# Supplementary Materials

# Quantitative diagnosis of breast tumors by morphometric classification of microenvironmental myoepithelial cells using a machine learning approach

Yoichiro Yamamoto, § [1,2,3,4] Akira Saito, § [4,5,6] Ayako Tateishi,[1] Hisashi Shimojo,[1] Hiroyuki Kanno,[1] Shinichi Tsuchiya,[7] Ken-ichi Ito,[8] Eric Cosatto,[9] Hans Peter Graf, [9] Rodrigo R. Moraleda,[10,11] Roland Eils [2,3] and Niels Grabe [12,13]

§ Both authors contributed equally to this work

[1] Department of Pathology, Shinshu University School of Medicine, Nagano, Japan.

[2] Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[3] Department for Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology (IPMB) and Bioquant, University of Heidelberg, Heidelberg, Germany.

[4] RIKEN Center for Advanced Intelligence Project, Pathology Informatics Unit, Tokyo, Japan.

[5] Quantitative Pathology & Immunology, Tokyo Medical University, Shinjuku, Tokyo, Japan.

[6] Department of Molecular Pathology, Tokyo Medical University, Shinjuku, Tokyo, Japan.

[7] Diagnostic Pathology, Ritsuzankai Iida Hospital, Nagano, Japan.

[8] Division of Breast and Endocrine Surgery, Shinshu University School of Medicine, Nagano, Japan.

[9] Department of Machine Learning, NEC Laboratories America, NJ, USA.

[10] Applied Tumor Immunity Clinical Cooperation Unit, National Center for Tumor Diseases, German Cancer Research Center, Heidelberg, Germany.

[11] Department of informatics, Technical University Federico Santa Maria Valparaiso, Chile.

[12] Department of Medical Oncology, National Center for Tumor Diseases, University of Heidelberg, Heidelberg, Germany.

[13] Hamamatsu Tissue Imaging and Analysis Center, Bioquant, University of Heidelberg, Heidelberg, Germany.

**Table S1. Characteristics of objects.**

**Total number of nuclei: 11661**

| | Cells (p63) | Cells (HE) | ROIs (p63) | ROIs (HE) | Cases | Age |
|---|---|---|---|---|---|---|
| Normal | 2749 | 1289 | 14 | 14 | 7 | 65.4±13.9 |
| UDH | 2307 | 1073 | 16 | 16 | 5 | 56.6±7.6 |
| LG-DCIS | 1819 | 666 | 21 | 21 | 5 | 65.0±15.6 |
| HG-DCIS | 1203 | 555 | 19 | 19 | 5 | 63.8±11.6 |

(Mean±S.D.)

UDH: Usual ductal hyperplasia

LG-DCIS: Low Grade Ductal carcinoma in situ

HG-DCIS: High Grade Ductal carcinoma in situ

**Table S2. Analyzed ROIs**

| | Total Cases | ROIs | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Case1 | Case2 | Case3 | Case4 | Case5 | Case6 | Case7 |
| Normal | 7 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| UDH | 5 | 4 | 2 | 3 | 3 | 4 | | |
| DCIS G1 | 5 | 4 | 7 | 4 | 4 | 2 | | |
| DCIS G3 | 5 | 4 | 3 | 5 | 3 | 4 | | |

## Table S3. Morphological features

**Area & Shape**

| | | | | |
|---|---|---|---|---|
| Area | Compactness | Eccentricity | Form Factor | Major Axis Length |
| Maximum Feret Diameter | Maximum Radius | Mean Radius | Median Radius | Minimum Feret Diameter |
| Minor Axis Length | Perimeter | Solidity | Ratio of Feret Diameter | Ratio of Axis Length |
| FracAtD 1of4* | FracAtD 2of4* | FracAtD 3of4* | FracAtD 4of4* | |

**Nuclear texture (GLCM)**

| | | | | |
|---|---|---|---|---|
| Angular Second Moment | Contrast | Correlation | Difference Entropy | Difference Variance |
| Entropy | Info. Measure of Correlation1 | Info. Measure of Correlation2 | Inverse Difference Moment | Sum Average |
| Sum Entropy | Sum Variance | Variance | | |

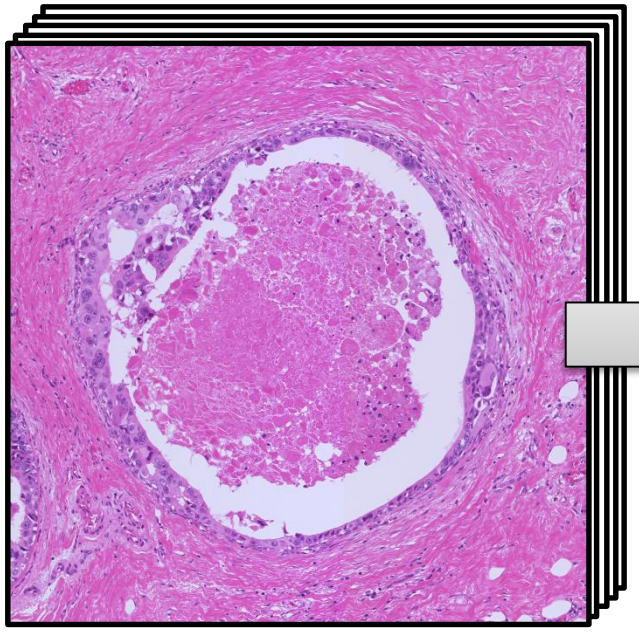
* FracAtD: Fraction of total stain in an object at a given radius. The distribution is measured from the center of the object, where the center is defined as the point farthest from any edge. The numbering is from 1 (innermost) to 4 (outermost).

**Table S4. Possible candidate molecules of crosstalk between luminal cells and myoepithelial cells.**

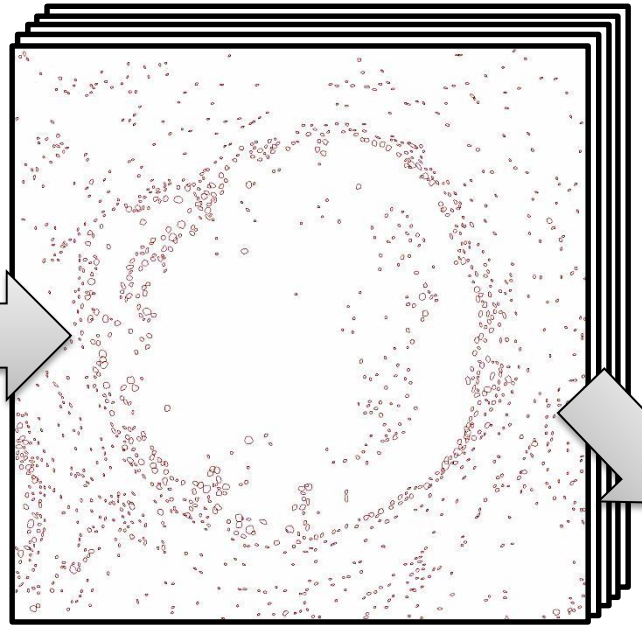
| Ligand (Luminal cell) | Receptor (Myoepithelial cell) |
|---|---|
| ADAM15, ADAM17, COL18A1 | ITGB1 |
| CDH1 | ITGB7 |
| DUSP18 | ITGA3 |
| DUSP18 | ITGA7 |
| FN1 | SDC2 |
| IL15 | IL15RA |
| **SLIT2** | **ROBO1** |
| **SHH** | **SMO** |
| WNT7B | FZD1 |

# Figure S1

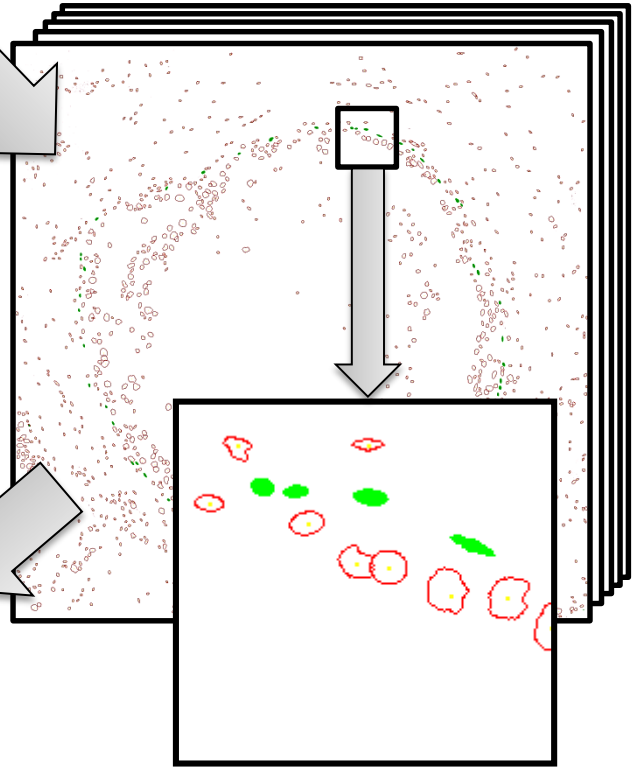
1.Raw data (Serial sections of p63 IHC slides)



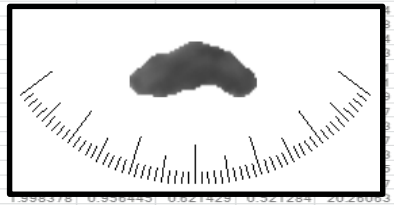
2.Segmentation of all cells by program of NEC



3. Manually picking up myoepithelial cells

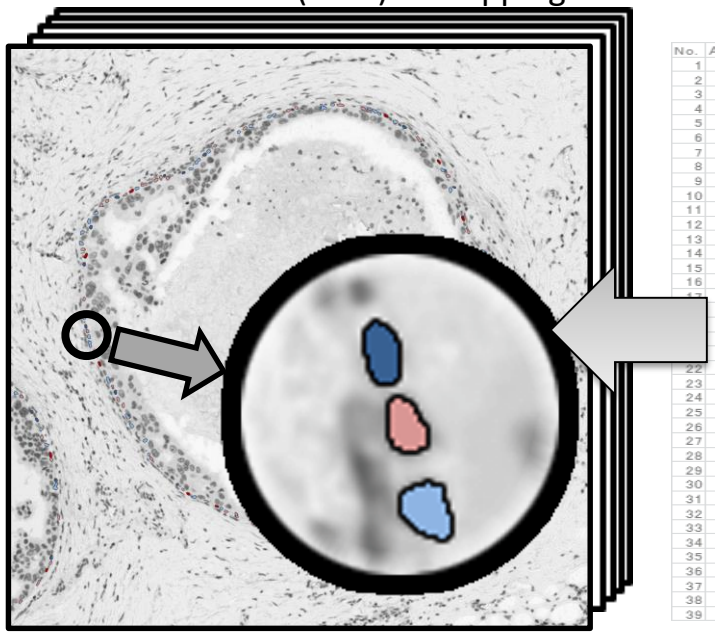


4.Measurement by CellProfiler

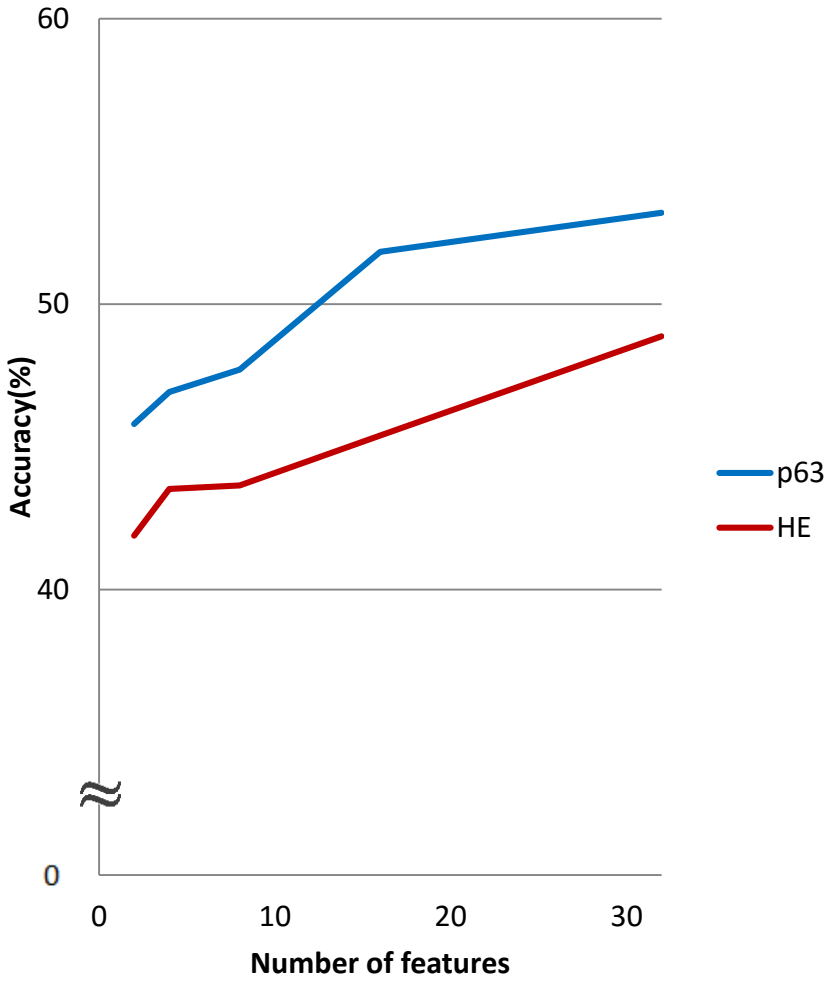


| No. | Area | Compactne | Eccentricit | Extent | FormFacto | MajorAxisL | MaxFeretD |
|---|---|---|---|---|---|---|---|
| 1 | 153 | 1.774397 | 0.948938 | 0.68 | 0.537721 | 25.12586 | 24.18677 |
| 2 | 156 | 1.264323 | 0.86517 | 0.619048 | 0.738498 | 20.09701 | 19.72308 |
| 3 | 134 | 1.708976 | 0.911633 | 0.553719 | 0.389348 | 22.39436 | 21.0238 |
| 4 | 147 | 1.648768 | 0.932858 | 0.480392 | 0.566413 | 23.42876 | 23.34524 |
| 5 | 126 | 1.441525 | 0.913324 | 0.7875 | 0.660322 | 19.98287 | 19.41649 |
| 6 | 87 | 1.542556 | 0.914052 | 0.639706 | 0.59216 | 17.20261 | 16.12452 |
| 7 | 85 | 1.681075 | 0.90786 | 0.664063 | 0.56573 | 17.66234 | 15.52417 |
| 8 | 142 | 1.623574 | 0.91986 | 0.44375 | 0.379272 | 22.61481 | 21.47091 |
| 9 | 86 | 1.661746 | 0.933837 | 0.632353 | 0.55903 | 18.03145 | 16.12452 |
| 10 | 94 | 2.959781 | 0.980617 | 0.29375 | 0.404999 | 26.16462 | 23.60085 |
| 11 | 107 | 2.002178 | 0.960625 | 0.607955 | 0.453488 | 22.56063 | 21.37756 |
| 12 | 134 | 2.143986 | 0.966961 | 0.406061 | 0.487632 | 26.256 | 24.69818 |
| 13 | 129 | | | | | | 24.16605 |
| 14 | 129 | | | | | | 21.37756 |
| 15 | 227 | | | | | | 24.69818 |
| 16 | 81 | | | | | | 22.47221 |
| 17 | 89 | | | | | | 17.02939 |
| 18 | 128 | | | | | | 23.85372 |
| 19 | 119 | | | | | | 22.36068 |
| 20 | 162 | | | | | | 17.80449 |
| 21 | 87 | | | | | | 17.26268 |
| 22 | 152 | | | | | | 17.80449 |
| 23 | 102 | | | | | | 18.78829 |
| 24 | 132 | | | | | | 16.12452 |
| 25 | 138 | | | | | | 18.60108 |
| 26 | 87 | | | | | | 19.23538 |
| 27 | 90 | 2.050249 | 0.962401 | 0.592105 | 0.553329 | 20.97888 | 18.68154 |
| 28 | 107 | 1.619323 | 0.936211 | 0.509524 | 0.627779 | 19.88177 | 19.10497 |
| 29 | 128 | 1.47595 | 0.904151 | 0.547009 | 0.625903 | 20.23429 | 18.86796 |
| 30 | 110 | 1.674226 | 0.918816 | 0.482456 | 0.505745 | 20.20779 | 18.24829 |
| 31 | 179 | 1.059069 | 0.614502 | 0.699219 | 0.775235 | 17.31132 | 17.20465 |
| 32 | 132 | 2.346707 | 0.971567 | 0.347368 | 0.44822 | 27.37303 | 24.83948 |
| 33 | 86 | 1.187822 | 0.816857 | 0.735043 | 0.766295 | 14.04825 | 13.41641 |
| 34 | 86 | 1.132375 | 0.730476 | 0.735043 | 0.763446 | 13.08082 | 12.16553 |
| 35 | 159 | 1.178992 | 0.827222 | 0.623529 | 0.801271 | 19.11294 | 17.80449 |
| 36 | 94 | 1.248771 | 0.851312 | 0.696296 | 0.745863 | 15.38524 | 14.86607 |
| 37 | 156 | 1.264323 | 0.86517 | 0.619048 | 0.738498 | 20.09701 | 19.72308 |
| 38 | 134 | 1.708976 | 0.911633 | 0.553719 | 0.389348 | 22.39436 | 21.0238 |
| 39 | 147 | 1.648768 | 0.932858 | 0.480392 | 0.566413 | 23.42876 | 23.34524 |

5. Classification (SVM) & Mapping



# Figure S2



# Figure S3

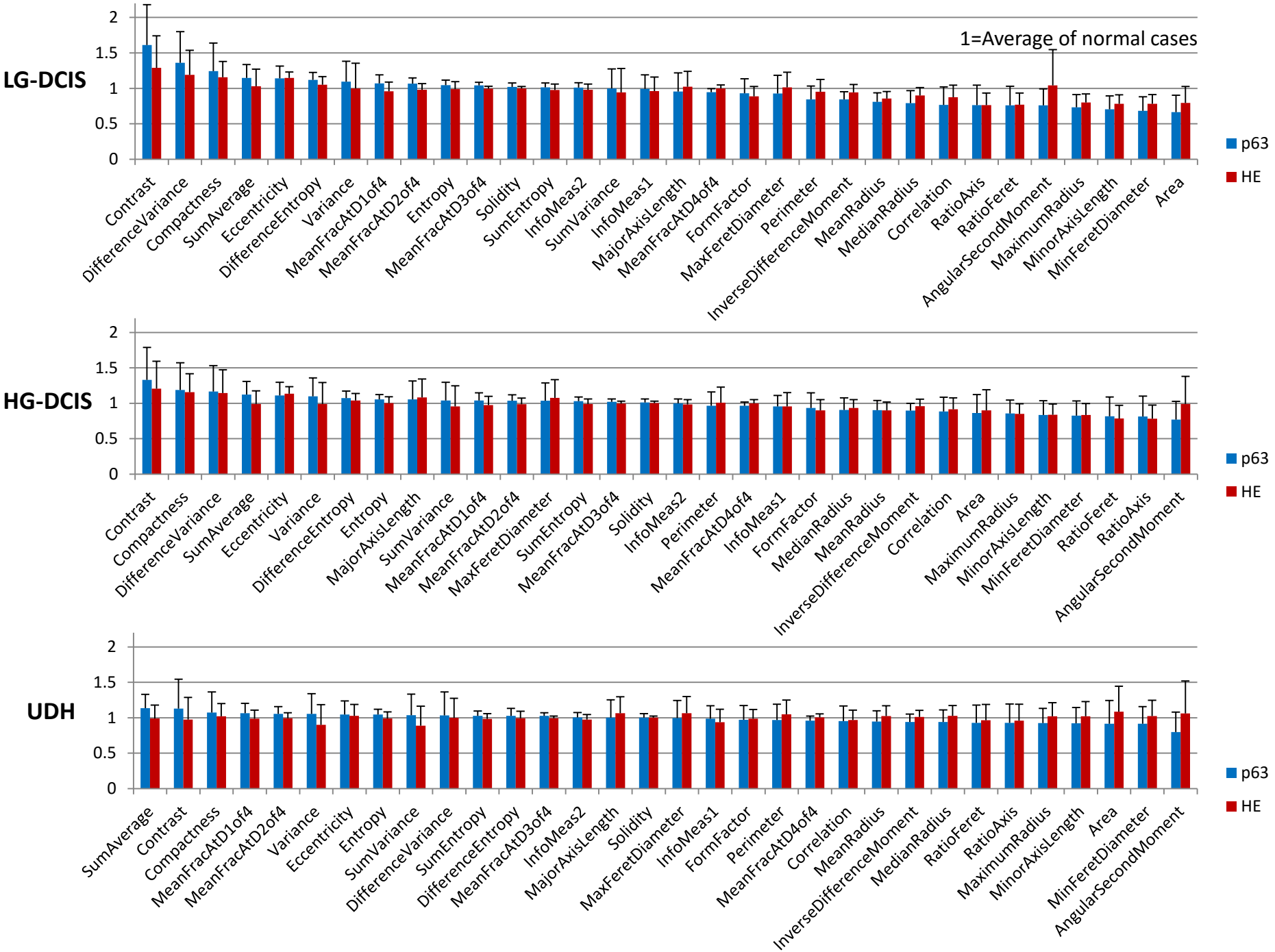
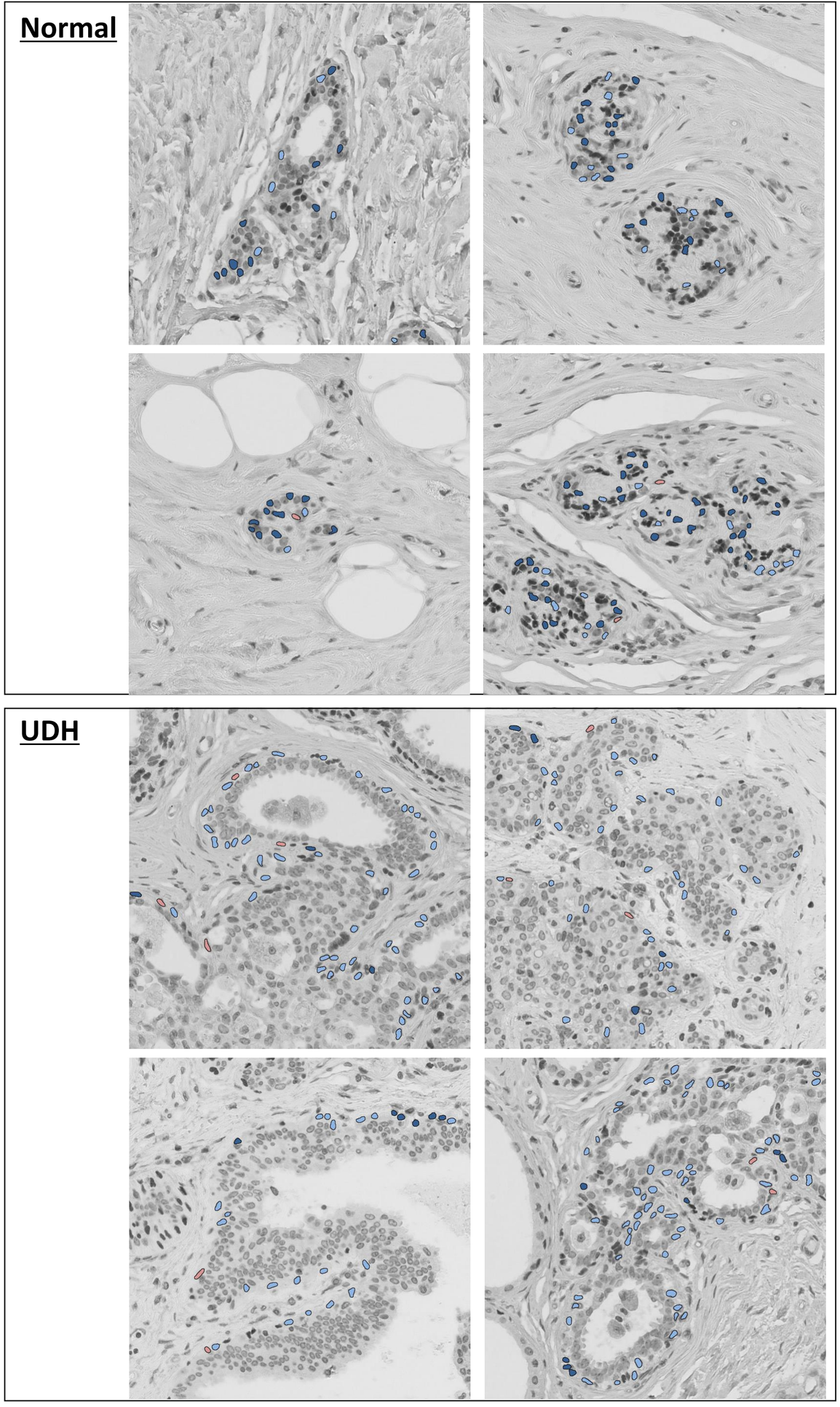
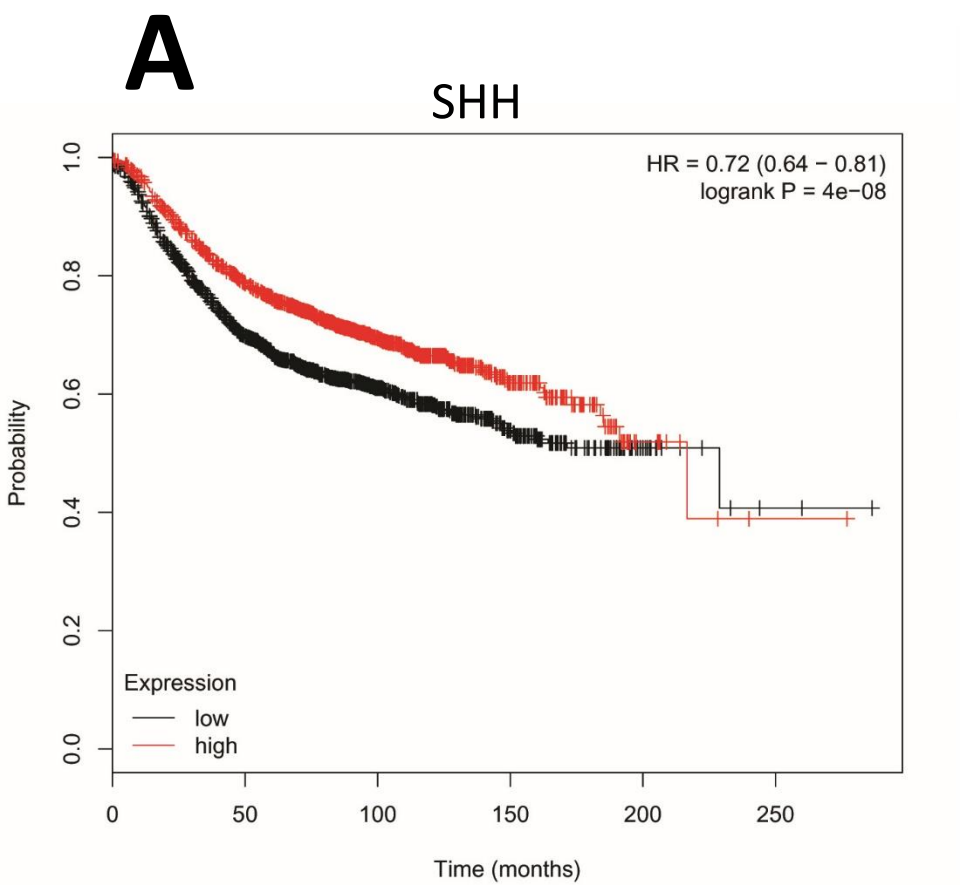
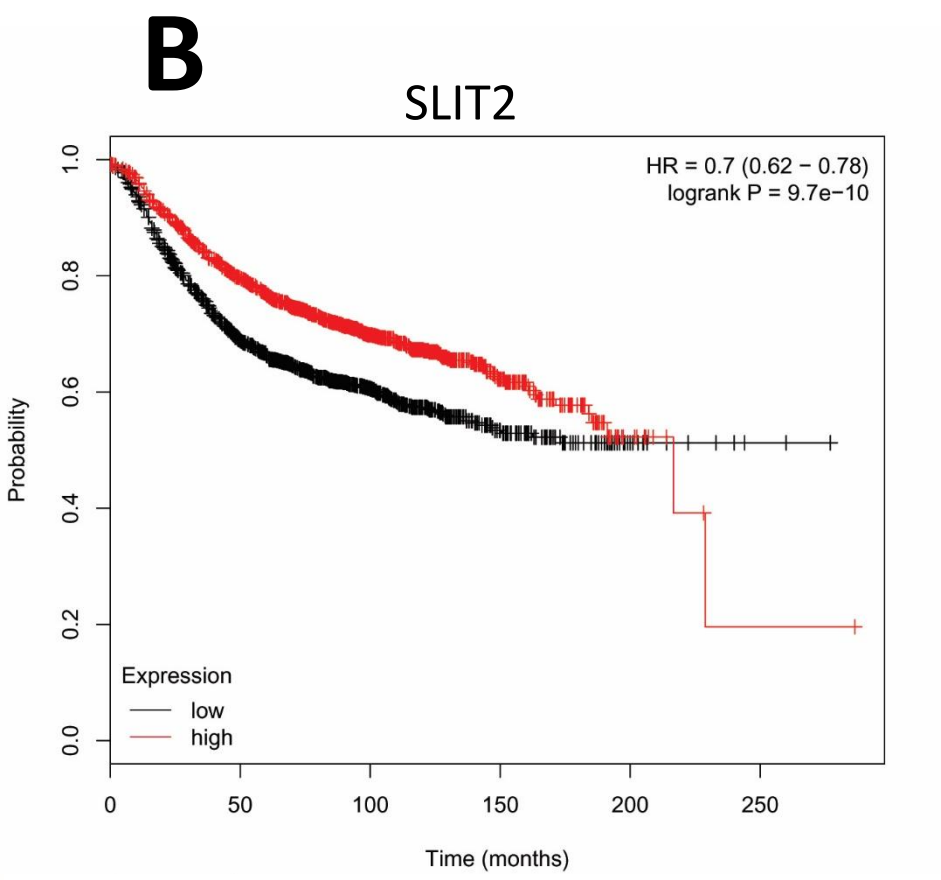

LG-DCIS
1=Average of normal cases
2
1.5
1
0.5
0
Contrast
DifferenceVariance
Compactness
SumAverage
Eccentricity
DifferenceEntropy
Variance
MeanFracAtD1of4
MeanFracAtD2of4
Entropy
MeanFracAtD3of4
Solidity
SumEntropy
InfoMeas2
SumVariance
InfoMeas1
MajorAxisLength
MeanFracAtD4of4
FormFactor
MaxFeretDiameter
Perimeter
InverseDifferenceMoment
MeanRadius
MedianRadius
Correlation
RatioAxis
RatioFeret
AngularSecondMoment
MaximumRadius
MinorAxisLength
MinFeretDiameter
Area
p63
HE
HG-DCIS
2
1.5
1
0.5
0
Contrast
Compactness
DifferenceVariance
SumAverage
Eccentricity
Variance
DifferenceEntropy
Entropy
MajorAxisLength
SumVariance
MeanFracAtD1of4
MeanFracAtD2of4
MaxFeretDiameter
SumEntropy
MeanFracAtD3of4
Solidity
InfoMeas2
Perimeter
MeanFracAtD4of4
InfoMeas1
FormFactor
MedianRadius
MeanRadius
InverseDifferenceMoment
Correlation
Area
MaximumRadius
MinorAxisLength
MinFeretDiameter
RatioFeret
RatioAxis
AngularSecondMoment
p63
HE
UDH
2
1.5
1
0.5
0
SumAverage
Contrast
Compactness
MeanFracAtD1of4
MeanFracAtD2of4
Variance
Eccentricity
Entropy
SumVariance
DifferenceVariance
SumEntropy
DifferenceEntropy
MeanFracAtD3of4
InfoMeas2
MajorAxisLength
Solidity
MaxFeretDiameter
InfoMeas1
FormFactor
Perimeter
MeanFracAtD4of4
Correlation
MeanRadius
InverseDifferenceMoment
MedianRadius
RatioFeret
RatioAxis
MaximumRadius
MinorAxisLength
Area
MinFeretDiameter
AngularSecondMoment
p63
HE

**Figure S4**



# Figure S5


A
SHH
HR = 0.72 (0.64 − 0.81)
logrank P = 4e−08
Probability
1.0
0.8
0.6
0.4
0.2
0.0
Expression
low
high
0
50
100
150
200
250
Time (months)


B
SLIT2
HR = 0.7 (0.62 − 0.78)
logrank P = 9.7e−10
Probability
1.0
0.8
0.6
0.4
0.2
0.0
Expression
low
high
0
50
100
150
200
250
Time (months)