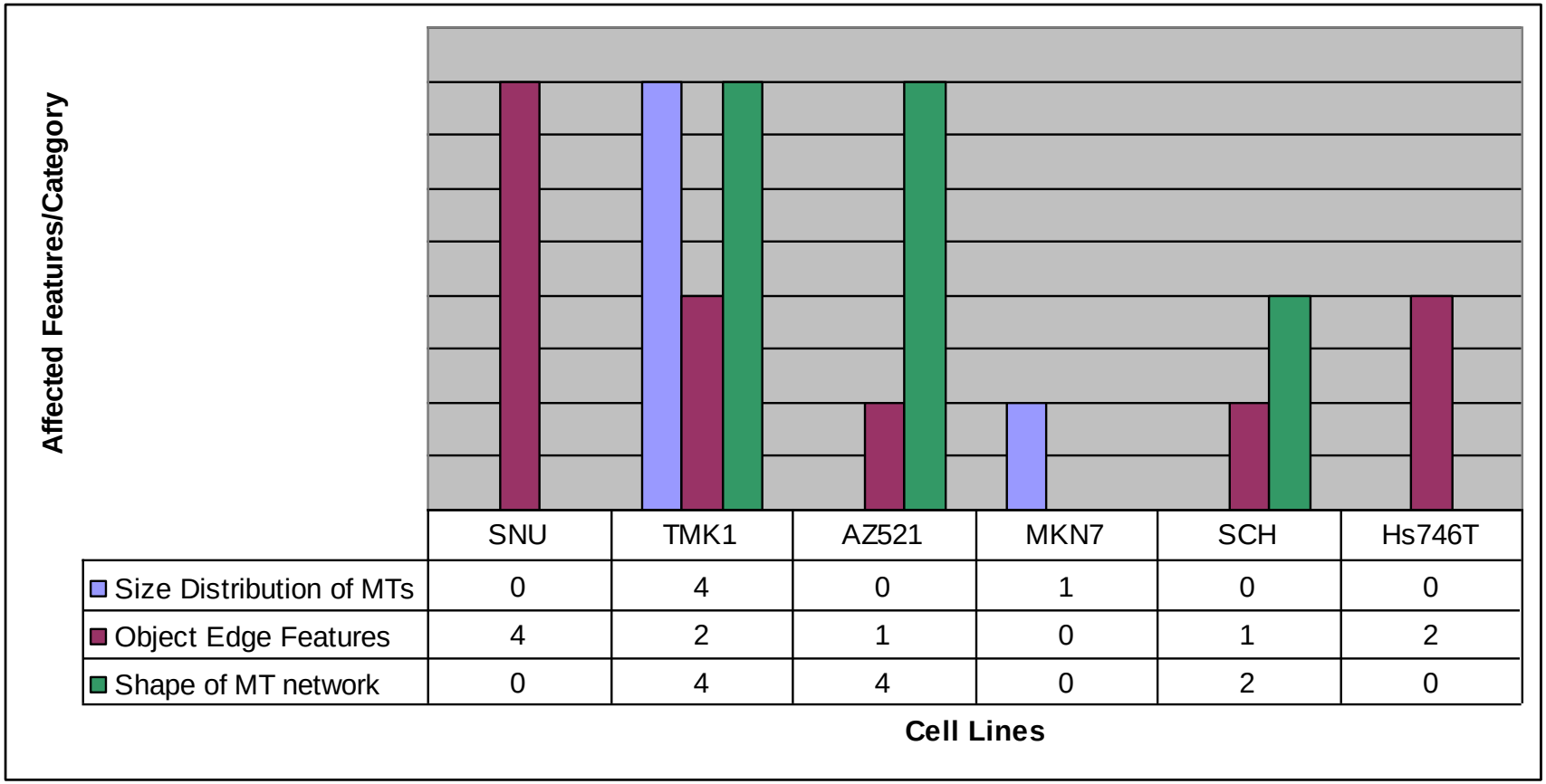
TMK1 and AZ521 sensitive cells change significantly the Average Length of the Morphological Skeleton of Objects (feature 9) after DTX treatment

- We have 3 main categories of features (in 3 different colors)

- Each category is represented by 5 features

- For the 3 resistant lines (on the right side), no more than 2 features per category are affected by DTX, which is 40% (2 of 5 or less), i.e. less than half

- For the 3 sensitive lines (on the left side), for at least 1 category 4 of 5 features are affected by DTX (80% per category)

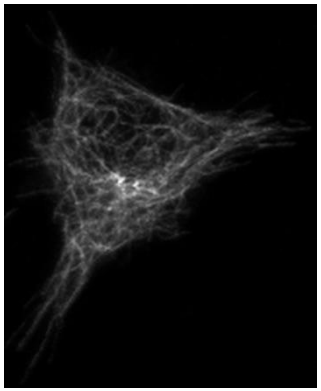


	SNU	TMK1	AZ521	MKN7	SCH	Hs746T
Size Distribution of MTs	0	4	0	1	0	0
Object Edge Features	4	2	1	0	1	2
Shape of MT network	0	4	4	0	2	0

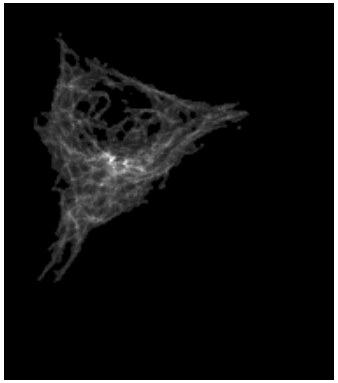
SNU1 ctl (#19, 10segment)

The ration A gives the total area of the image that is edges

Raw data

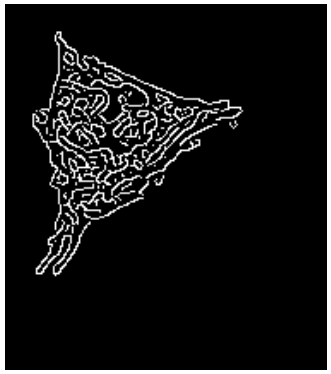


Bright pixels

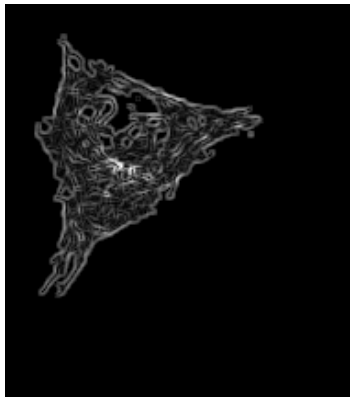


1/0 mask foreground

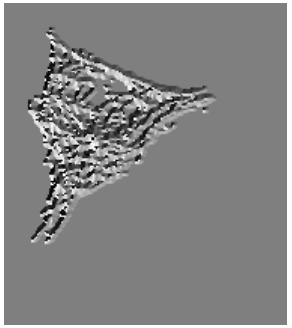
Edge map



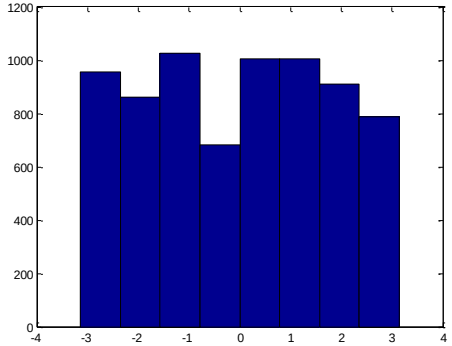
Gradient/Edge magnitude



Gradient dir (radians)

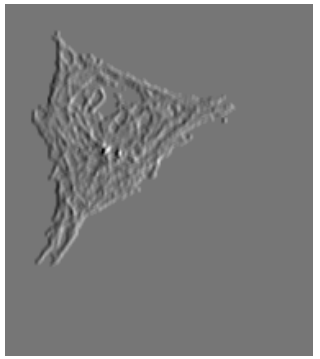


Hist Grad dir (radians) hist(v,8)

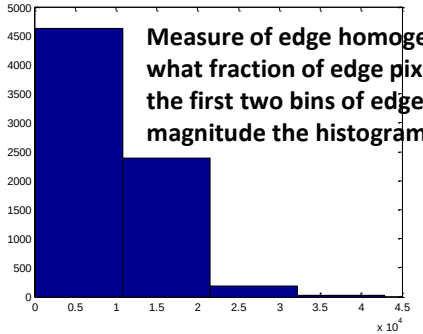


Y dir gradient

X dir gradient



Hist Edge magnitude hist(v_mag,4)



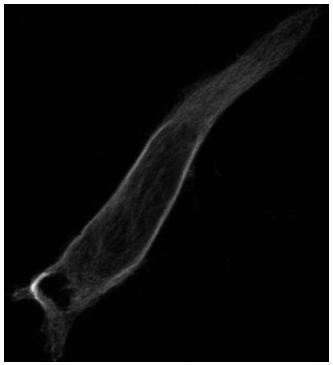
Measure of edge homogeneity - what fraction of edge pixels are in the first two bins of edge magnitude the histogram

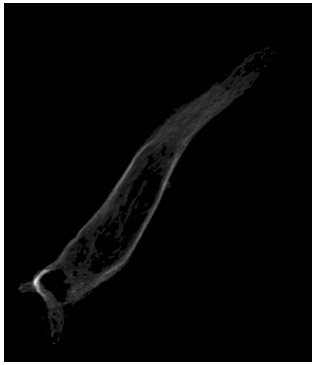
values = 0.3678 0.6397 1.5059 1.0199 0.2352
values = [A homogeneity maxminratio maxnextmaxratio sumdiff] ;
Feat # 9, 10, 11, 12, 13

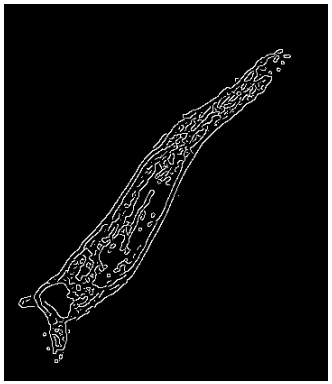
SNU1 dtx (#28, 20segment)

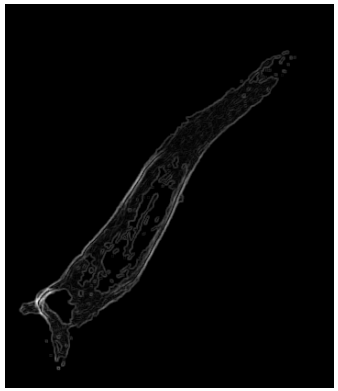
Raw data

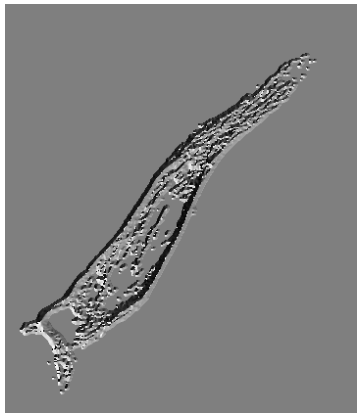
Bright pixels

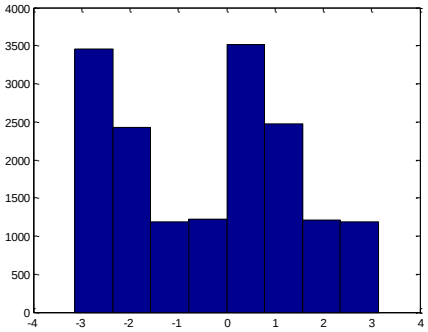




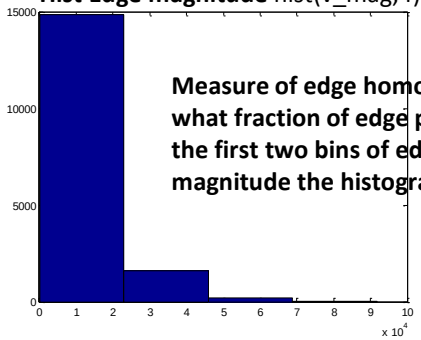








Hist Edge magnitude hist(v_mag,4)

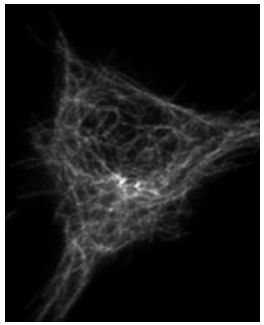


Measure of edge homogeneity - what fraction of edge pixels are in the first two bins of edge magnitude the histogram

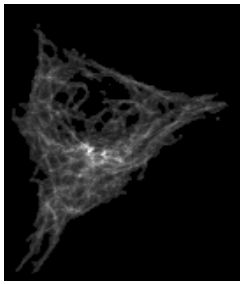
values =0.3599 0.8914 2.9646 1.0174 0.0462

values = [A homogeneity maxminratio maxnextmaxratio sumdiff] ;

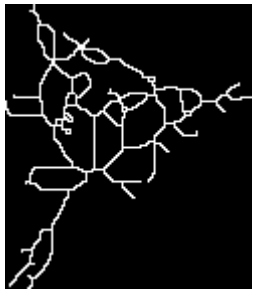
SNU1 ctl (#19, 10segment)



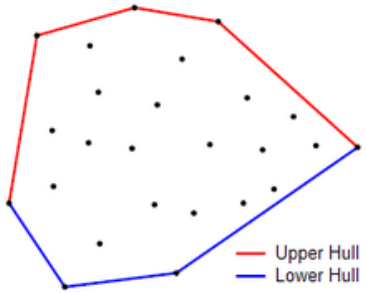
Obj Image



Obj Skeleton



Skel Hull



Morphological Skeleton is a skeleton (or medial axis) representation of a shape. The medial axis of an object is the set of all points having more than one closest point on the object's boundary. The skeleton usually emphasizes geometrical and topological properties of the shape; smallest possible set of lines that preserve the topology and are equidistant to the borders.
Originally referred to as the topological skeleton, it was introduced as a tool for biological shape recognition.

skellen – is the length of the object skeleton (#14)

skel_obj_area_ratio - is skellen / objsize (number of pixels of Obj Image) (#16)

skel_hull_area_ratio – is skellen / hullsize (number of pixels in Skell Hull) (#15)

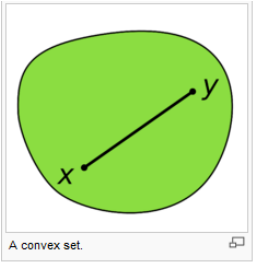
skel_fluor = sum(objimg(find(objskel))) sum of intensity of the skeleton pixels

obj_fluor = sum(objimg(:)) – sum of all intensity of the object

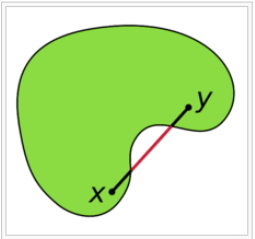
skel_obj_fluor_ratio = skel_fluor/obj_fluor; (#17)

Ratio # branch_points / length skeleton (#18)

Convex hull or convex envelope of a set X of points is the smallest convex set that contains X;
An object is convex if for every pair of points within the object, every point on the straight line segment that joins the pair of points is also within the object
Computing Convex hulls finds its practical applications in pattern recognition, image processing, statistics, etc.



A convex set.

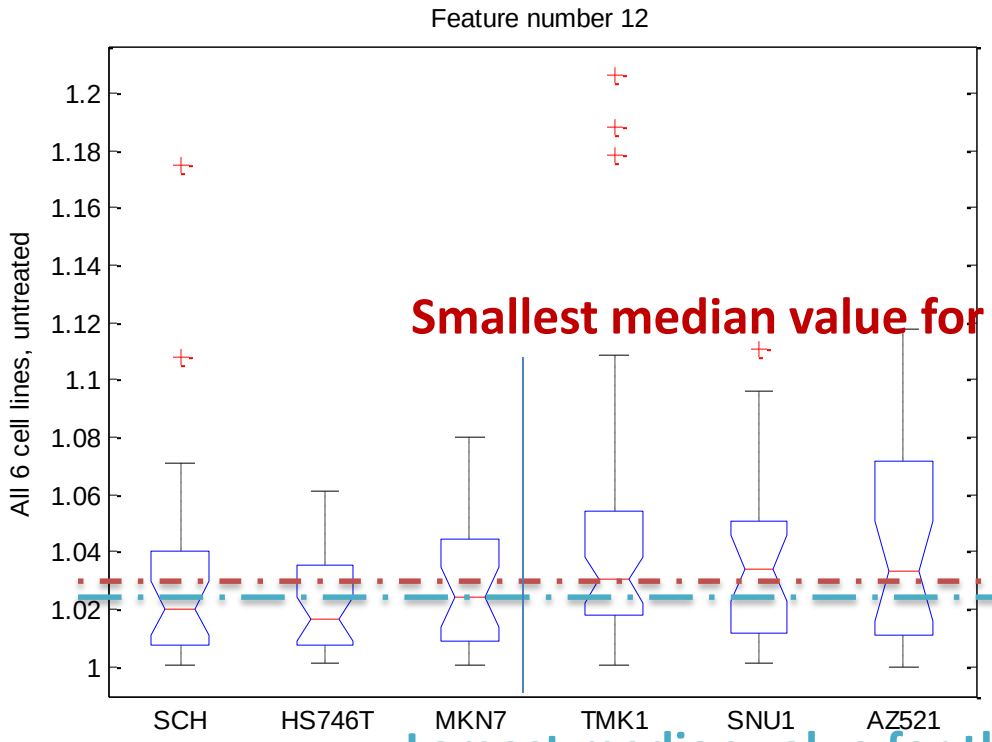


A non-convex set, with a line-segment outside the set.

Feats #14, 15, 16, 17, 18 = [skellen / skel_hull_area_ratio / skel_obj_area_ratio / skel_obj_fluor_ratio / no_of_branch_points/skellen];

feats = 1089 0.1168 0.1809 0.1717 0.0735

Baseline differences, Feature 12



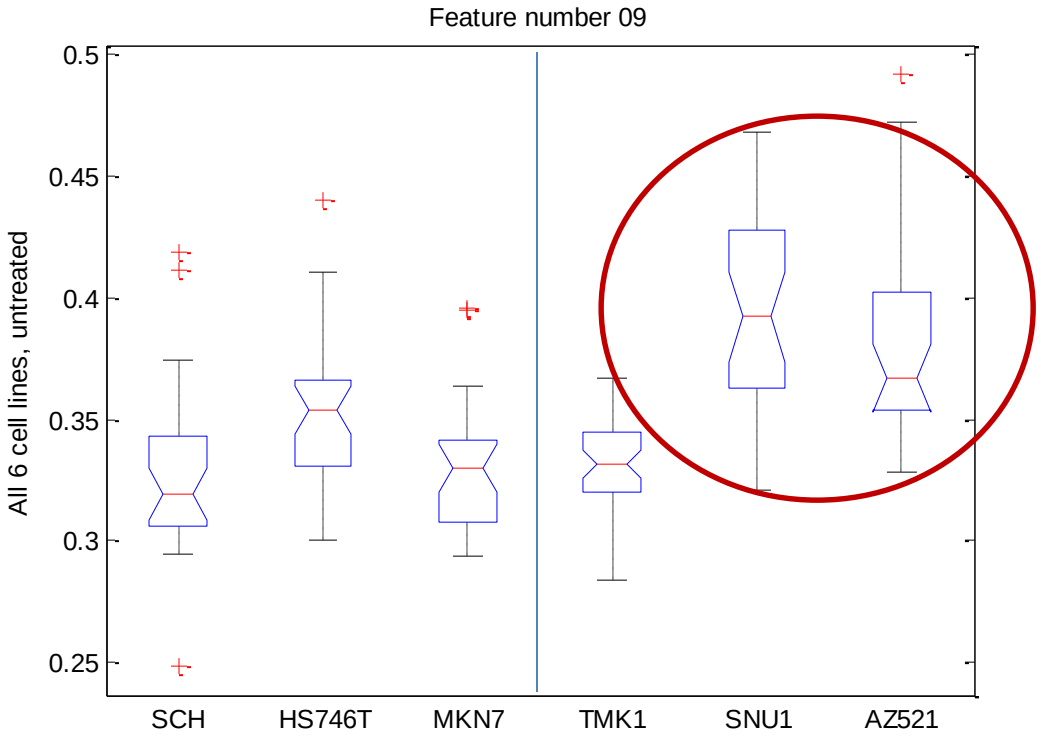
Smallest median value for the sensitive cell lines

Largest median value for the resistant cell lines

patterns containing edges oriented predominantly along a particular direction

(ratio of the largest to the next largest value in the edge direction histogram)

Baseline differences, Feature 09

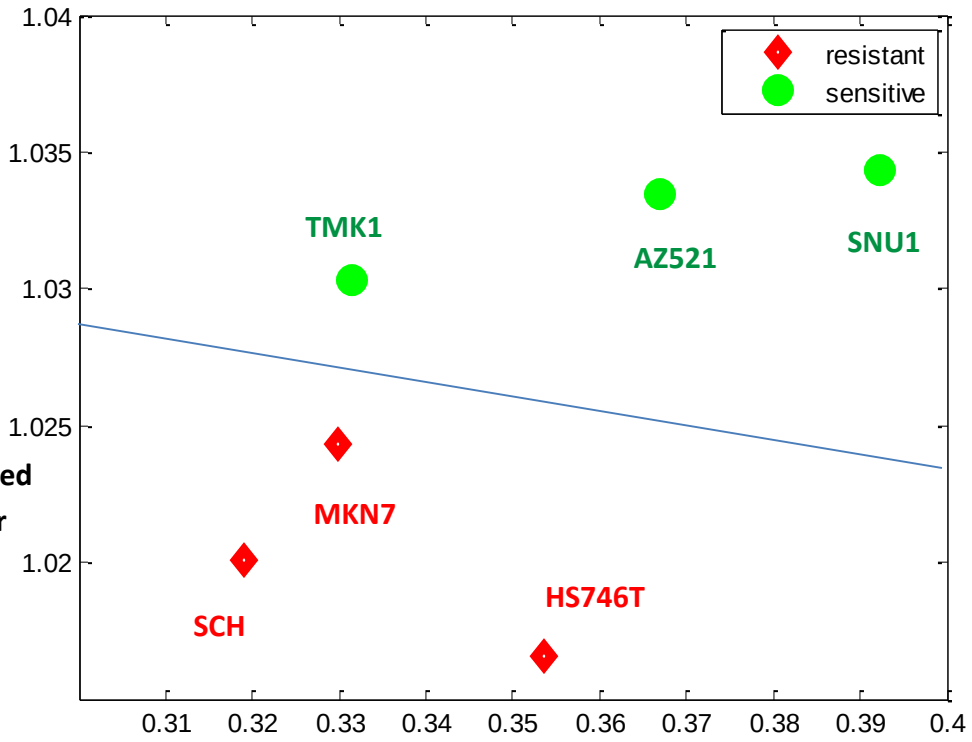


distinguishes protein that localizes along the edges

Scatter plot, Features 9 + 12 at baseline

Feature 12

patterns containing edges oriented predominantly along a particular direction



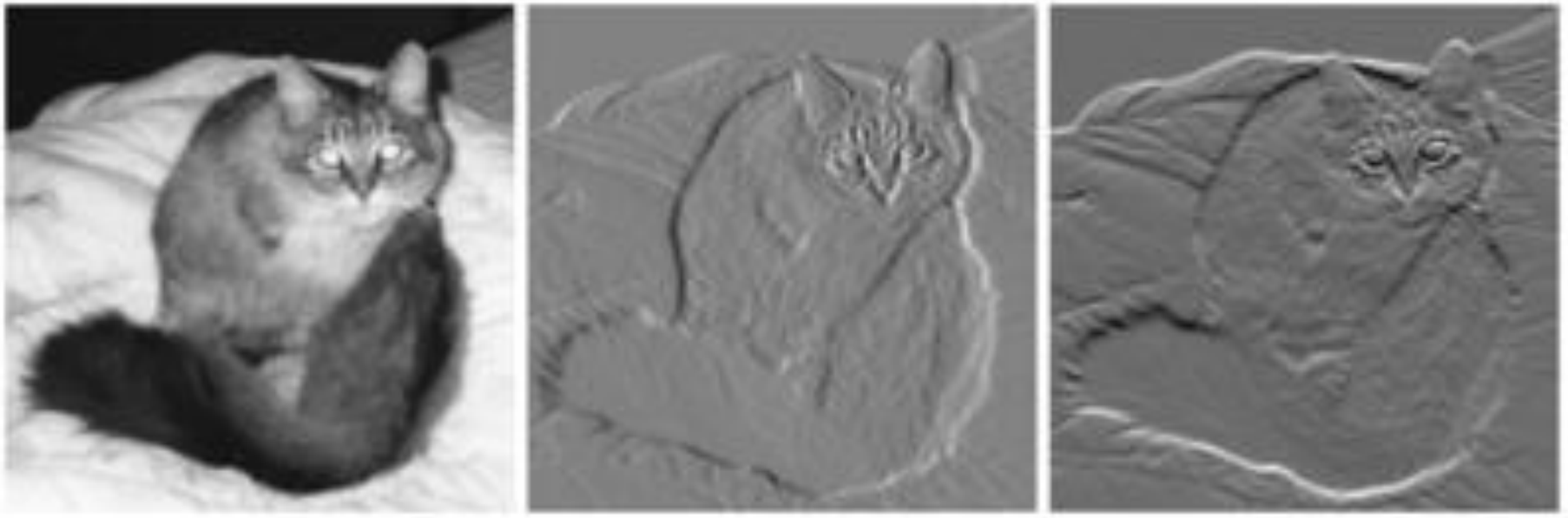
Feature 9

distinguishes protein that localizes along the edges

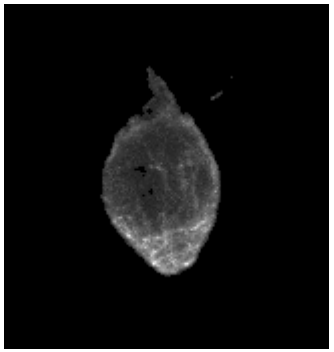
Next steps

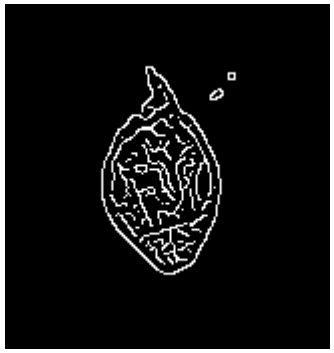
- **Extend the analysis to texture analysis (Harralick and Zernike)**
- **Extract another set of different features for the DAPI channel**
- **Develop new custom features**

Back up slides - Example of image gradient

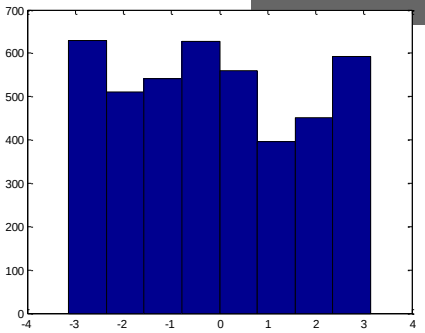


On the left, an intensity image of a cat. In the center, a gradient image in the x direction measuring horizontal change in intensity. On the right, a gradient image in the y direction measuring vertical change in intensity. Gray pixels have a small gradient; black or white pixels have a large gradient.

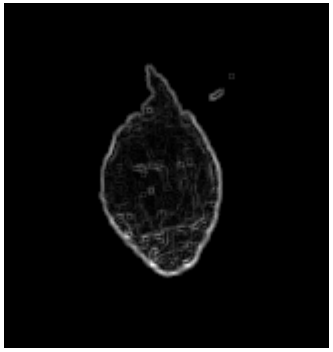




edge

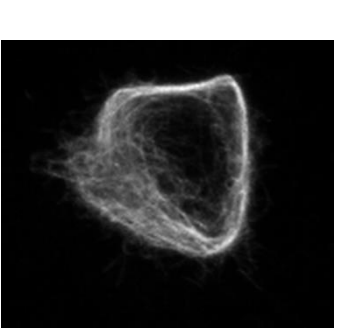
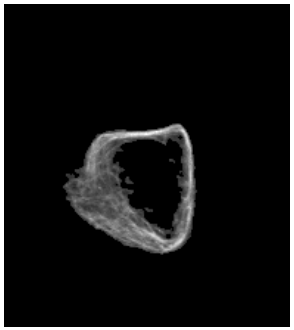
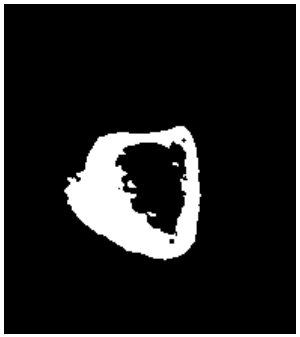
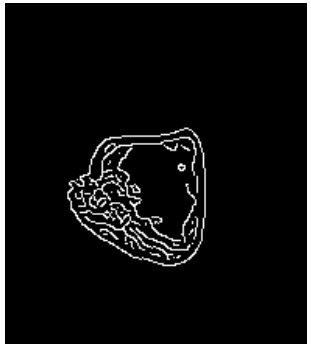
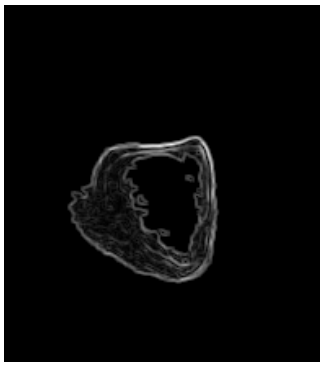


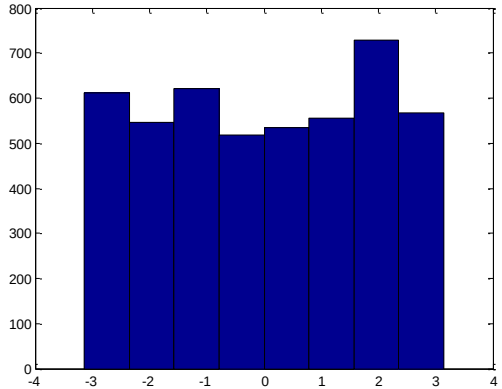

```
% Total area of the image that is edges
A = bwarea(edge(imageproc,'canny',[]))/bwarea(im2bw(imageproc)) ;

% Directional edge filters
N = [1 1 1 ; 0 0 0 ; -1 -1 -1] ; y direction
W = [1 0 -1 ; 1 0 -1 ; 1 0 -1] ; x direction
•       iprocN = filter2(N,imageproc) ;
•       iprocW = filter2(W,imageproc) ;

% Calculate the magnitude and direction of the gradient
iprocmag = sqrt(iprocN.^2 + iprocW.^2) ;
iproctheta = atan2(iprocN, iprocW) ;
iproctheta(find(iproctheta==-pi))=pi;

% max/min ratio
[hmax maxidx] = max(h) ;
hmin = min(h) ;

if (hmin ~= 0),
  maxminratio = hmax/hmin ;
else
  maxminratio = 0 ;
end

htmp=h ;
htmp(maxidx) = 0 ;
hnextmax = max(htmp) ;
maxnextmaxratio=hmax/hnextmax ;
```

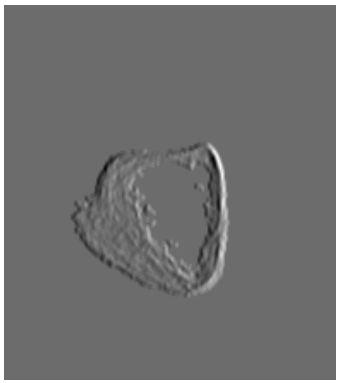


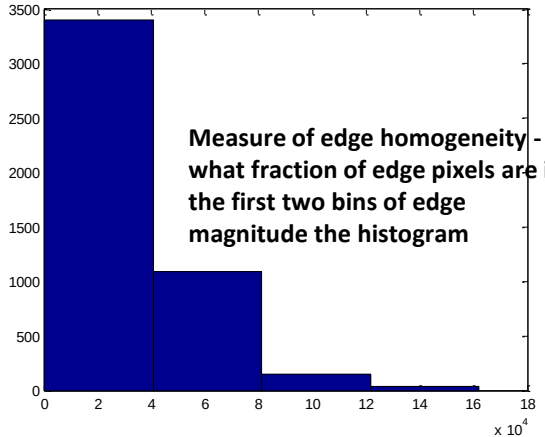
```
% Difference between bins of histogram at angle and angle+pi
%  In general, objects have an equal number of pixels at an angle
%  and that angle+pi. The differences are normalized to the sum of
%  the two directions.
diff = abs(h(1:4)-h(5:8))./abs(h(1:4)+h(5:8));
diff(abs(h(1:4)-h(5:8))==0) = 0;

sumdiff=sum(diff) ;

% Measure of edge homogeneity - what fraction of edge pixels are in
%  the first two bins of the histogram.
h_mag = hist(v_mag,4);
homogeneity = sum(h_mag(1))/sum(h_mag(:)) ;

names = [names cellstr('edges:area_fraction') ...
        cellstr('edges:homogeneity') ...
     cellstr('edges:direction_maxmin_ratio') ...
        cellstr('edges:direction_maxnextmax_ratio') ...
        cellstr('edges:direction_difference')] ;

values = [A homogeneity maxminratio maxnextmaxratio sumdiff] ;
```

SNU1 dtx (#, 11segment)



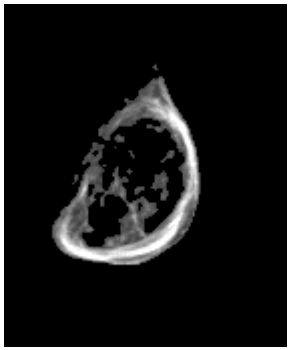



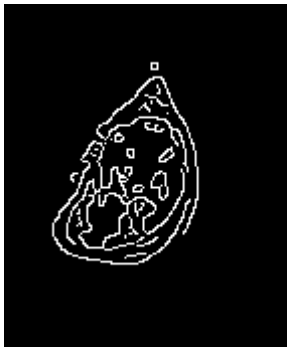
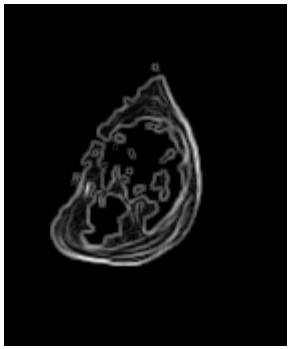


Measure of edge homogeneity - what fraction of edge pixels are in the first two bins of edge magnitude the histogram

values =0.3169 0.7264 1.4046 1.1739 0.2013
values = [A homogeneity maxminratio maxnextmaxratio sumdiff] ;

SNU1 dtx (#2, 22segment)

values =

0.3884 0.5480 1.9244 1.2962 0.2928

values = [A homogeneity maxminratio maxnextmaxratio sumdiff] ;

Balanced histogram thresholding

From Wikipedia, the free encyclopedia

In image processing, the **balanced histogram thresholding method** (BHT),[1] is a very simple method used for automatic image thresholding. Like Otsu's Method[2] and the **Iterative Selection Thesholding Method**,[3] this is a histogram based thresholding method. This approach assumes that the image is divided in two main classes: The **background** and the **foreground**. The **BHT** method tries to find the optimum threshold level that divides the histogram in two classes.

This method *weighs* the histogram, checks which of the two sides is heavier, and removes weight from the heavier side until it becomes the lighter. It repeats the same operation until the edges of the weighing scale meet.

Given its simplicity, this method is a good choice as a first approach when presenting the subject of *automatic image thresholding*.

Original image.

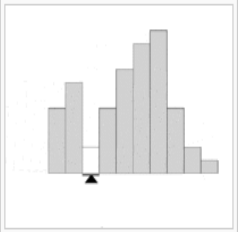
Algorithm [edit source | edit beta]

The following listing, in C notation, is a simplified version of the **Balanced Histogram Thresholding** method:

```
int BHThreshold(int[] histogram) {
    i_m = (int)((i_s + i_e) / 2.0f); // center of the weighing scale I_m
    w_l = get_weight(i_s, i_m + 1, histogram); // weight on the left W_l
    w_r = get_weight(i_m + 1, i_e + 1, histogram); // weight on the right W_r
    while (i_s <= i_e) {
        if (w_r > w_l) { // right side is heavier
            w_r -= histogram[i_e--];
            if (((i_s + i_e) / 2) < i_m) {
                w_r += histogram[i_m];
                w_l -= histogram[i_m--];
            }
        } else if (w_l >= w_r) { // left side is heavier
            w_l -= histogram[i_s++];
            if (((i_s + i_e) / 2) > i_m) {
                w_l += histogram[i_m + 1];
                w_r -= histogram[i_m + 1];
                i_m++;
            }
        }
    }
    return i_m;
}
```

Thresholded image.

This method may have problems when dealing with very noisy images, because the *weighing scale* may be misplaced. The problem can be minimized by ignoring the extremities of the histogram.[4]



Evolution of the method.

References [edit source | edit beta]

1. ^ A. Anjos and H. Shahbazkia. Bi-Level Image Thresholding - A Fast Method. BIOSIGNALS 2008. Vol:2. P:70-76.
2. ^ Nobuyuki Otsu (1979). "A threshold selection method from gray-level histograms". IEEE Trans. Sys., Man., Cyber. 9: 62–66.
3. ^ Ridler TW, Calvard S. (1978) Picture thresholding using an iterative selection method, IEEE Trans. System, Man and Cybernetics, SMC-8: 630-632.
4. ^ A. Anjos, R. Leite, M. L. Cancela, H. Shahbazkia. MAQ – A Bioinformatics Tool for Automatic Macroarray Analysis. International Journal of Computer Applications. 2010. Number 7 - Article 1.

External links [edit source | edit beta]

Correspondence

Picture Thresholding Using an Iterative Selection Method

T. W. RIDLER AND S. CALVARD

Abstract—**An object may be extracted from its background in a picture by threshold selection. Ideally, if the object has a different average gray level from that of its surrounding, the effect of thresholding will produce a white object with a black background or vice versa. In practice, it is often difficult, however, to select an appropriate threshold, and a technique is described whereby an optimum threshold may be chosen automatically as a result of an iterative process, successive iterations providing increasingly cleaner extractions of the object region. An application to low contrast images of handwritten text is discussed.**

I. Introduction

Features of interest in an image may often be extracted from their surroundings using a thresholding technique [1] in which all gray levels below the threshold are mapped into black, those levels above are mapped into white, or vice versa. The success of the technique depends on the object that is desired to be extracted occupying a range of gray levels distinct from that of the background. In practice it is difficult to select the optimum threshold, especially if a range of image scenes with widely differing properties is being considered, and some automatic means of threshold selection is required in each case. Thresholding at too high a level results in a loss of information, while thresholding at low levels can give rise to objectionable background clutter [2].

A common method of automatically deriving a threshold at which to segment a given picture is to examine its gray level

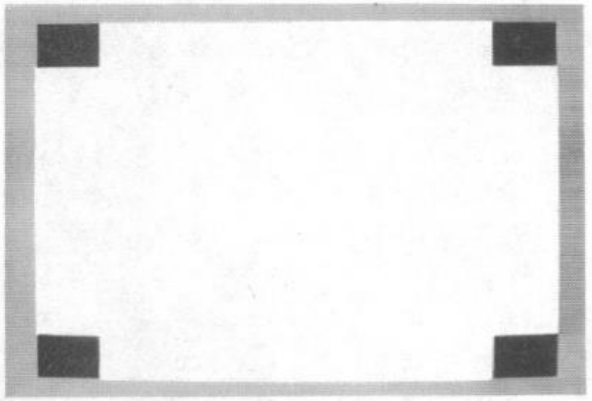


Fig. 1. Initial object-background estimation.

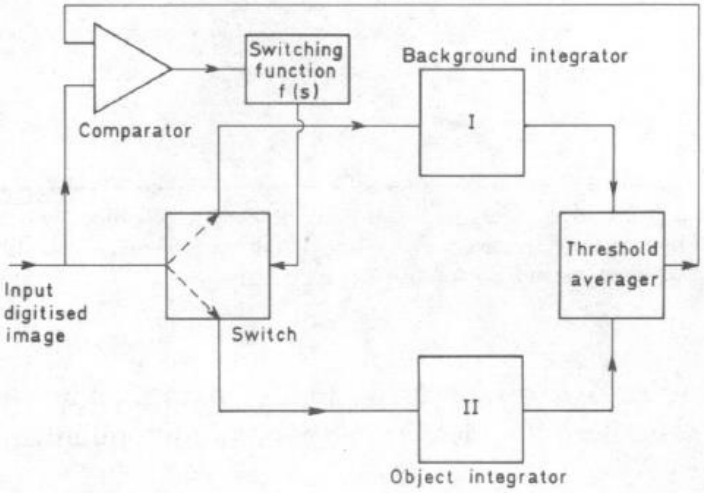
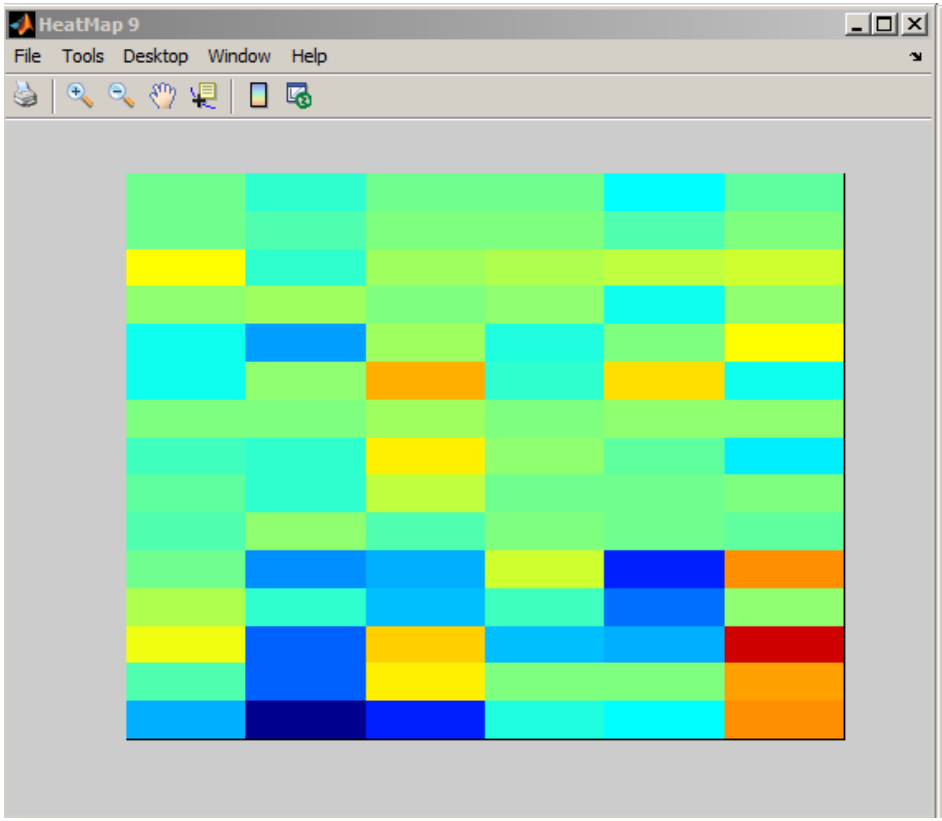
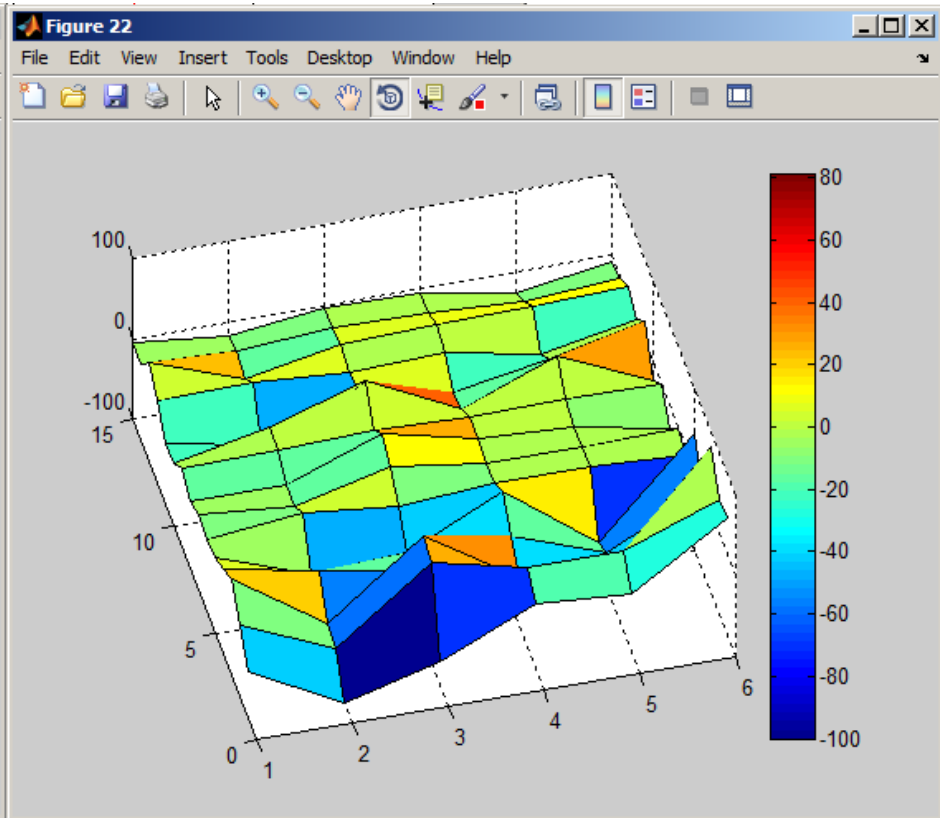


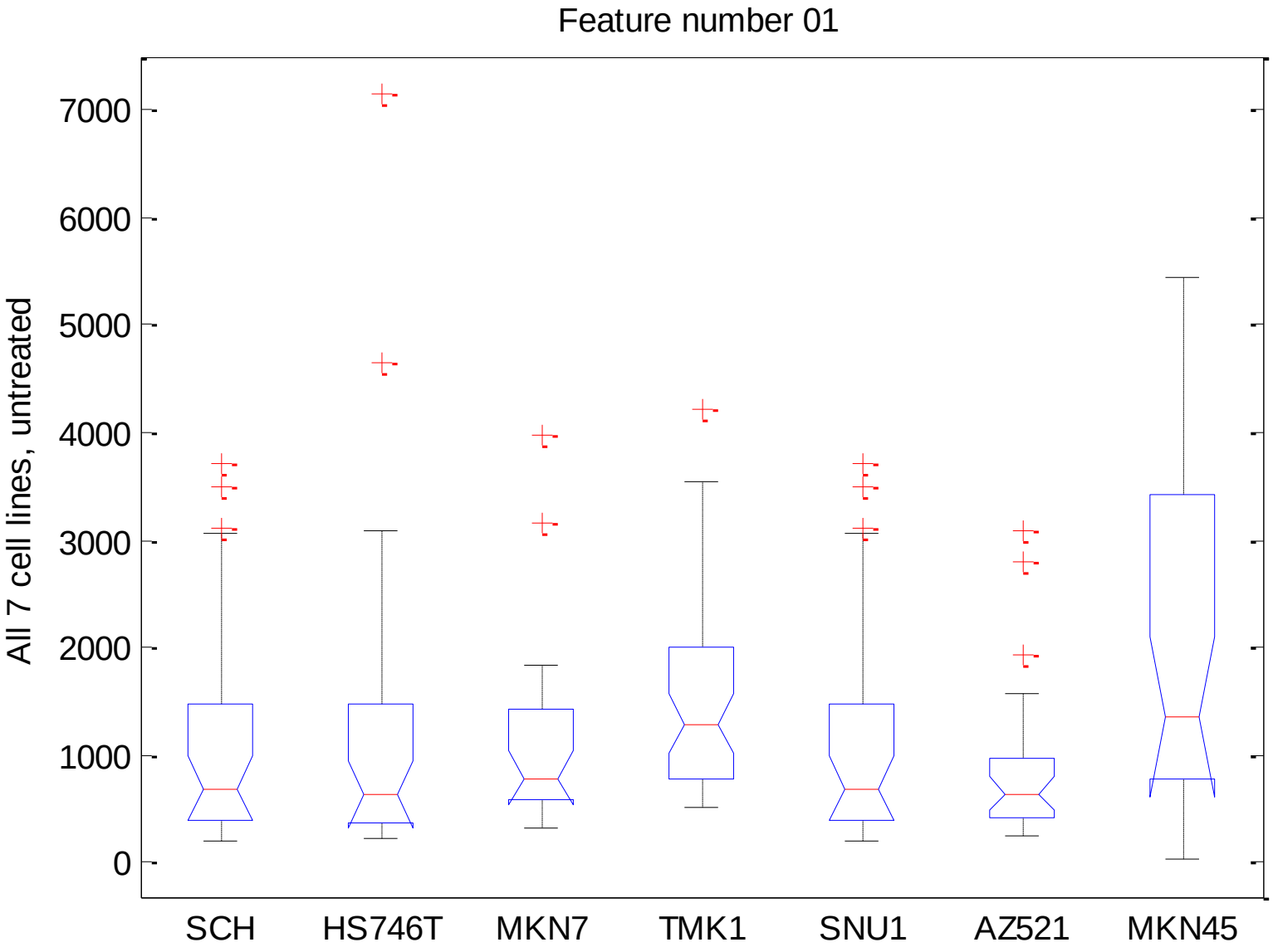
Fig. 2. Schematic image processor for iterative threshold selection.

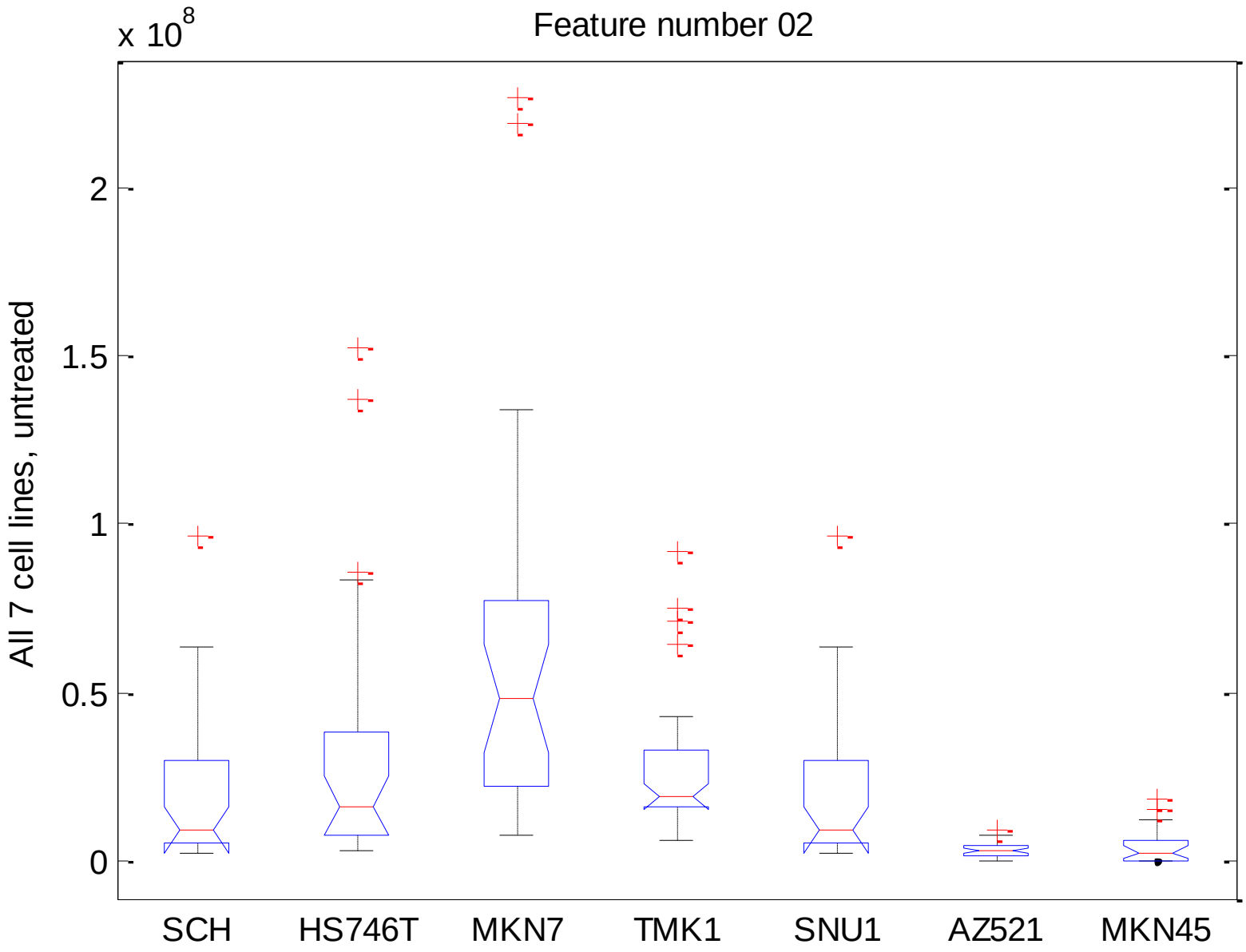
Heatmap, [med(Dtx)-med(Bsl)]/med(Bsl)*100

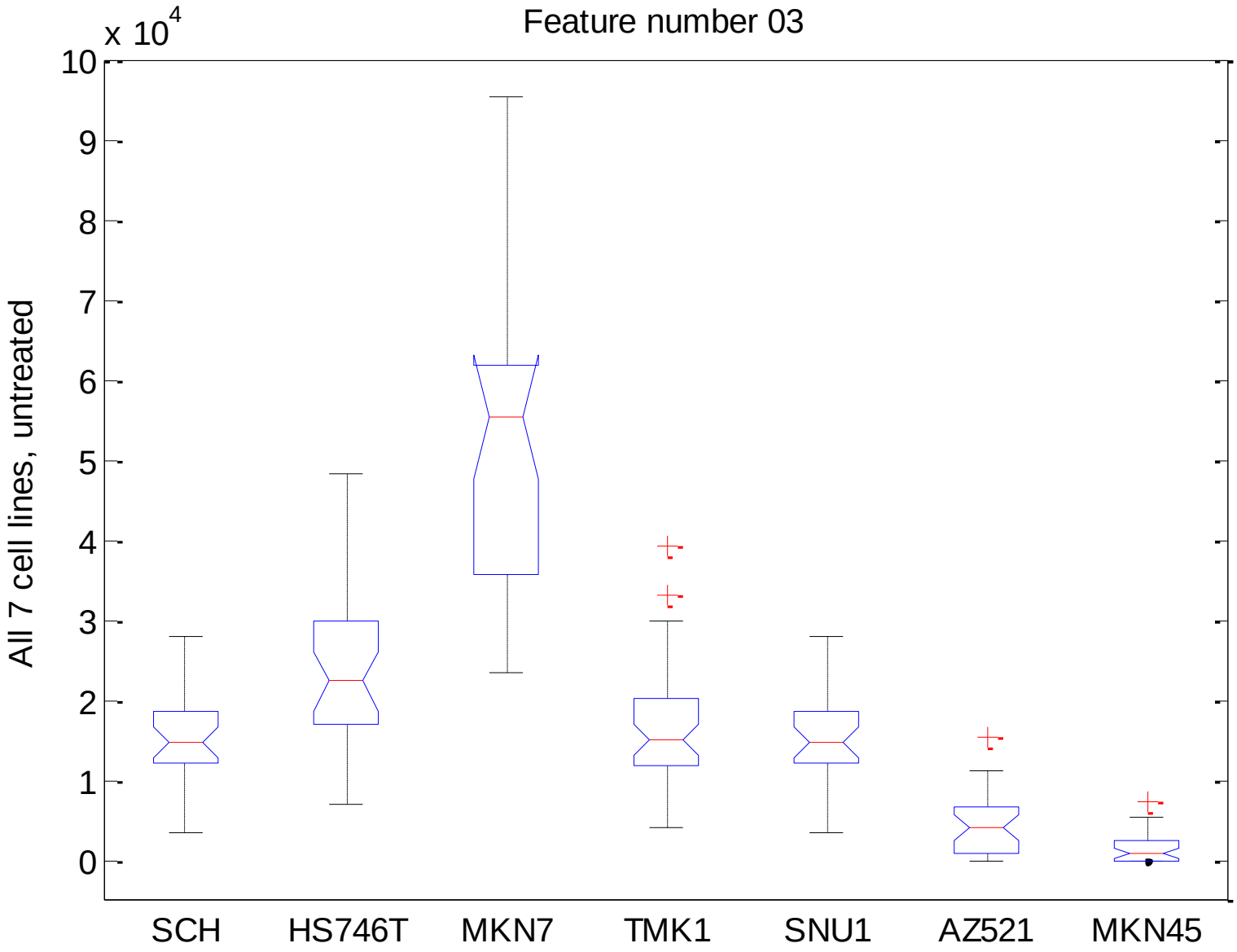
15 features x 6 cell lines

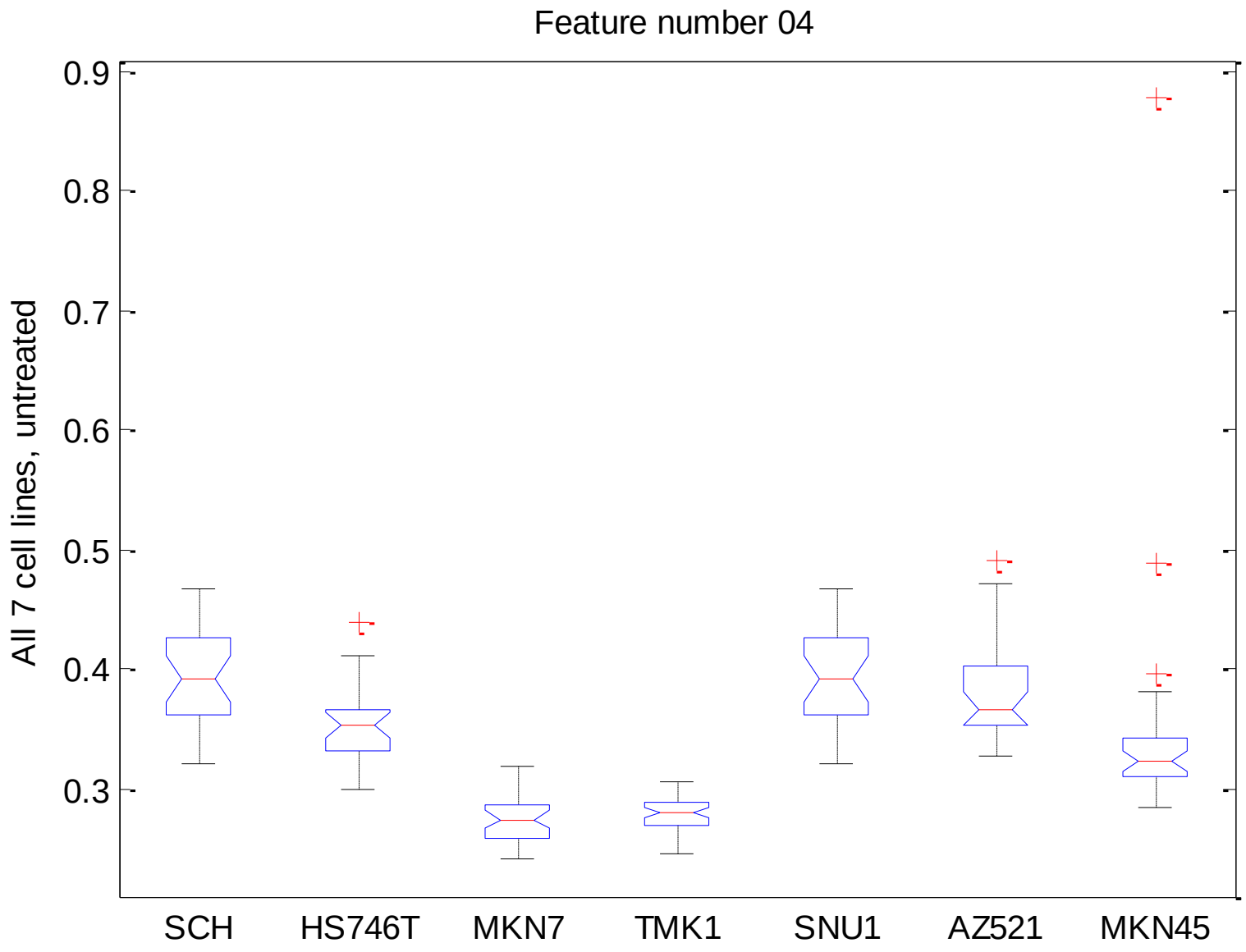


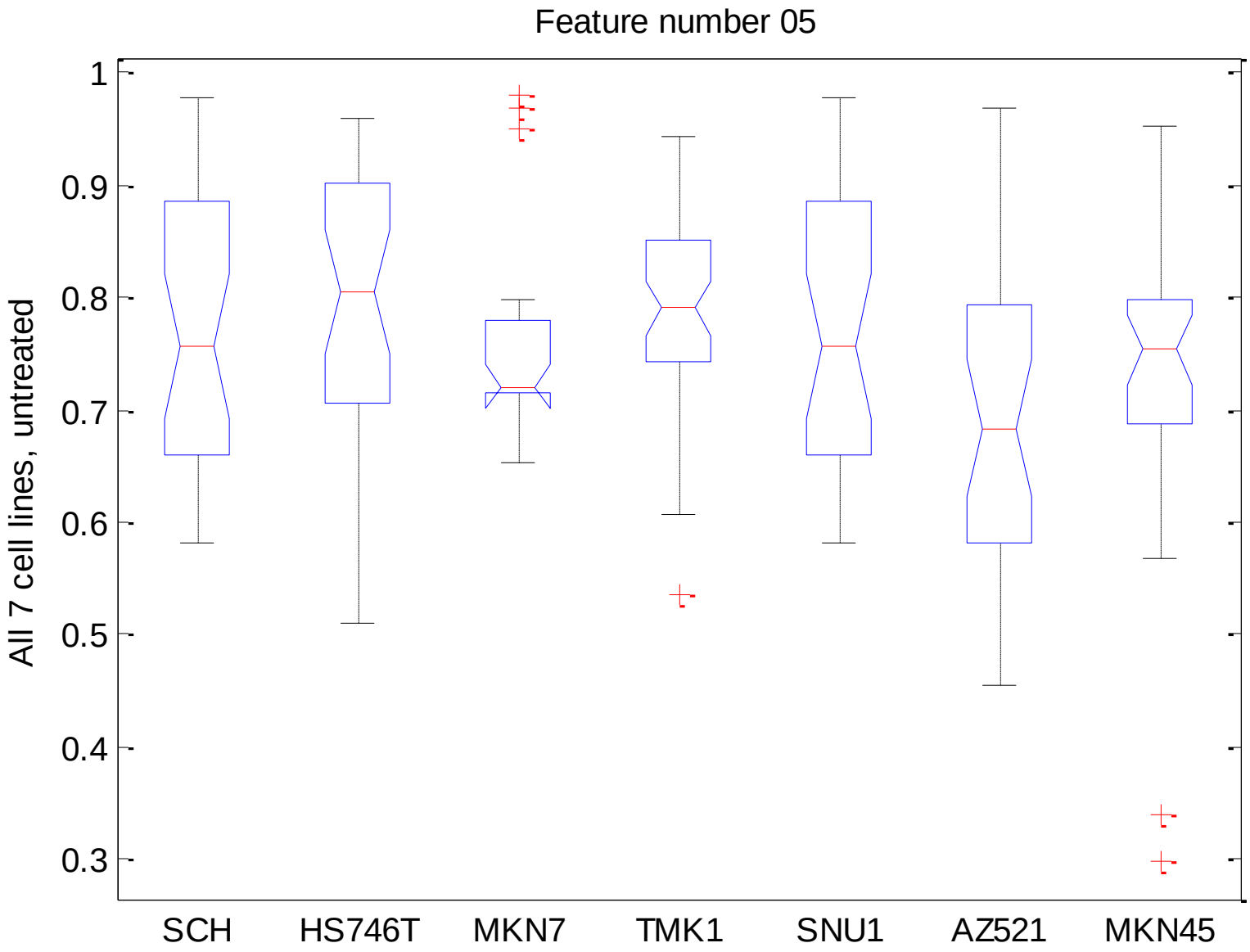


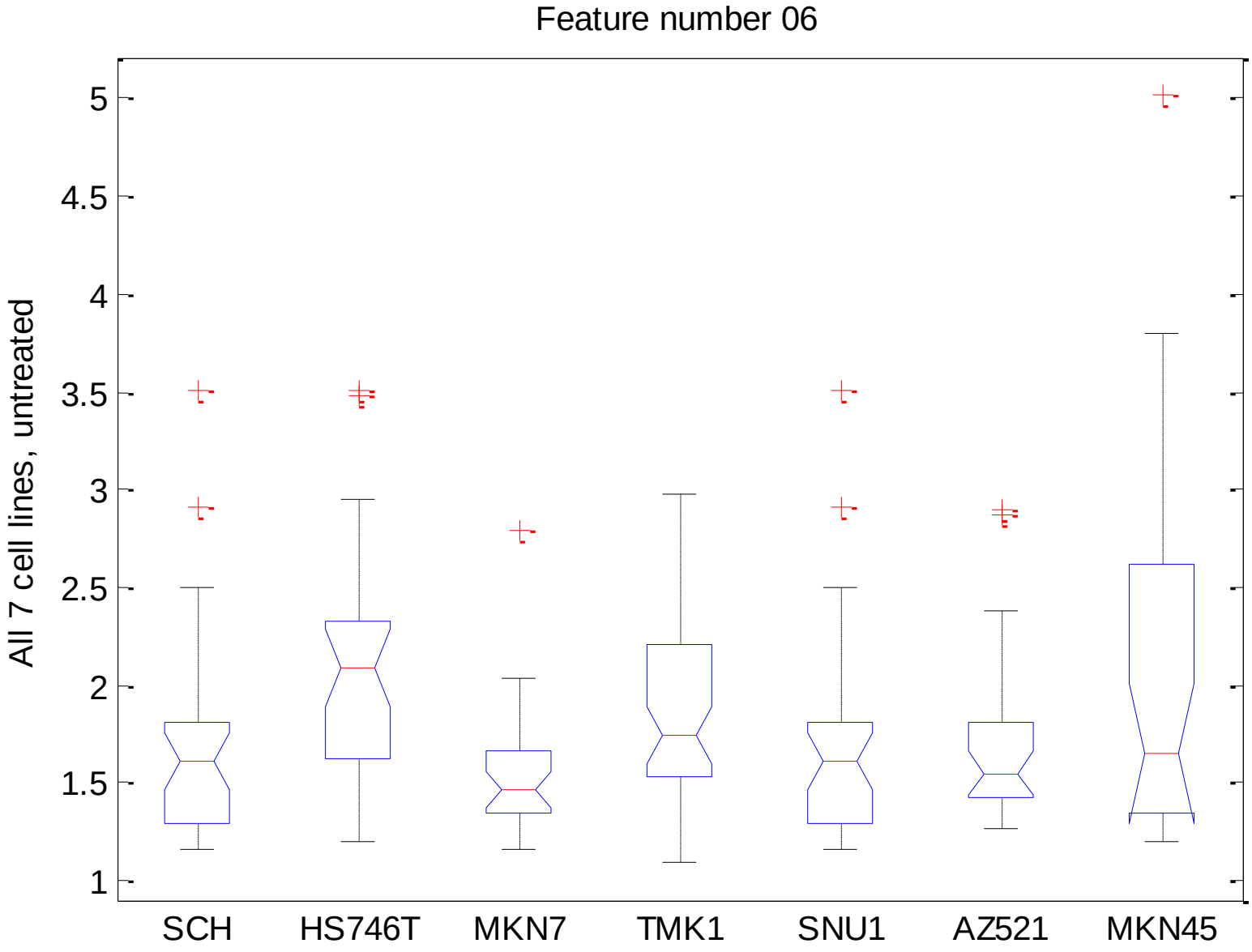

Feature number 01
All 7 cell lines, untreated
7000
6000
5000
4000
3000
2000
1000
0
SCH
HS746T
MKN7
TMK1
SNU1
AZ521
MKN45

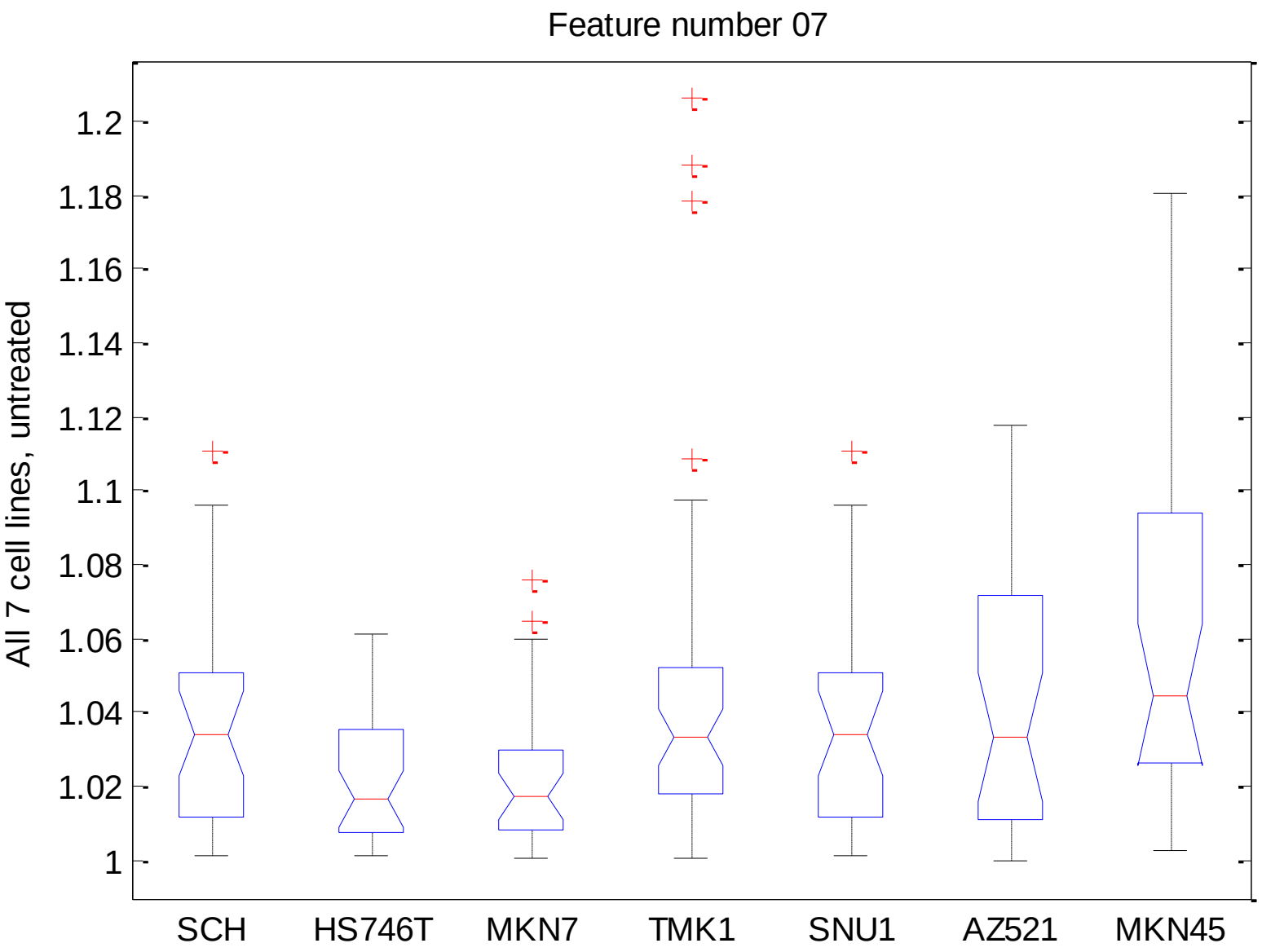

Feature number 02
x 10^8
All 7 cell lines, untreated
2
1.5
1
0.5
0
SCH
HS746T
MKN7
TMK1
SNU1
AZ521
MKN45

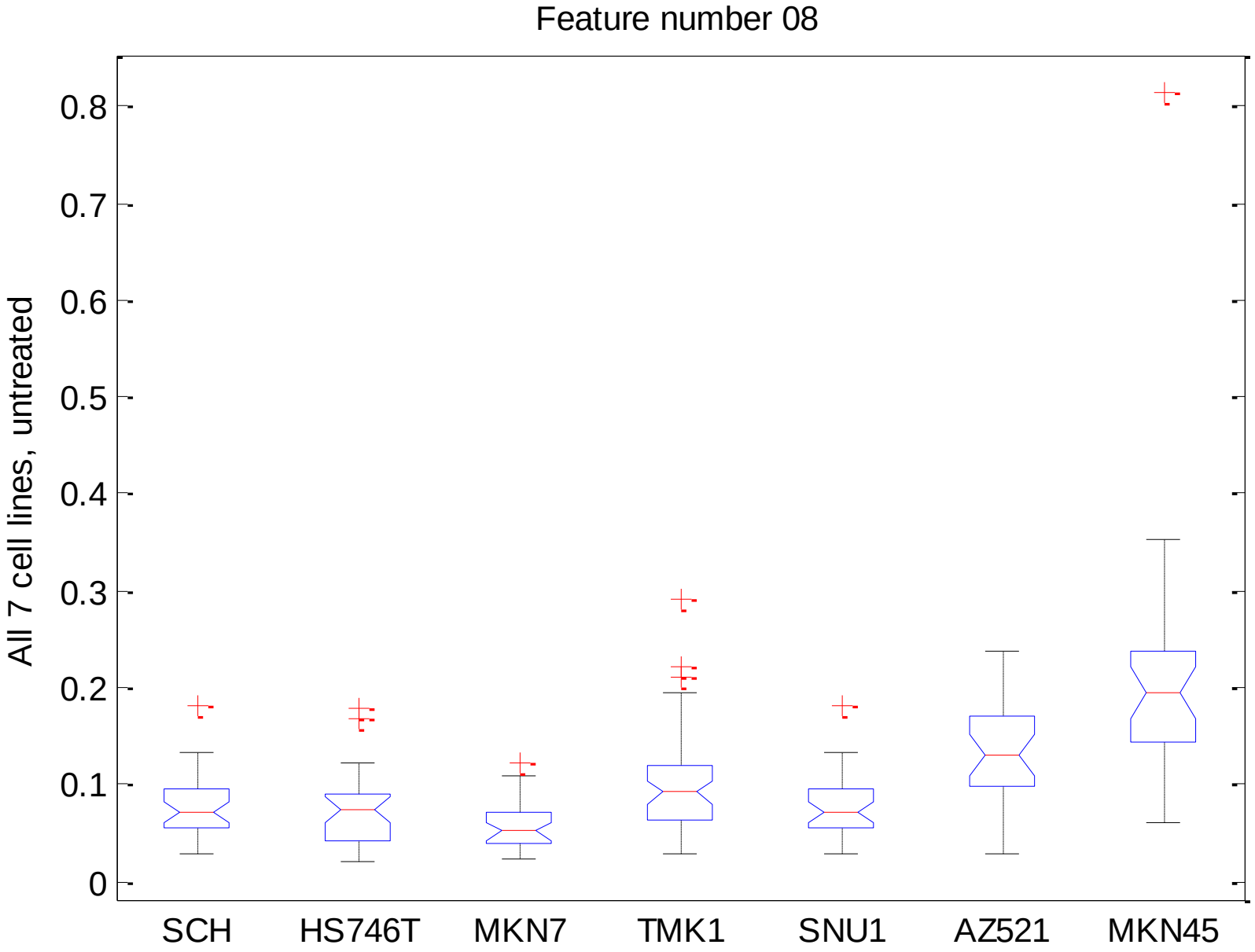

Feature number 03
x 10^4
10
9
8
7
6
5
4
3
2
1
0
All 7 cell lines, untreated
SCH
HS746T
MKN7
TMK1
SNU1
AZ521
MKN45

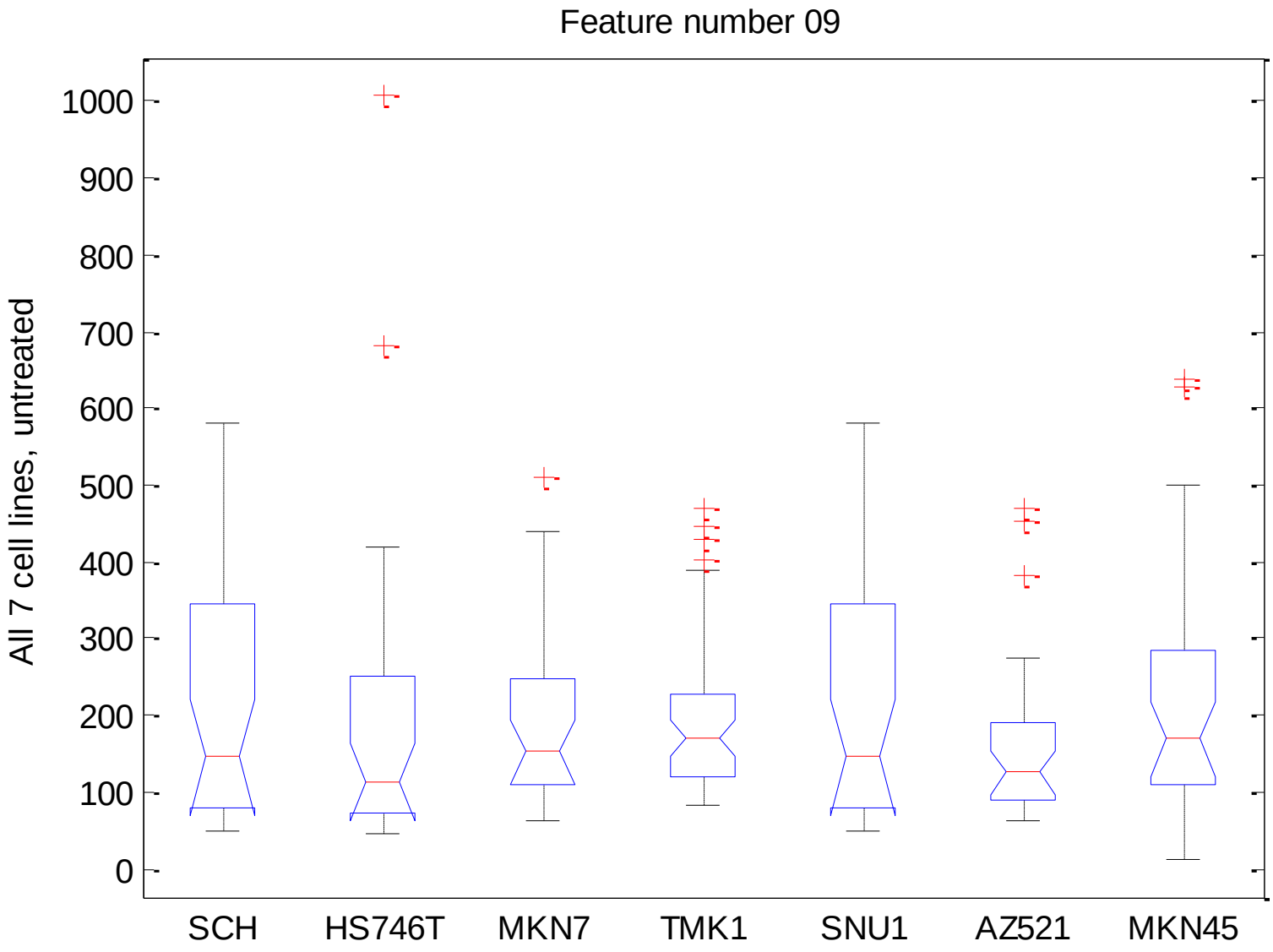

Feature number 04
All 7 cell lines, untreated
0.9
0.8
0.7
0.6
0.5
0.4
0.3
SCH
HS746T
MKN7
TMK1
SNU1
AZ521
MKN45

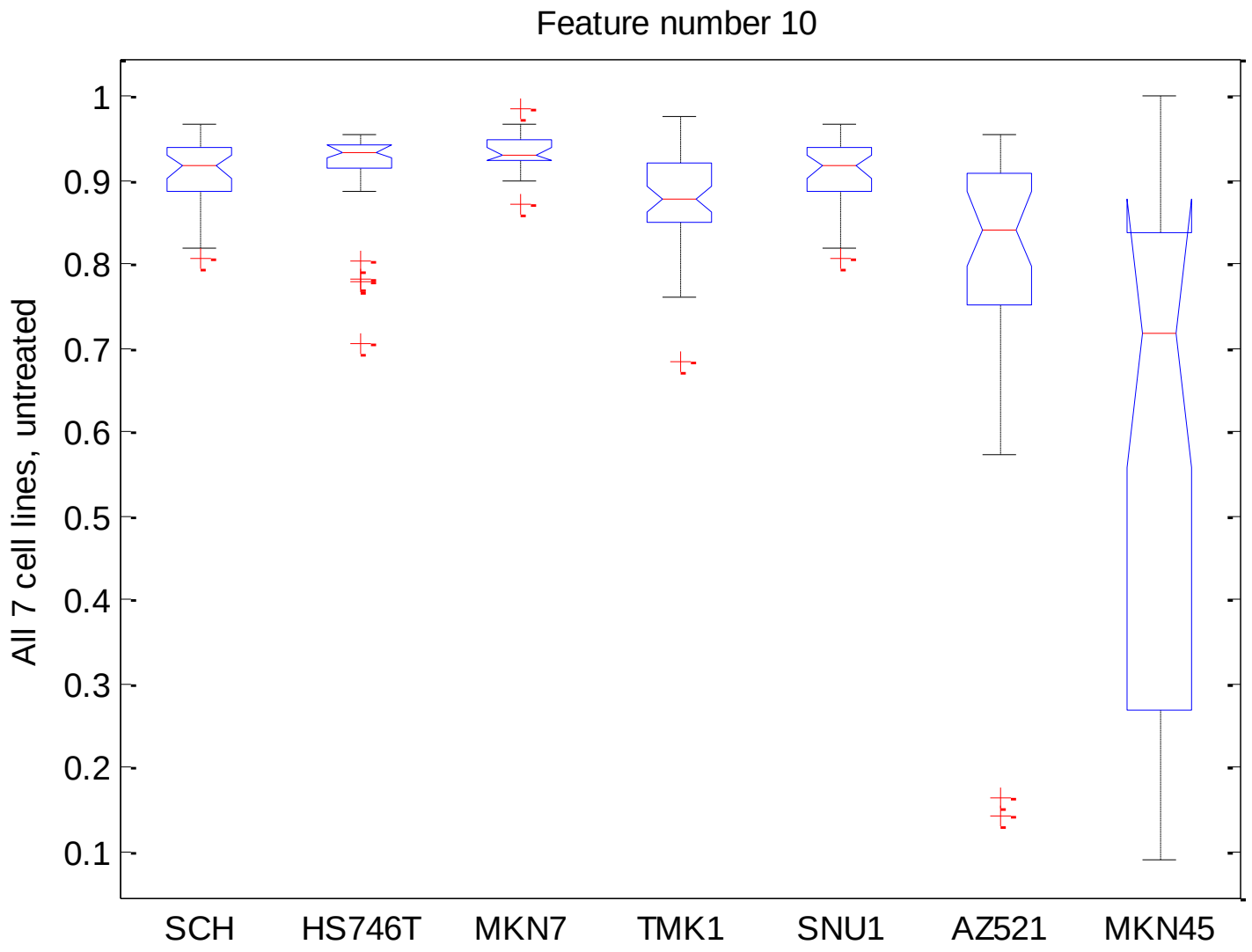

Feature number 05
All 7 cell lines, untreated
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
SCH
HS746T
MKN7
TMK1
SNU1
AZ521
MKN45

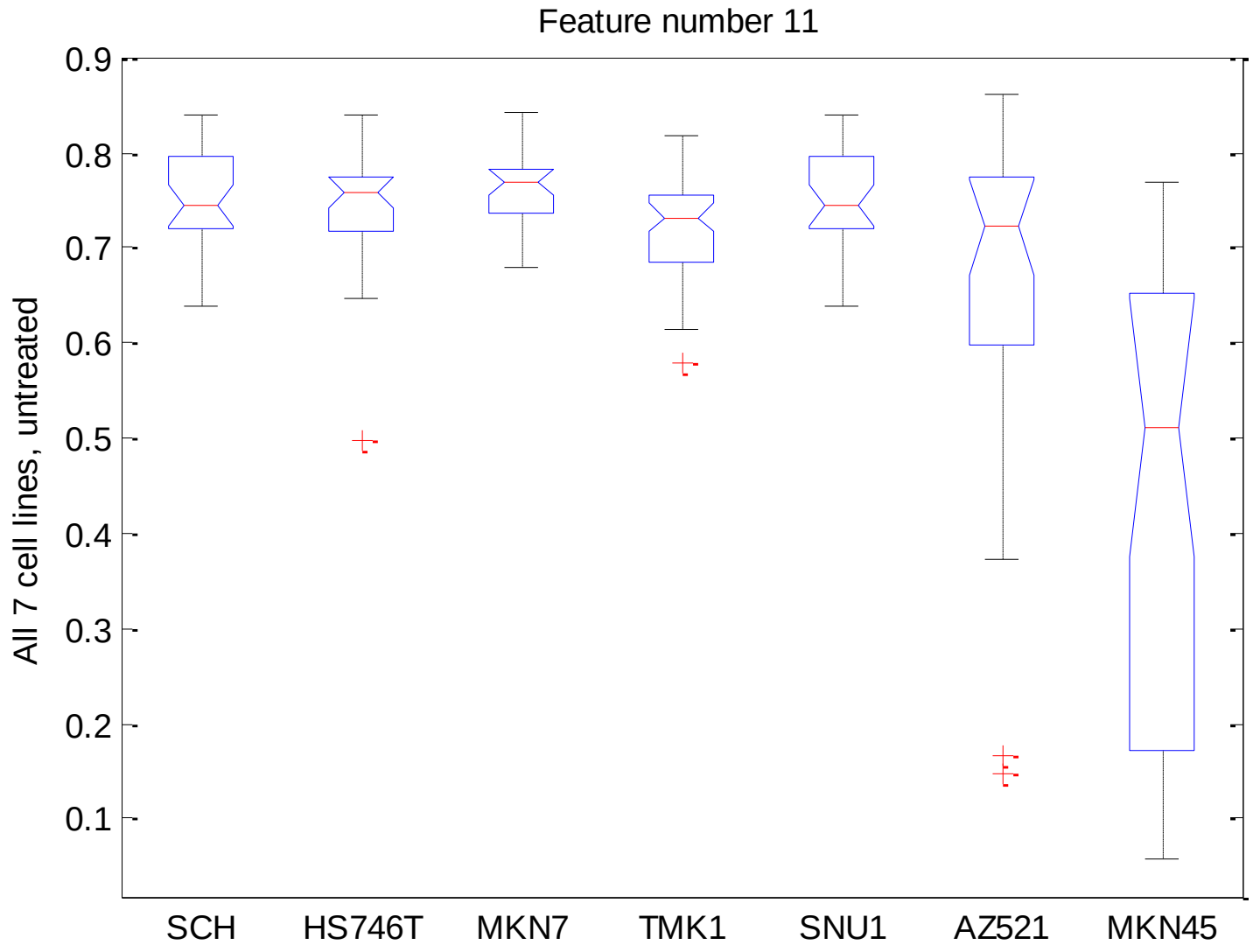

Feature number 06
All 7 cell lines, untreated
5
4.5
4
3.5
3
2.5
2
1.5
1
SCH
HS746T
MKN7
TMK1
SNU1
AZ521
MKN45

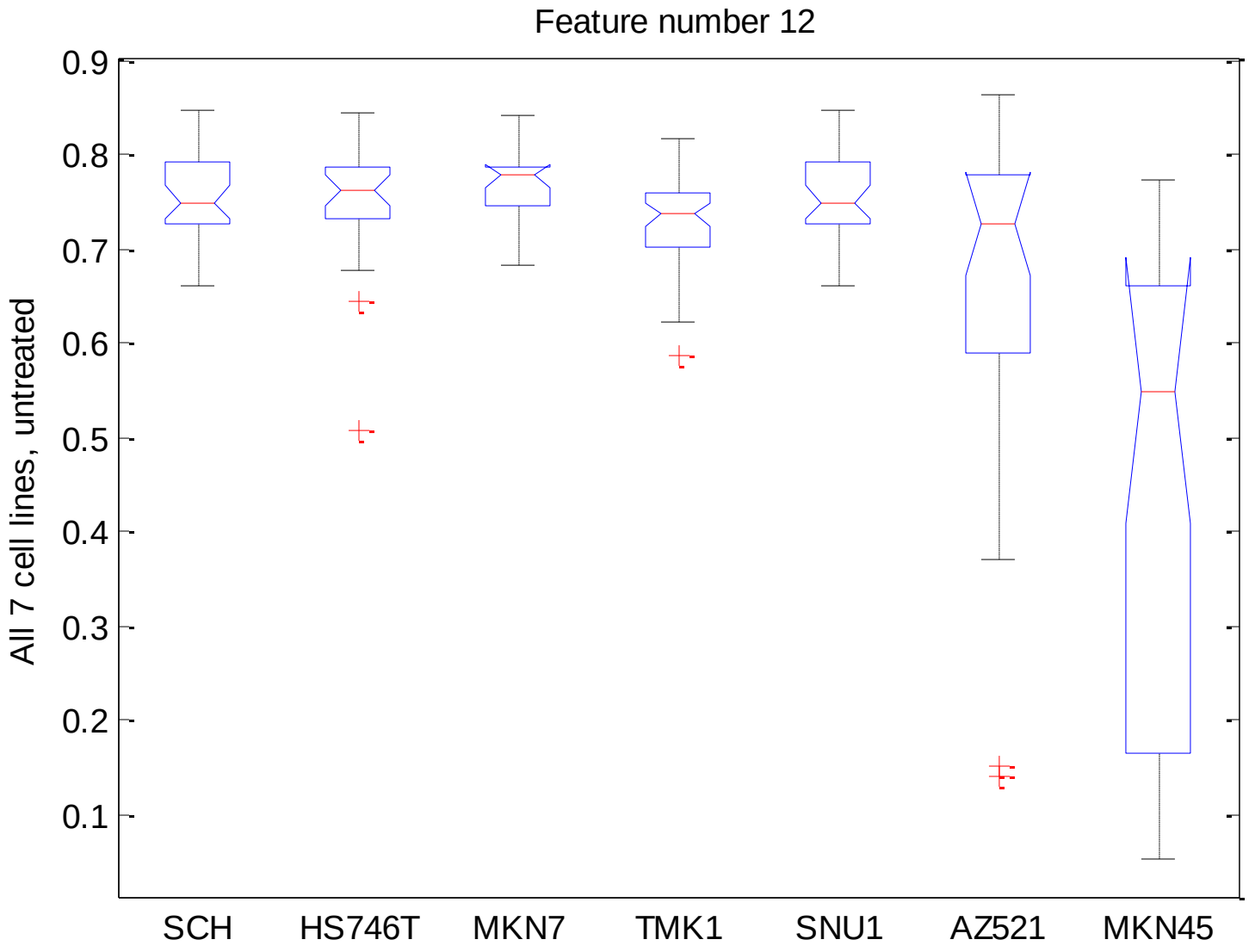

Feature number 07
All 7 cell lines, untreated
1.2
1.18
1.16
1.14
1.12
1.1
1.08
1.06
1.04
1.02
1
SCH
HS746T
MKN7
TMK1
SNU1
AZ521
MKN45

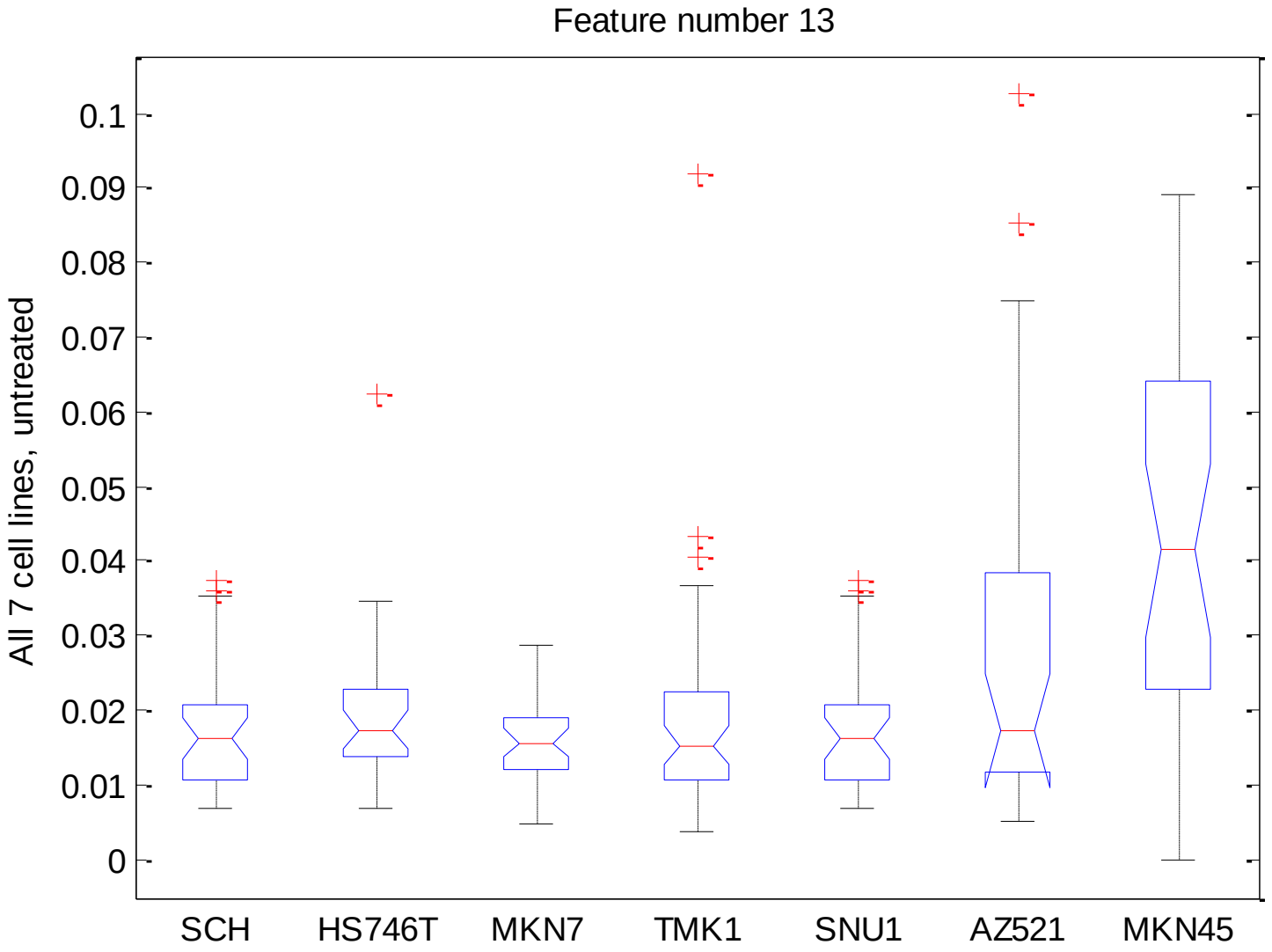

Feature number 08
All 7 cell lines, untreated
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
SCH
HS746T
MKN7
TMK1
SNU1
AZ521
MKN45


Feature number 09
All 7 cell lines, untreated
1000
900
800
700
600
500
400
300
200
100
0
SCH
HS746T
MKN7
TMK1
SNU1
AZ521
MKN45

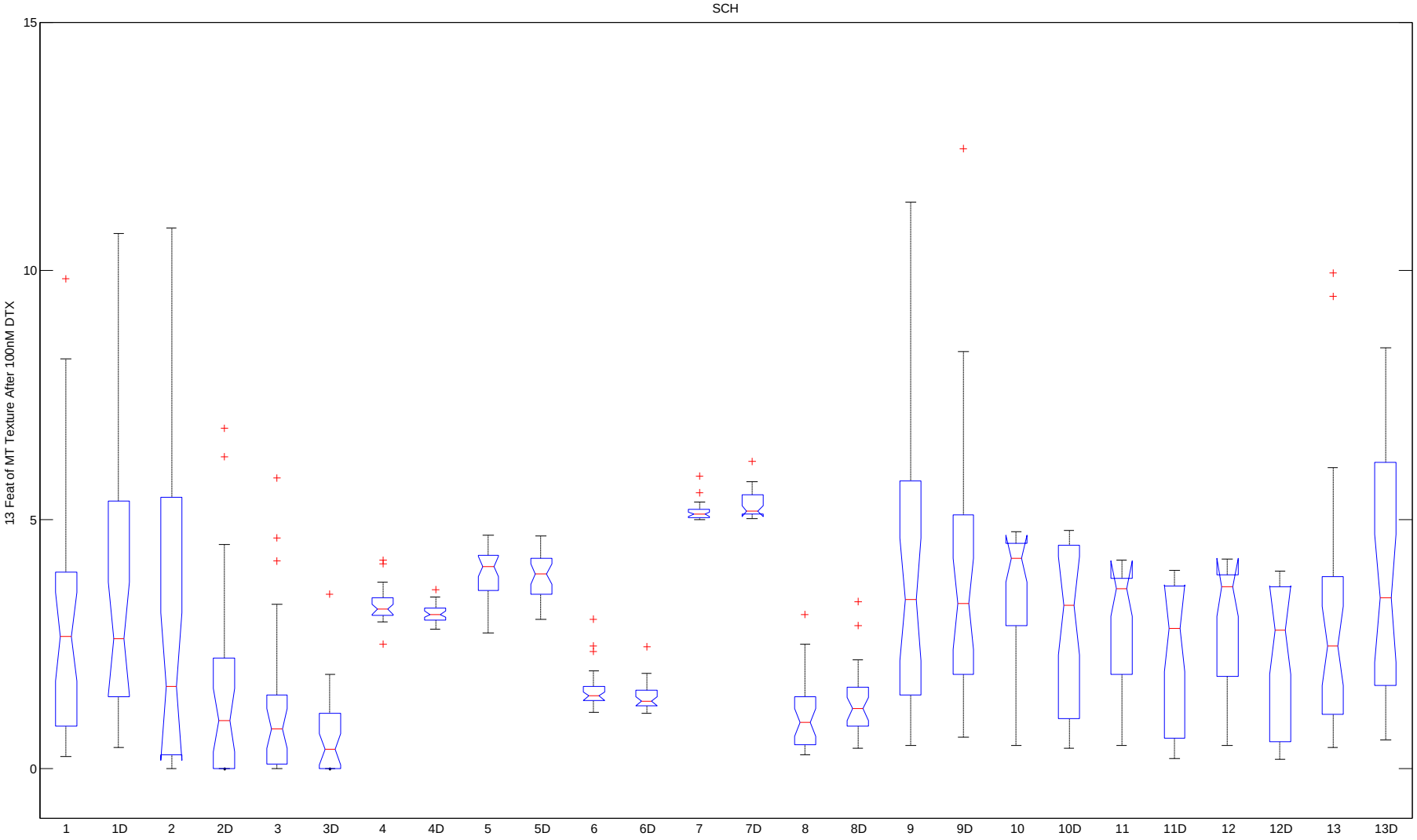

Feature number 10
All 7 cell lines, untreated
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
SCH
HS746T
MKN7
TMK1
SNU1
AZ521
MKN45


Feature number 11
All 7 cell lines, untreated
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
SCH
HS746T
MKN7
TMK1
SNU1
AZ521
MKN45


Feature number 12
All 7 cell lines, untreated
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
SCH
HS746T
MKN7
TMK1
SNU1
AZ521
MKN45

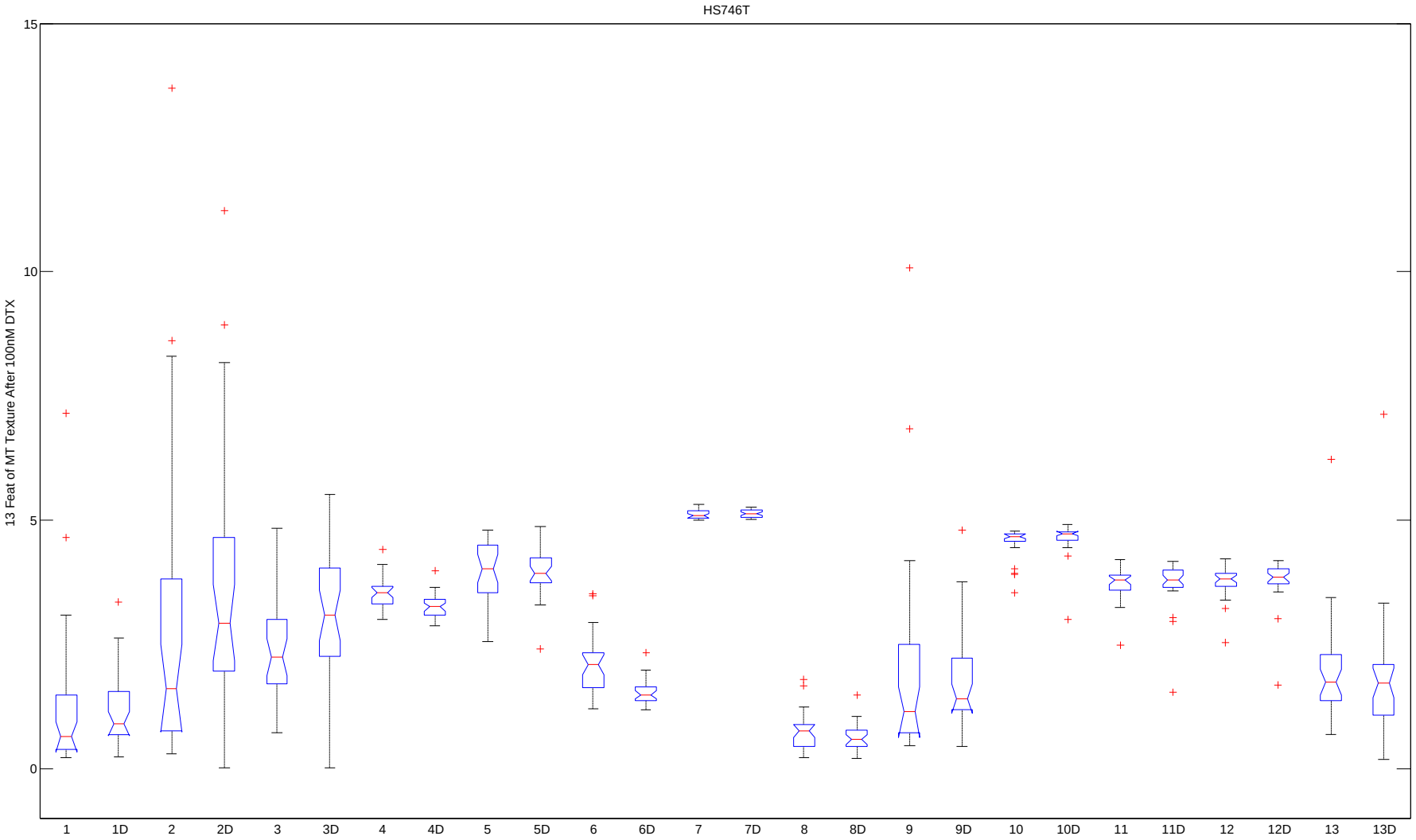

Feature number 13
All 7 cell lines, untreated
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
SCH
HS746T
MKN7
TMK1
SNU1
AZ521
MKN45

Cell Line : SCH – resistant



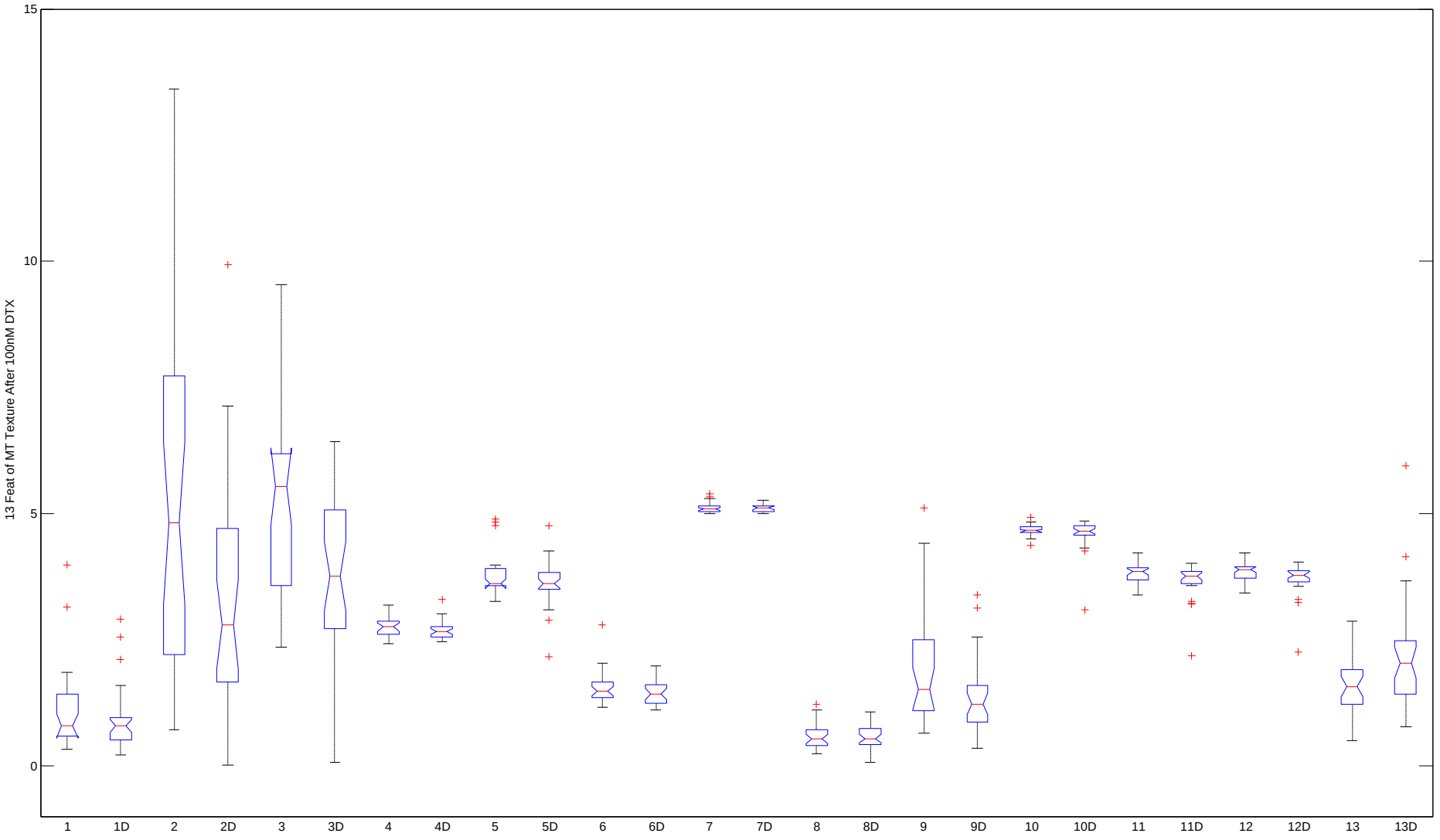
- Cell Line:SCH
- FeatureName p<0.05 p
- object_size:average 0 0.9385
- object_size:variance 0 0.1852
- object_size:ratio 0 0.0530
- edges:area_fraction 1 0.0177
- edges:homogeneity 0 0.9207
- edges:direction_maxmin_ratio 0 0.1135
- edges:direction_maxnextmax_ratio 1 0.0073
- edges:direction_difference 0 0.2082
- obj_skel_len 0 0.5825
- obj_skel_hull_area_ratio 0 0.1697
- obj_skel_obj_area_ratio 0 0.1122
- obj_skel_obj_fluor_ratio 0 0.0993
- obj_skel_branch_per_len 0 0.2128

Cell Line : HS746T – resistant



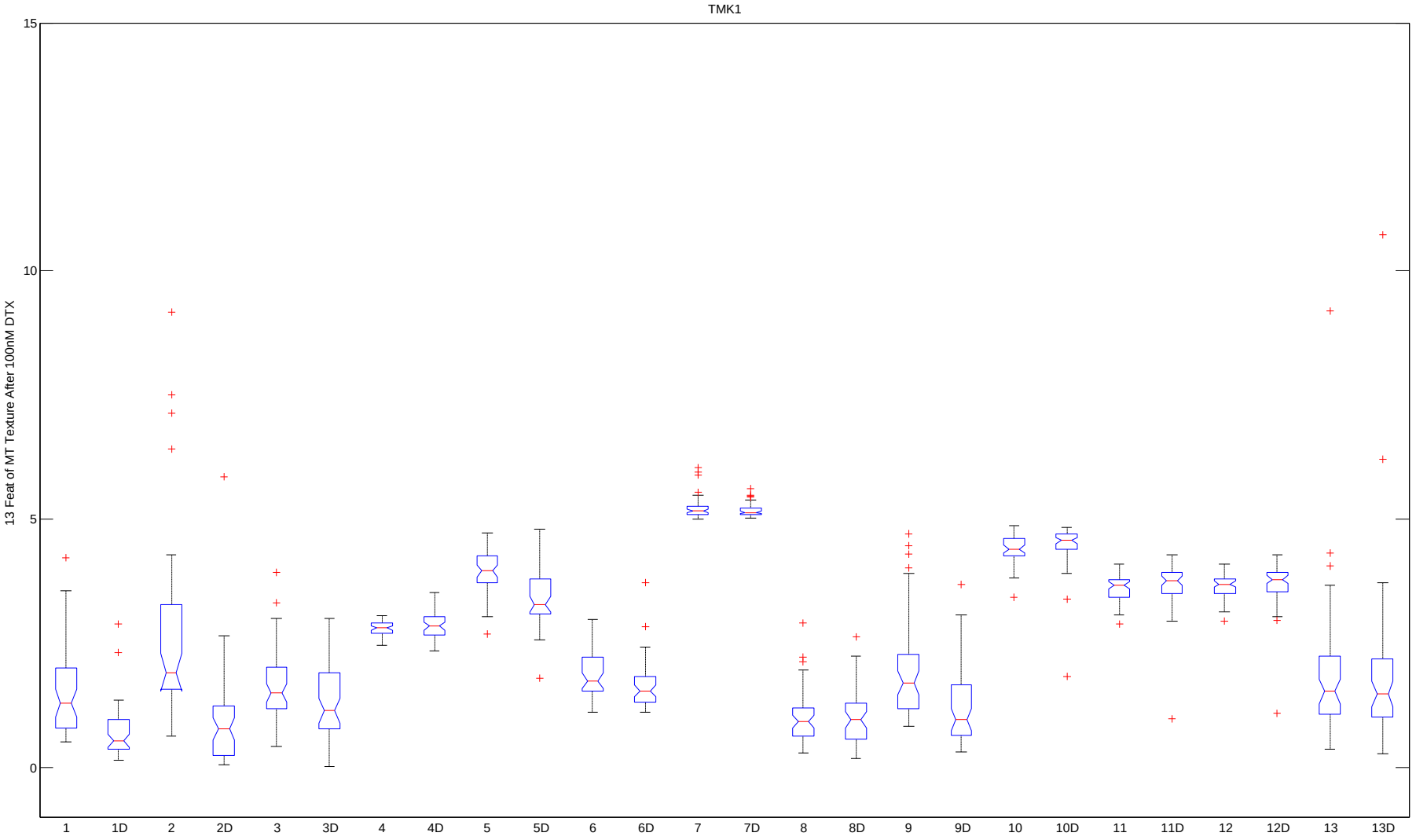
- Cell Line:HS746T
- FeatureName p<0.05 p
- object_size:average 0 0.6876
- object_size:variance 0 0.5237
- object_size:ratio 1 0.0279
- edges:area_fraction 1 0.0011
- edges:homogeneity 0 0.7999
- edges:direction_maxmin_ratio 1 0.0004
- edges:direction_maxnextmax_ratio 0 0.4087
- edges:direction_difference 0 0.1352
- obj_skel_len 0 0.5460
- obj_skel_hull_area_ratio 0 0.3784
- obj_skel_obj_area_ratio 0 0.9464
- obj_skel_obj_fluor_ratio 0 0.9765
- obj_skel_branch_per_len 0 0.4570

Cell Line : MKN7 – resistant



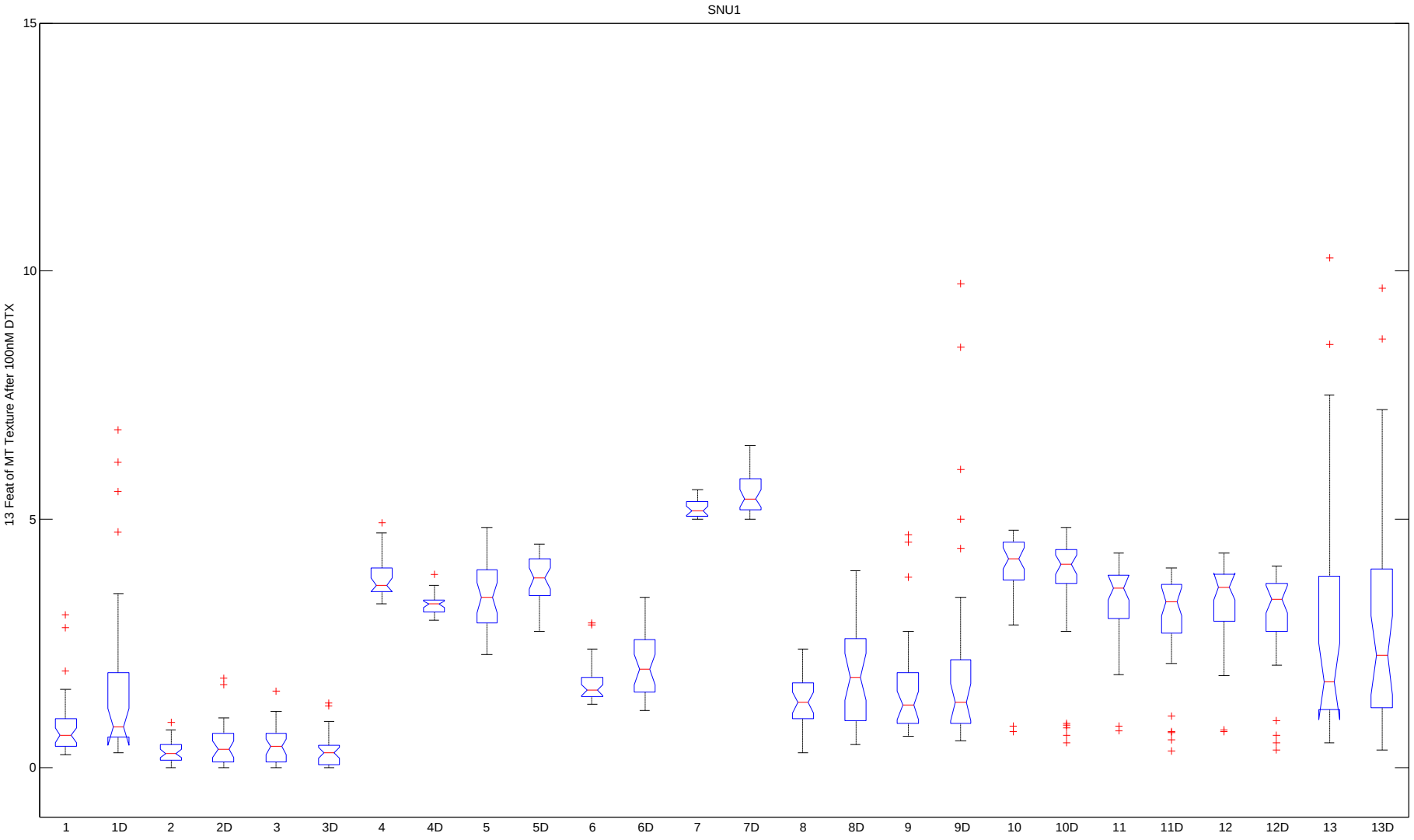
- Cell Line:MKN7
- FeatureName p<0.05 p
- object_size:average 0 0.3804
- object_size:variance 1 0.0271
- object_size:ratio 1 0.0010
- edges:area_fraction 0 0.2227
- edges:homogeneity 0 0.1884
- edges:direction_maxmin_ratio 0 0.0769
- edges:direction_maxnextmax_ratio 0 0.6265
- edges:direction_difference 0 0.5017
- obj_skel_len 0 0.1380
- obj_skel_hull_area_ratio 0 0.2841
- obj_skel_obj_area_ratio 0 0.0624
- obj_skel_obj_fluor_ratio 0 0.0621
- obj_skel_branch_per_len 1 0.0288

Cell Line : TMK1 – sensitive



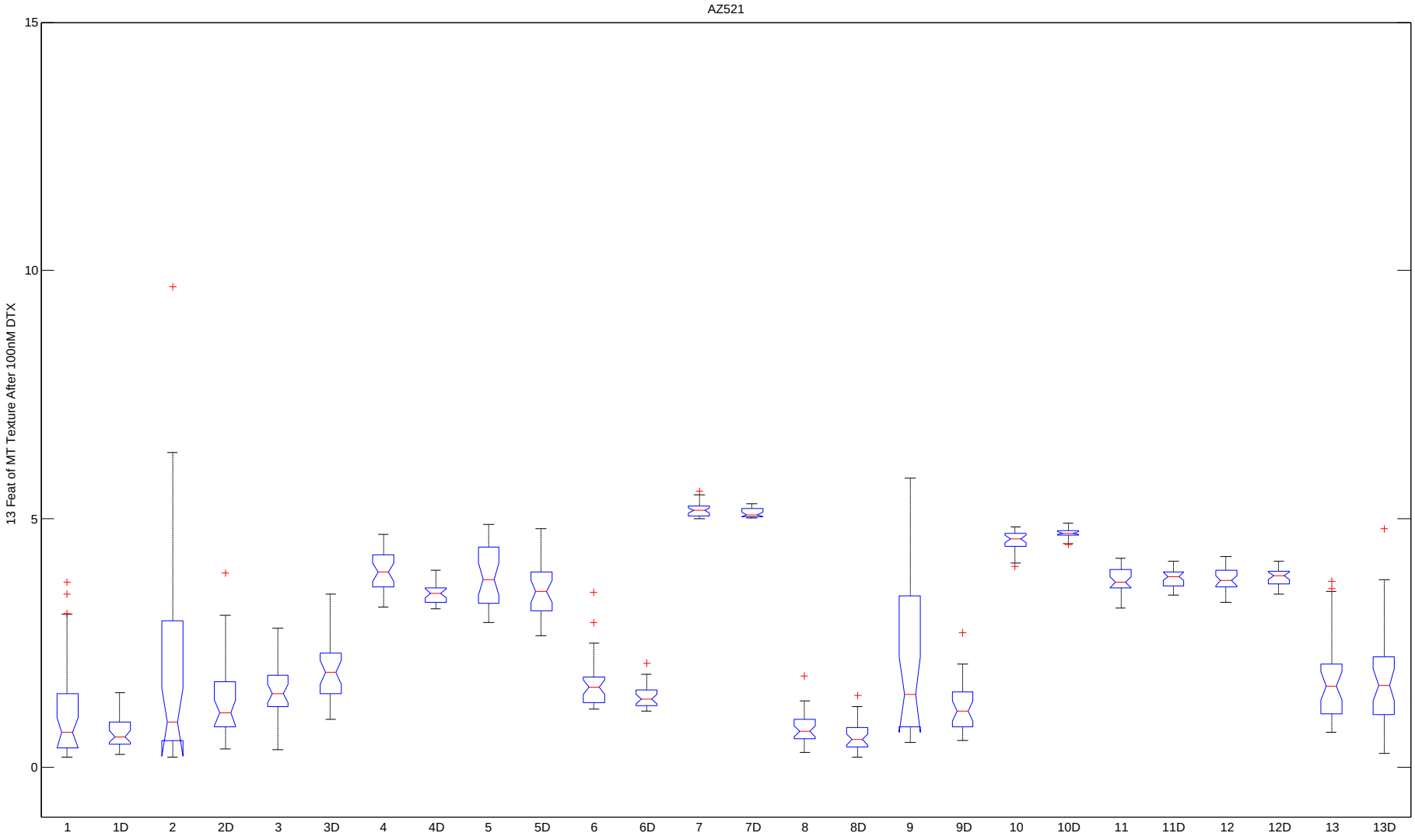
- Cell Line:TMK1
- FeatureName p<0.05 p
- object_size:average 1 0.0000
- object_size:variance 1 0.0000
- object_size:ratio 1 0.0079
- edges:area_fraction 0 0.0586
- edges:homogeneity 1 0.0000
- edges:direction_maxmin_ratio 1 0.0267
- edges:direction_maxnextmax_ratio 0 0.2249
- edges:direction_difference 0 0.6824
- obj_skel_len 1 0.0001
- obj_skel_hull_area_ratio 0 0.2916
- obj_skel_obj_area_ratio 0 0.4505
- obj_skel_obj_fluor_ratio 0 0.4701
- obj_skel_branch_per_len 0 0.8401

Cell Line : SNU1 - sensitive



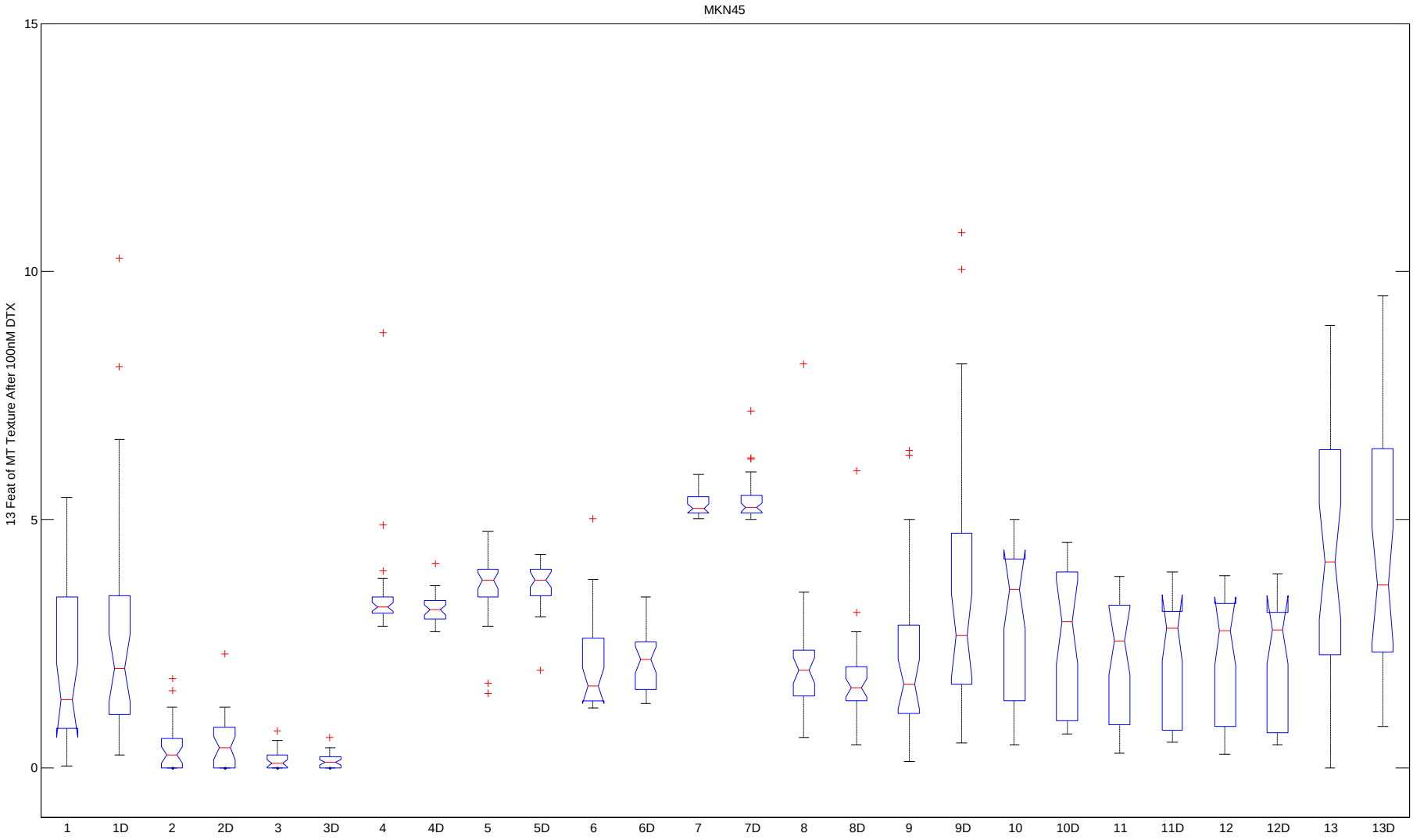
- Cell Line:SNU1
- FeatureName p<0.05 p
- object_size:average 1 0.0390
- object_size:variance 0 0.1368
- object_size:ratio 0 0.3181
- edges:area_fraction 1 0.0000
- edges:homogeneity 1 0.0435
- edges:direction_maxmin_ratio 1 0.0021
- edges:direction_maxnextmax_ratio 1 0.0003
- edges:direction_difference 1 0.0080
- obj_skel_len 0 0.1965
- obj_skel_hull_area_ratio 0 0.2808
- obj_skel_obj_area_ratio 0 0.1603
- obj_skel_obj_fluor_ratio 0 0.1551
- obj_skel_branch_per_len 0 0.8024

Cell Line : AZ521 - sensitive

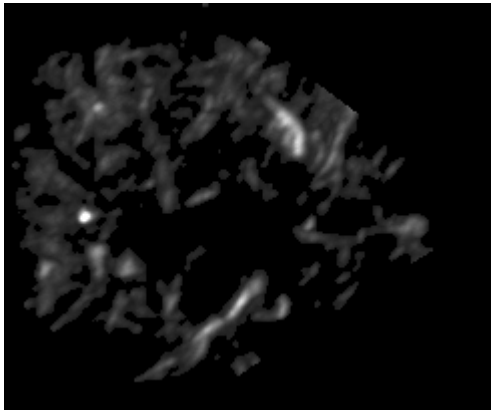


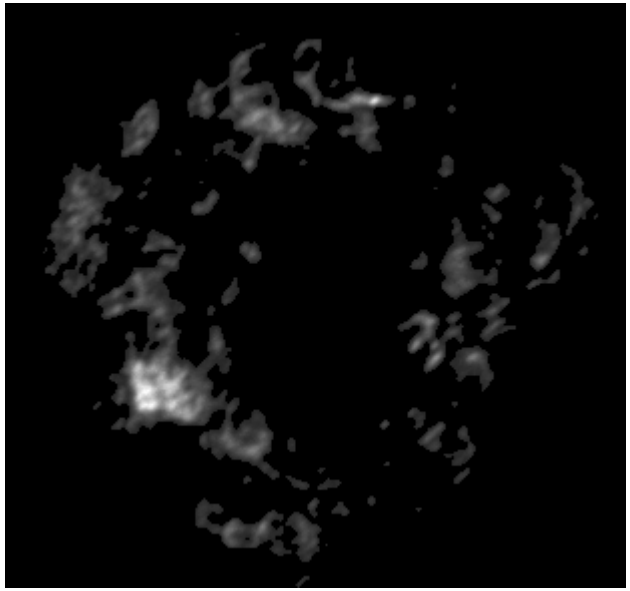
- Cell Line:AZ521
- FeatureName p<0.05 p
- object_size:average 1 0.0254
- object_size:variance 0 0.1188
- object_size:ratio 1 0.0263
- edges:area_fraction 1 0.0000
- edges:homogeneity 0 0.1146
- edges:direction_maxmin_ratio 1 0.0136
- edges:direction_maxnextmax_ratio 0 0.0723
- edges:direction_difference 0 0.1317
- obj_skel_len 1 0.0038
- obj_skel_hull_area_ratio 1 0.0003
- obj_skel_obj_area_ratio 0 0.6782
- obj_skel_obj_fluor_ratio 0 0.8323
- obj_skel_branch_per_len 0 0.9507

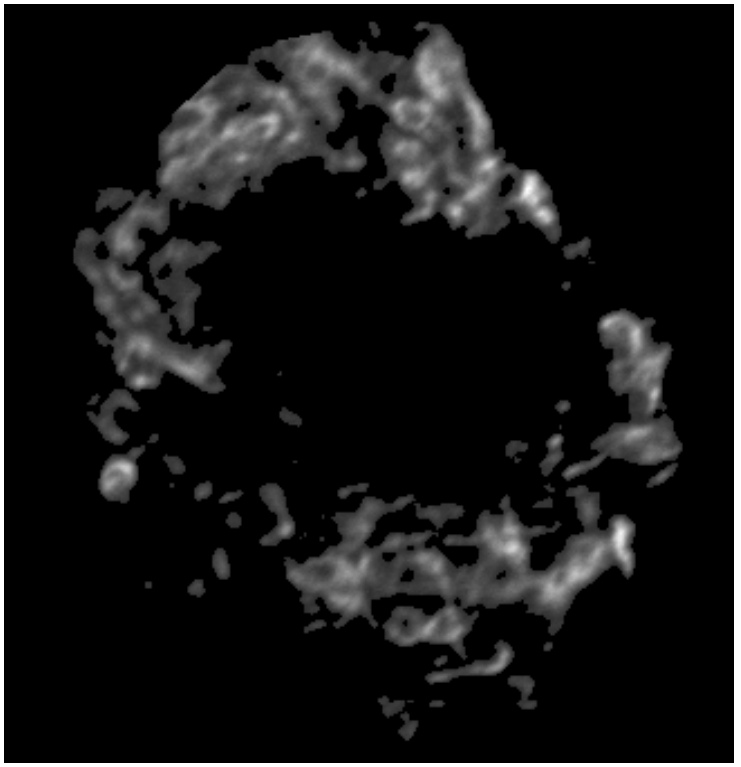
Cell Line : MKN45 – sensitive, intestinal



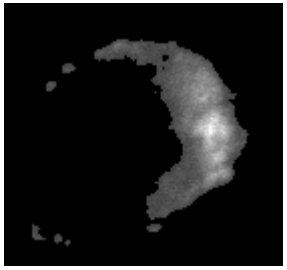
- Cell Line:MKN45
- FeatureName p<0.05 p
- object_size:average 0 0.2418
- object_size:variance 0 0.6713
- object_size:ratio 0 0.7595
- edges:area_fraction 0 0.1960
- edges:homogeneity 0 0.6897
- edges:direction_maxmin_ratio 0 0.5853
- edges:direction_maxnextmax_ratio 0 0.3975
- edges:direction_difference 0 0.3530
- obj_skel_len 1 0.0490
- obj_skel_hull_area_ratio 0 0.4678
- obj_skel_obj_area_ratio 0 0.9842
- obj_skel_obj_fluor_ratio 0 0.9266
- obj_skel_branch_per_len 0 0.7770

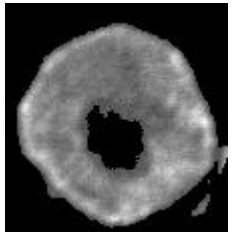






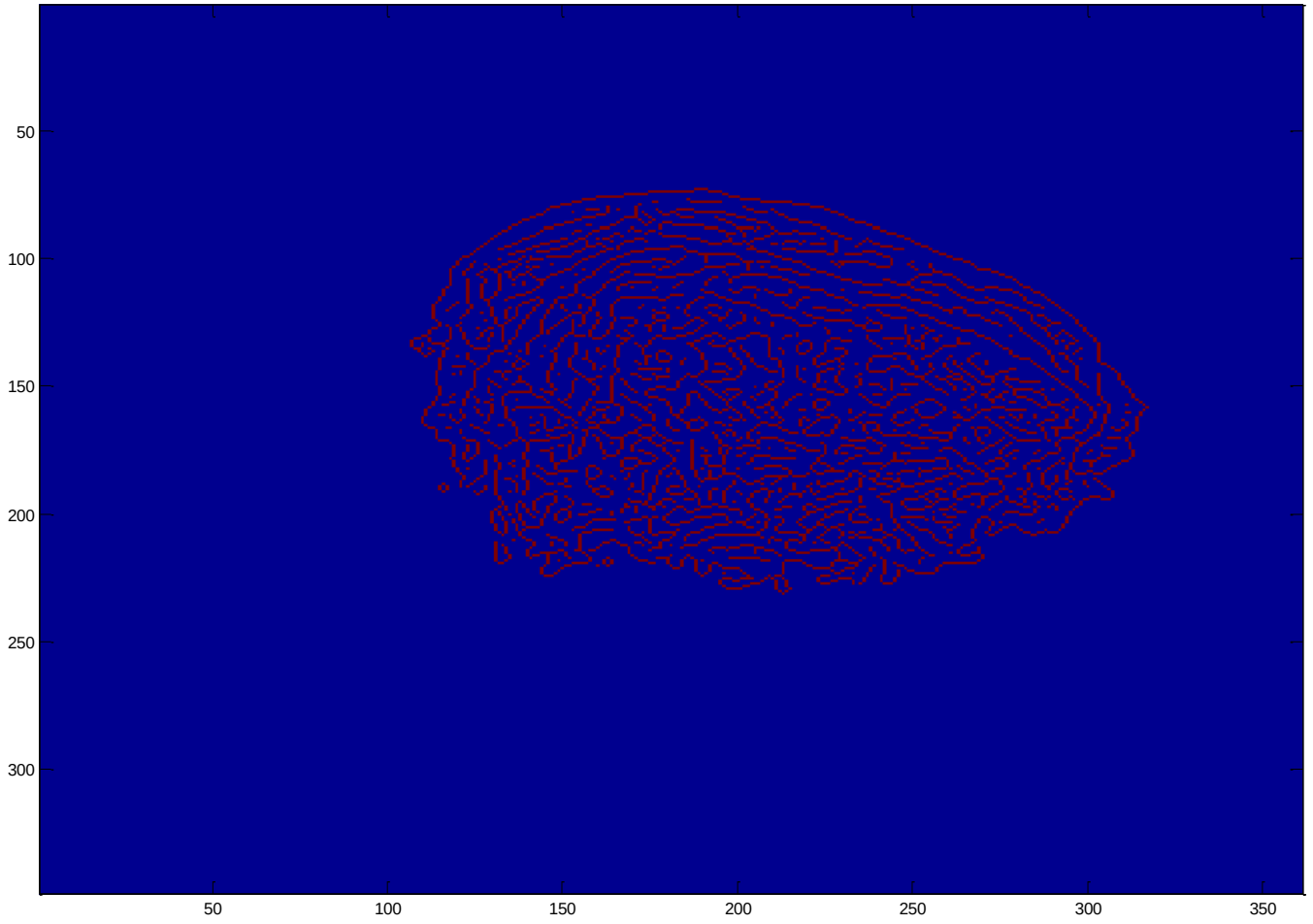
1, 2, 4



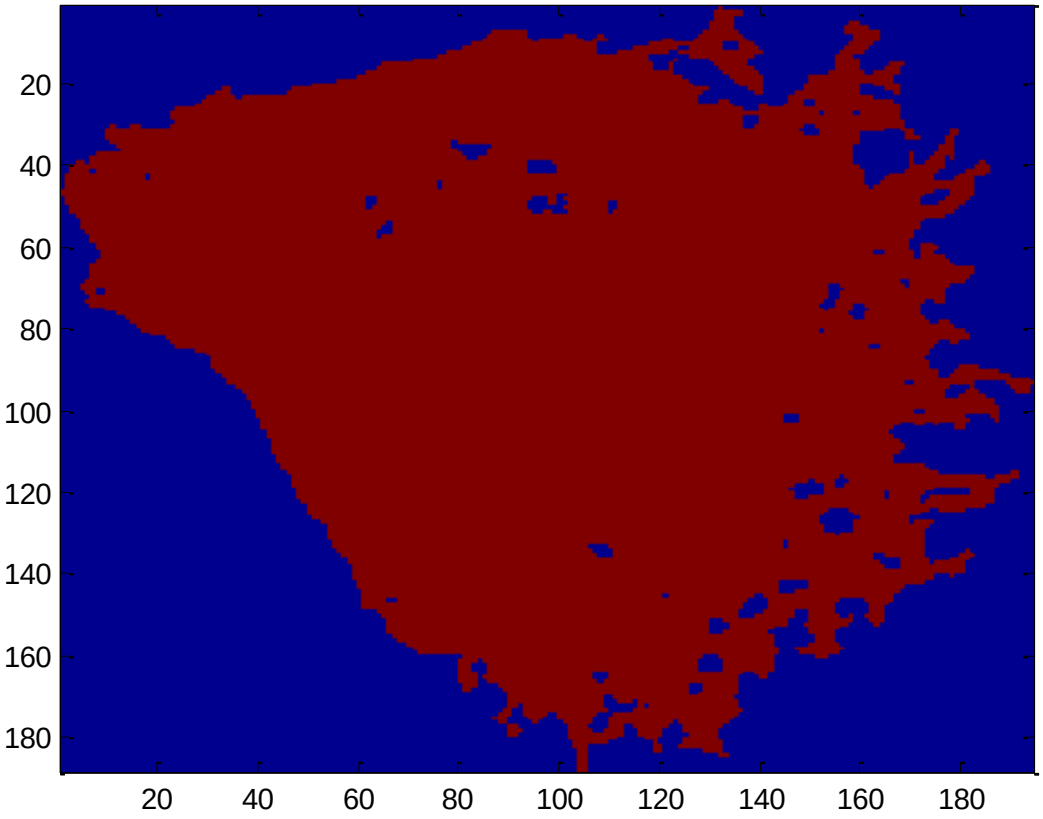


1, 3, 6

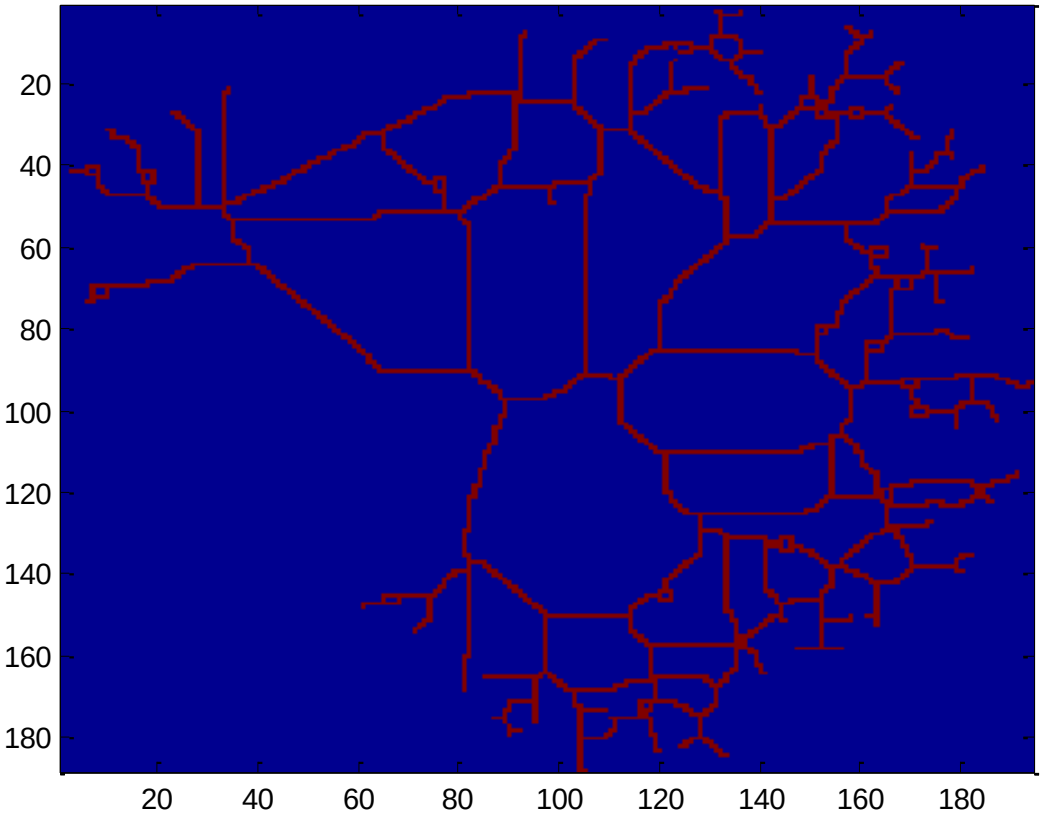
e = edge(imageproc, 'canny', []); figure, imagesc(e);



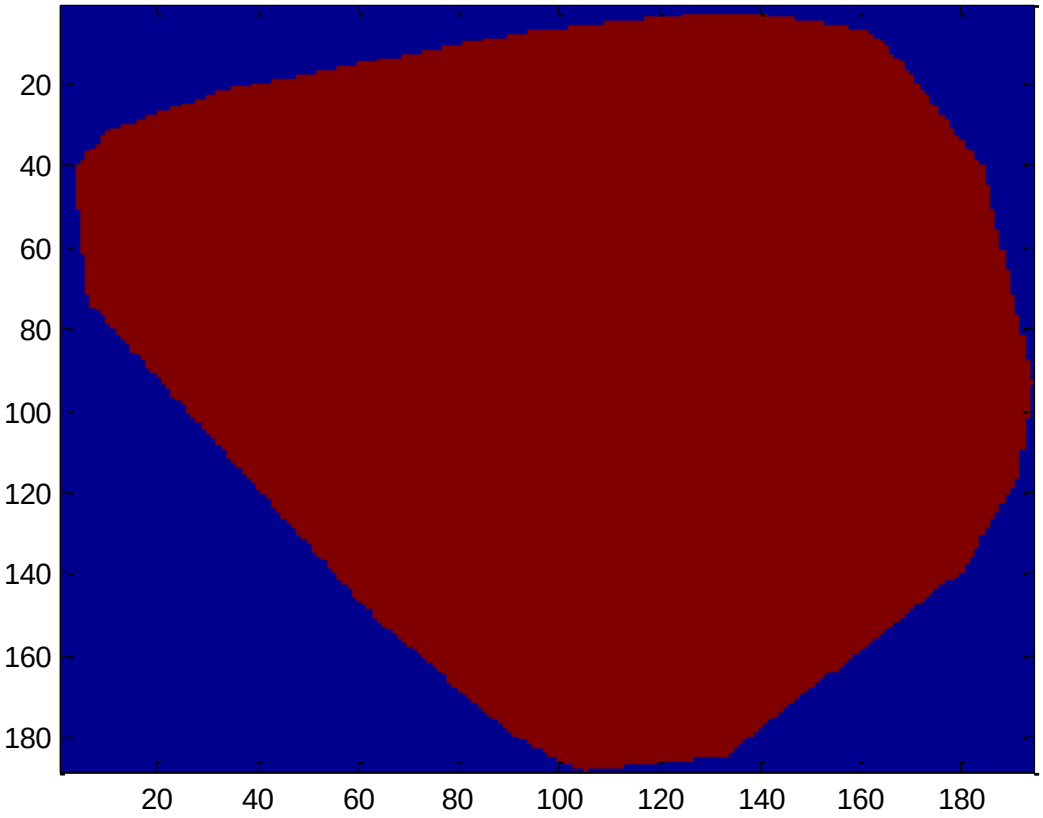
segmentation



2546



Convex hul



Mk7 threshold figure(1), imagesc(imageproc)