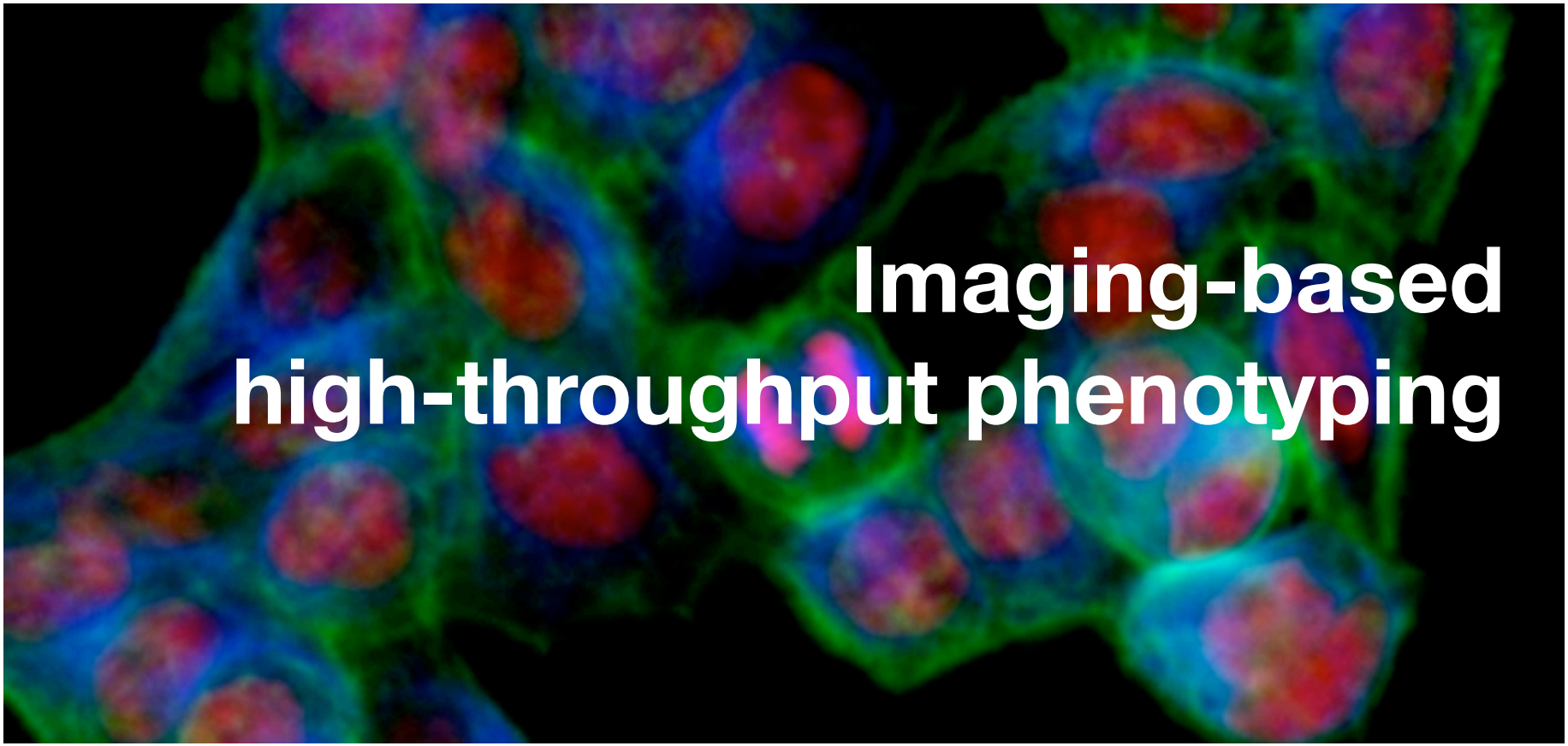


Imaging-based high-throughput phenotyping

Wolfgang Huber

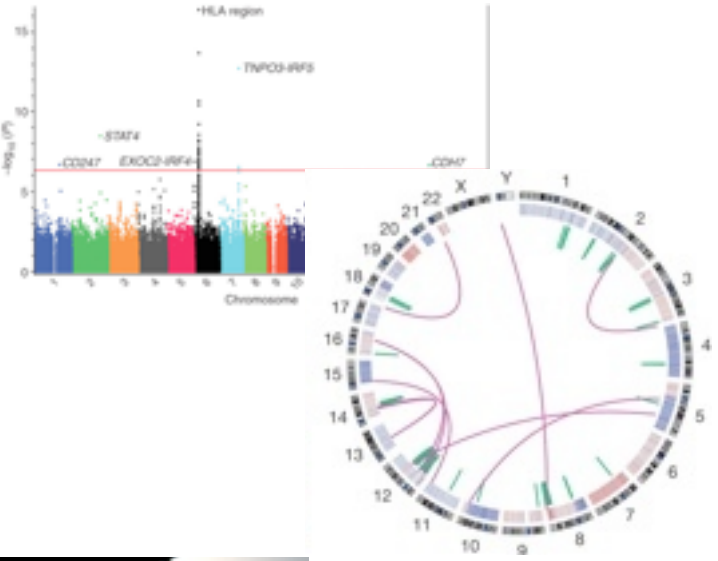
EMBL

How do we know which genes do what?

Forward genetics

from phenotypes to genes

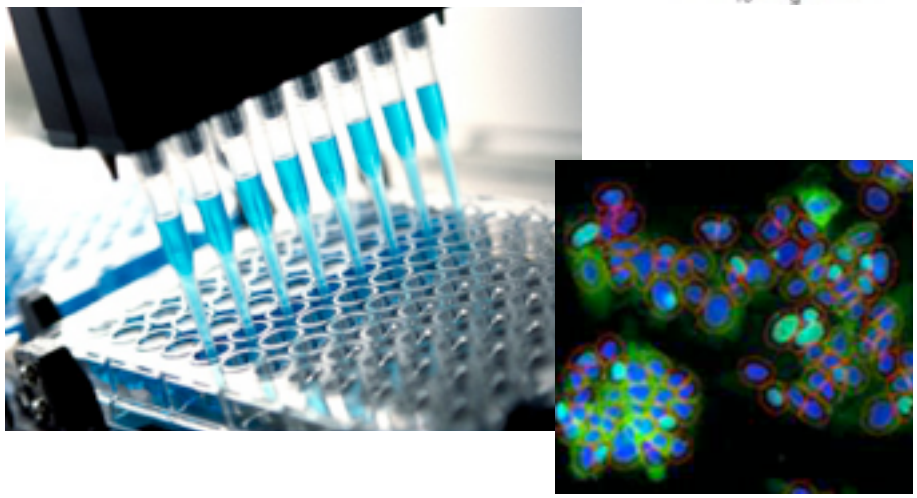
- → **genome-wide association studies**
- → **sporadic/rare mutations**
- → **cancer genome sequencing**



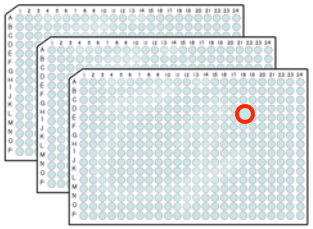
Reverse genetics

from genes to phenotypes

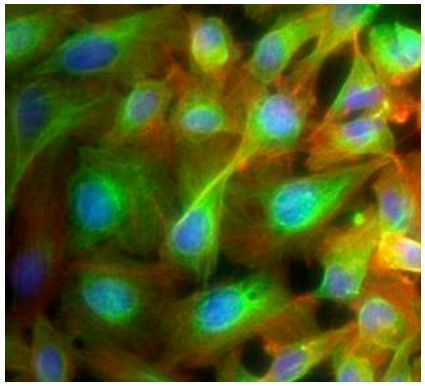
- → **deletion libraries**
- → **high-throughput RNAi**



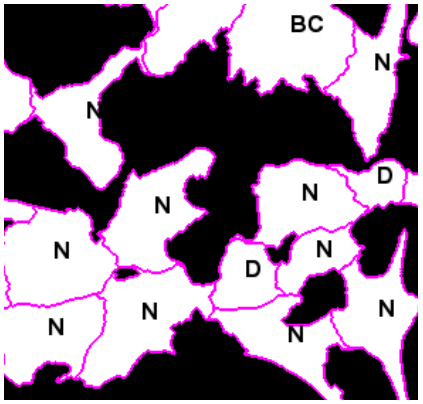
High-throughput RNAi and automated cellular phenotyping



RNAi or drug library



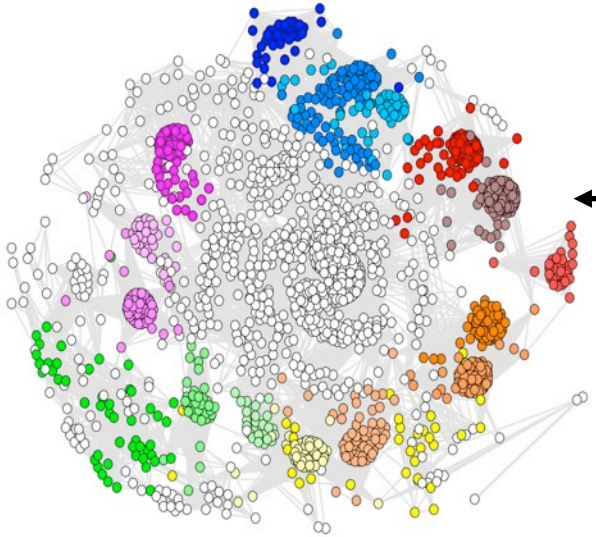
Segmentation



Feature extraction

	g.x	g.y	g.s	g.p	g.pdm
[1,]	123.1391	3.288660	194	67	9.241719
[2,]	206.7460	9.442248	961	153	20.513190
[3,]	502.9589	7.616438	219	60	8.286918
[4,]	20.1919	22.358418	1568	157	22.219461
[5,]	344.7959	45.501992	2259	233	35.158966
[6,]	188.2611	50.451863	2711	249	28.732680
[7,]	269.7996	46.404036	2131	180	26.419631
[8,]	106.6127	58.364243	1348	143	21.662879
[9,]	218.5582	77.299007	1913	215	25.724580
[10,]	19.1766	81.840147	1908	209	26.303760
[11,]	6.3558	62.017647	340	68	10.314127
[12,]	58.9873	86.034128	2139	214	27.463158
[13,]	245.1087	94.387405	1048	123	18.280901
[14,]	411.2741	109.198678	2572	225	28.660816
[15,]	167.8151	107.966014	1942	160	24.671533
[16,]	281.7084	121.609892	2871	209	31.577270

Quantitative cell and organelle features



multivariate phenotypic landscape

Boutros, Bras, Huber, **Genome Biol.** 2006
Fuchs, Pau et al. **Mol. Sys. Biol.** 2010
Pau, Fuchs et al. **Bioinf.** 2010
Neumann et al. **Nature** 2010
Kuttenkeuler et al. **J. Innate Imm.** 2010
Axelsson et al. **BMC Bioinf.** 2011
Horn et al. **Nature Methods** 2011

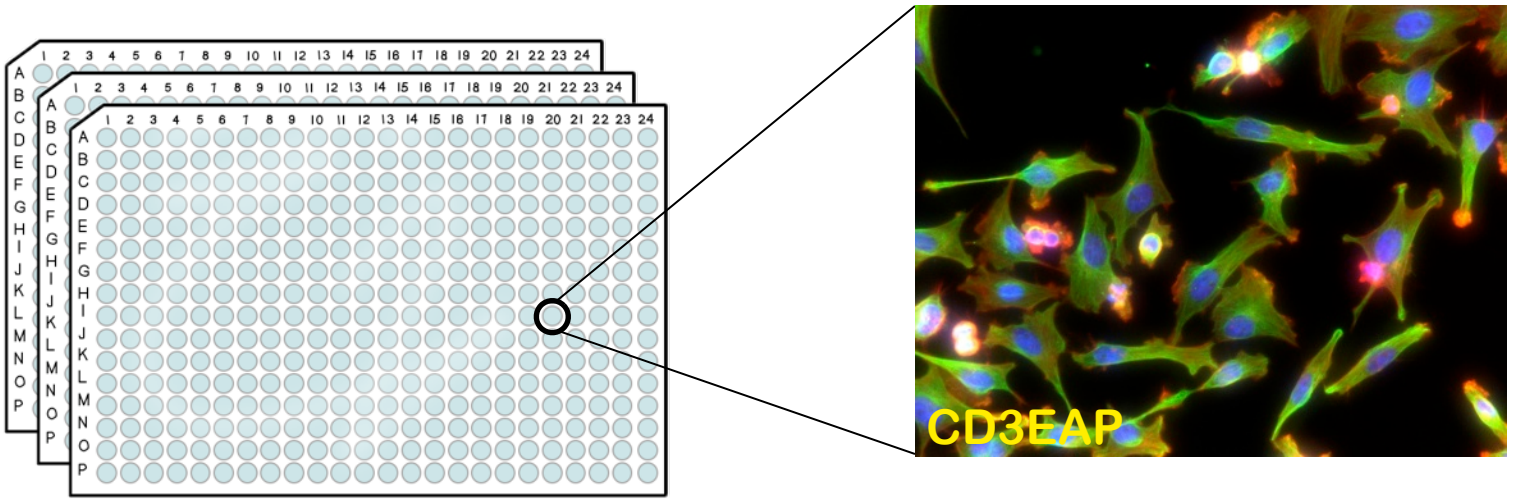
An example

with G.Pau; F. Fuchs, C. Budjan, Michael Boutros (DKFZ)

Genome-wide RNAi library (Dharmacon, 22k siRNA-pools)

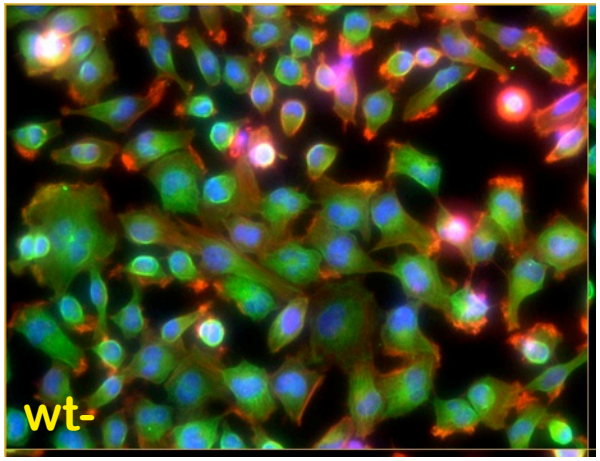
HeLa cells, incubated 48h, then fixed and stained

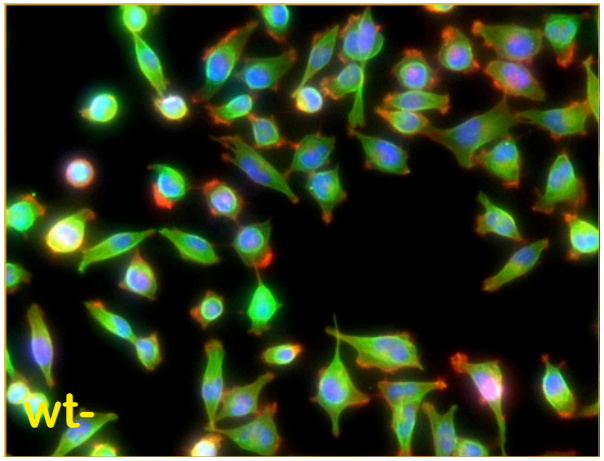
Microscopy readout: DNA (DAPI), tubulin (Alexa), actin (TRITC)

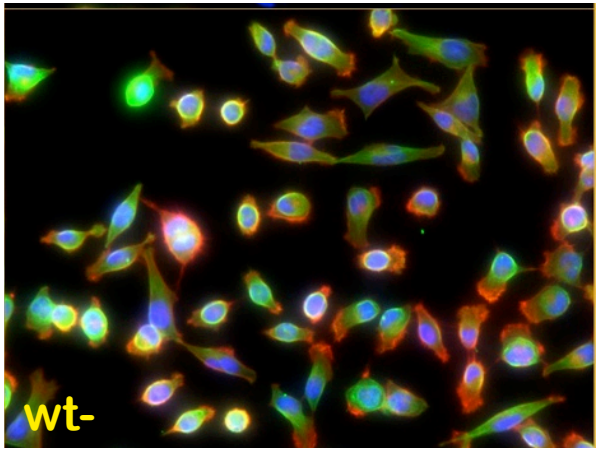


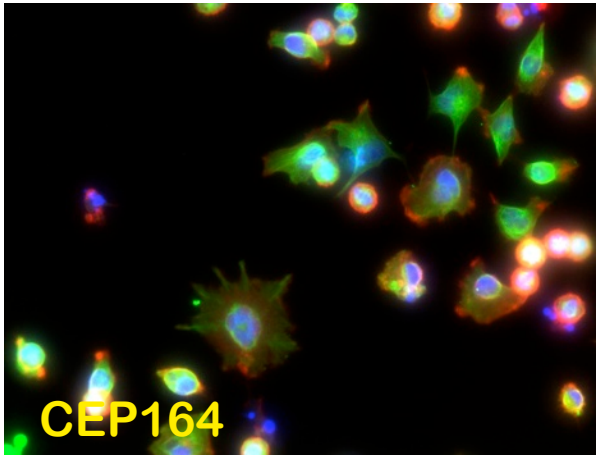
Fuchs, Pau, et al. Molecular Systems Biology (2010)

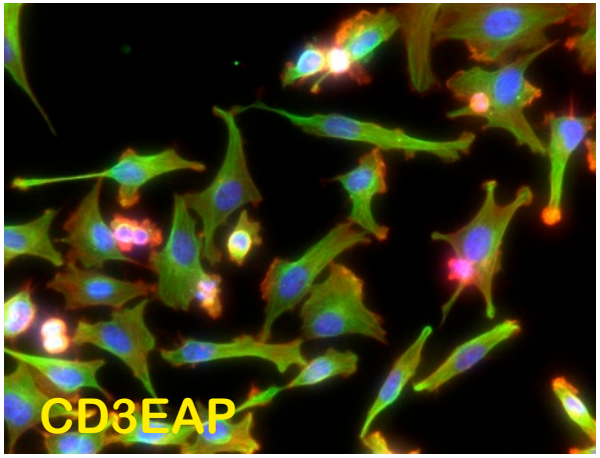
siRNA perturbation phenotypes are observed by automated microscopy

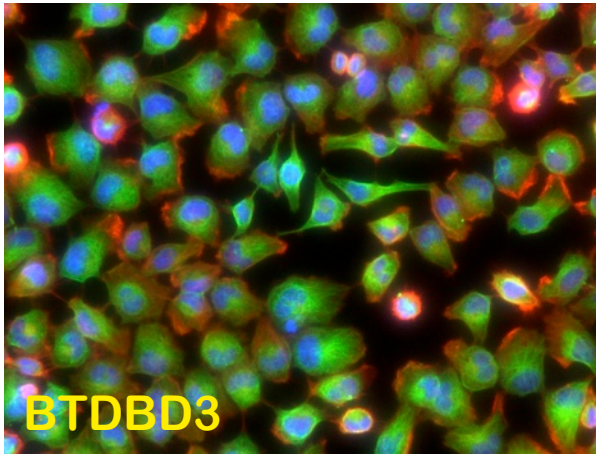








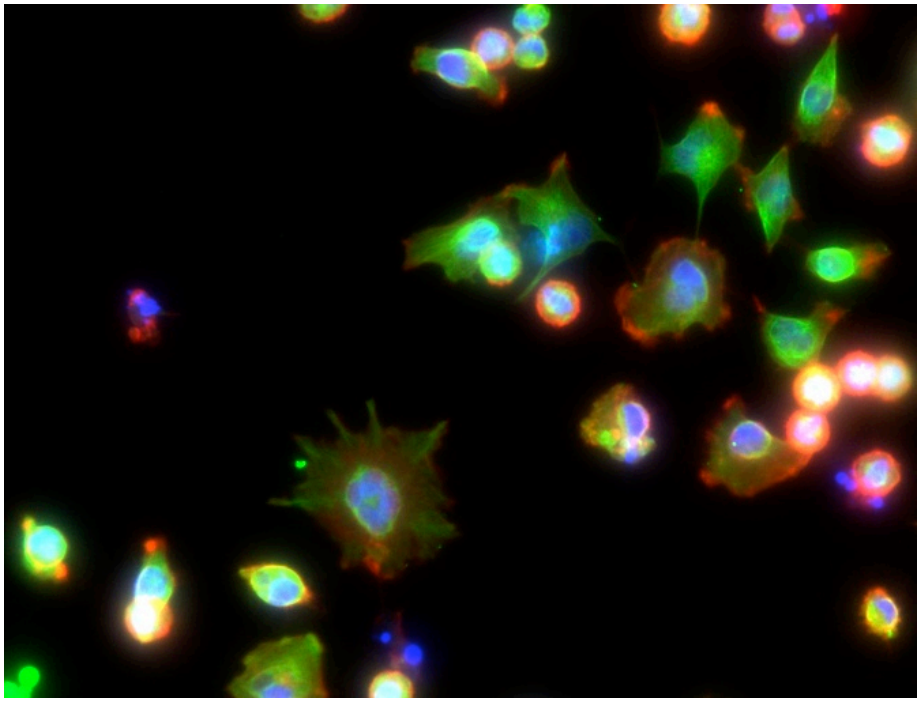


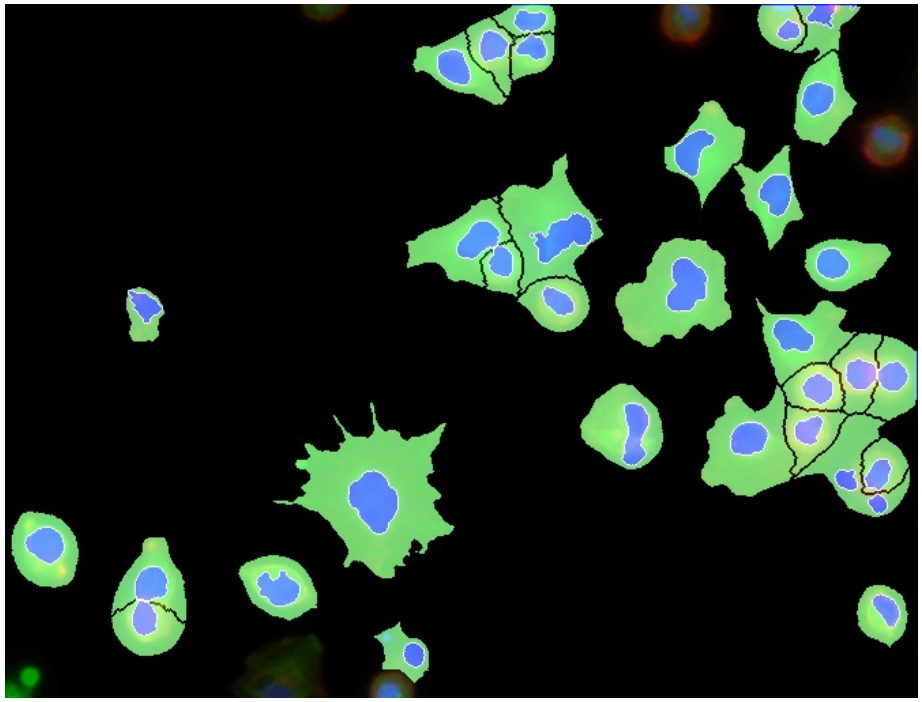


22839 wells DNA, tubulin, actin

4 images per well, each with 3 colours, 1344 x 1024 pixel at 12 bit

Segmentation



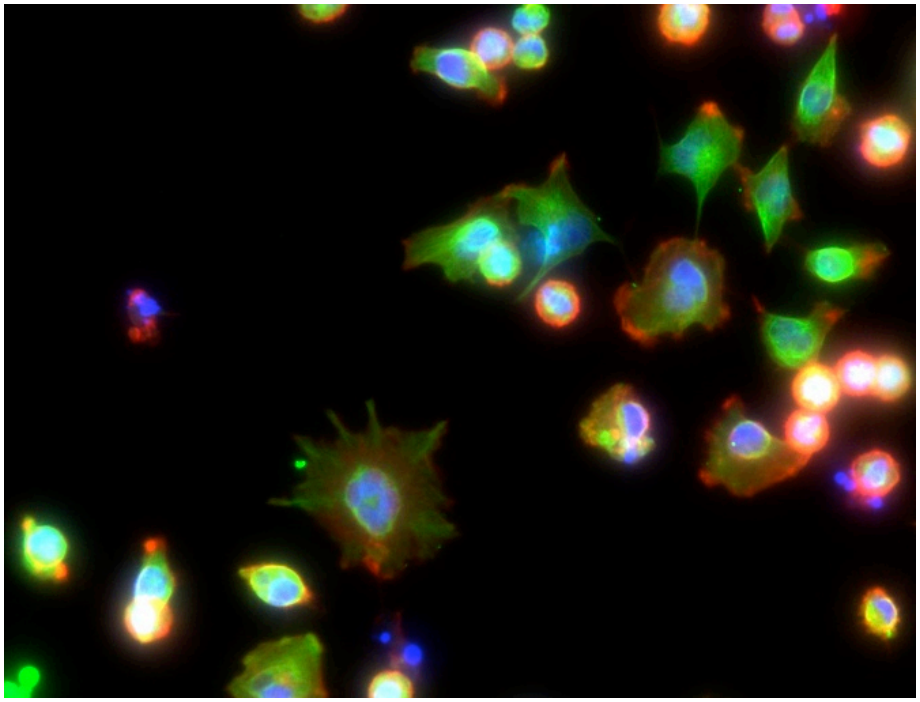


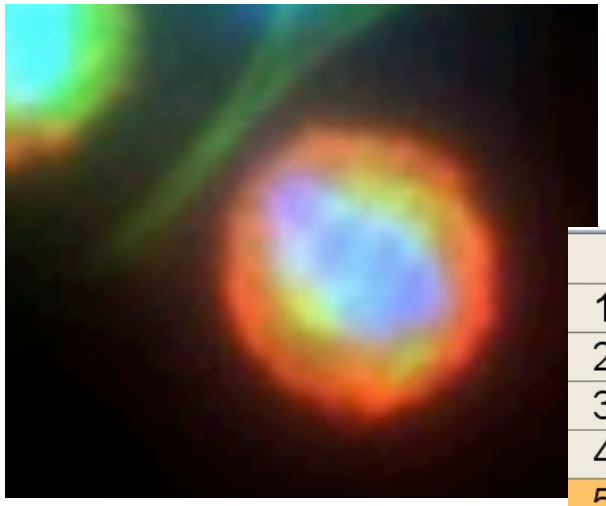
CellProfiler (GUI)
EBImage R package

Extraction of quantitative cell descriptors

- geometry (intensity, size, perimeter, eccentricity…)
- texture (Haralick, Zernike moments…) on each channel
- relative positions/densities

Translation and rotation invariant (?)





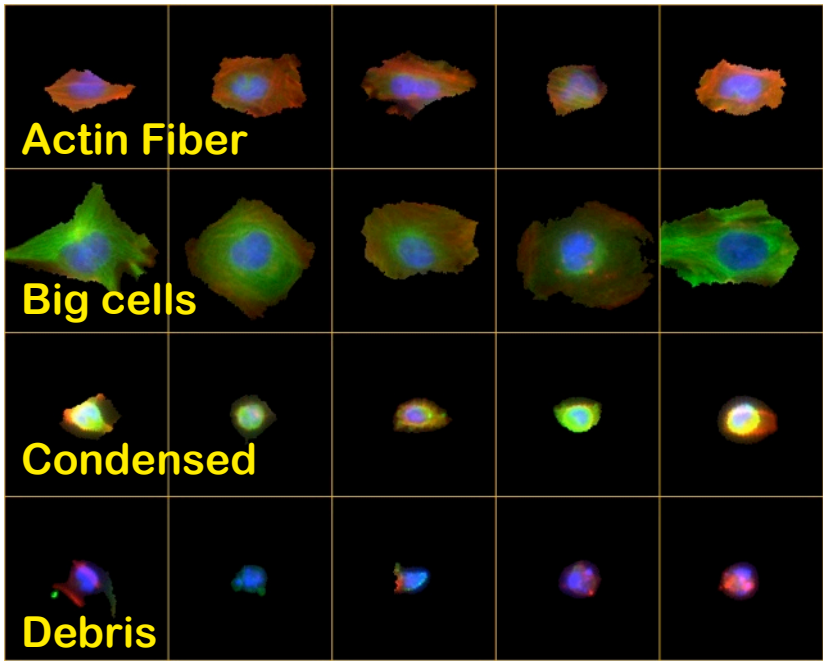
	A
1	202.12
2	11.31
3	2.22
4	4.01
5	3.14
6	15.7
7	-0.911

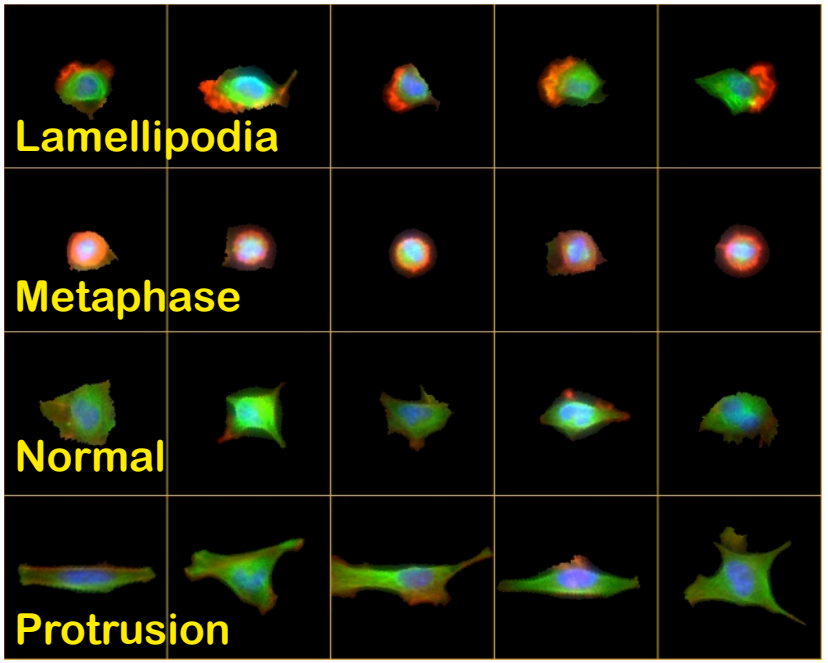
Classification, Tagging: categorial ‘features’

based on the numeric descriptors

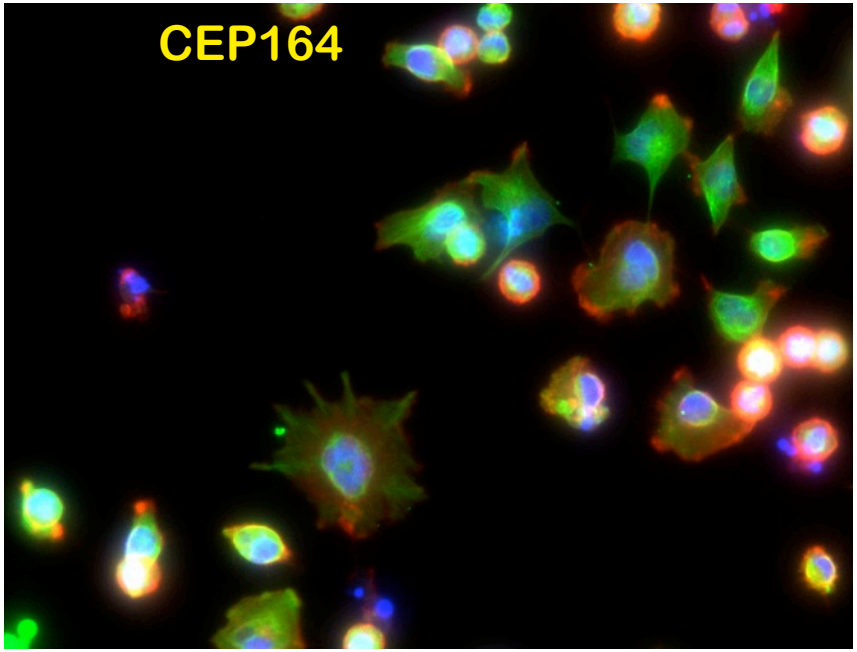
supervised learning

can be a way to reduce noise / focus on biological signal



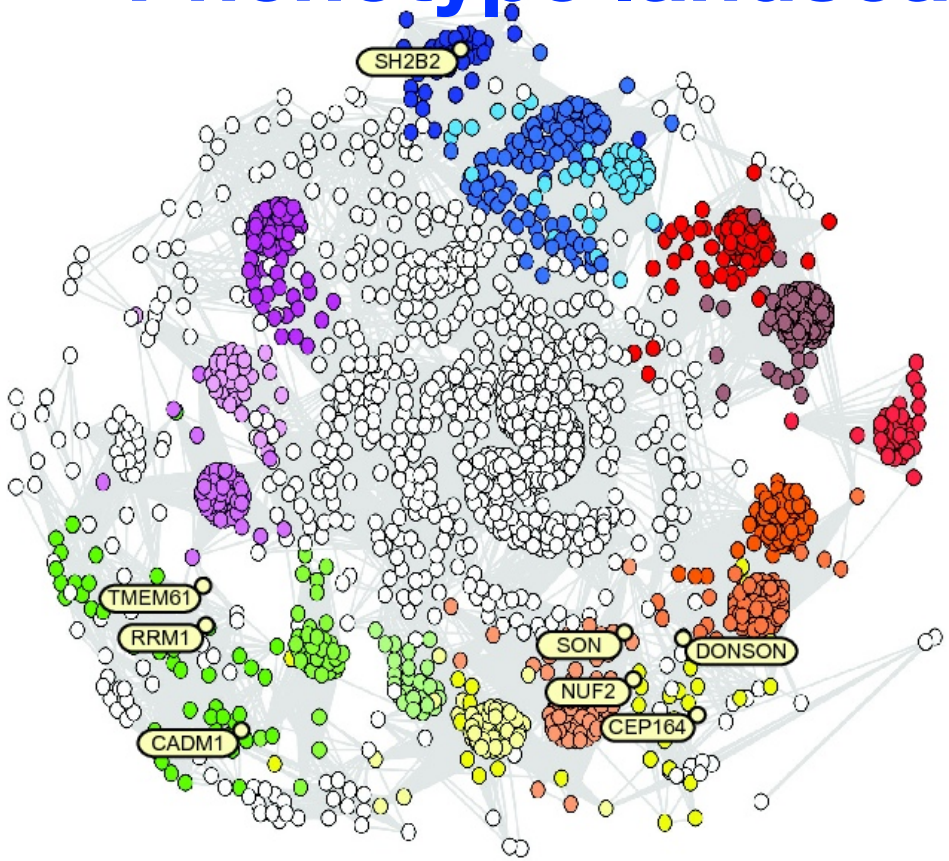


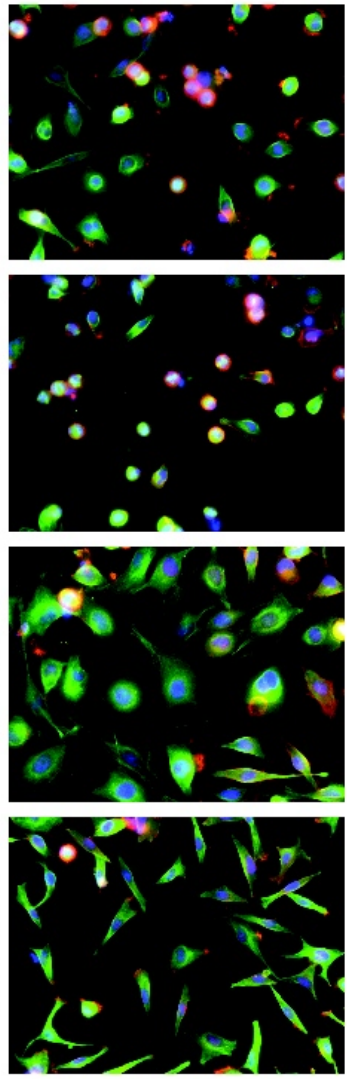
Per-cell vs per-well (population) features



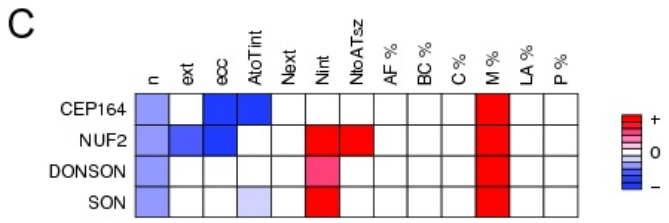
number of cells	128
average intensity	1054.8
average nuclear intensity	1225.6
average cell size	842.3
average nuclear size	278.7
average eccentricy	0.649
avg. nuclear / cell size	2.91
# AF (actin fibers)	2
# BC (big)	7
# M (mitotic)	15
# LA (lamellipodia)	0
# P (with protrusions)	17
# Z (telophase)	2

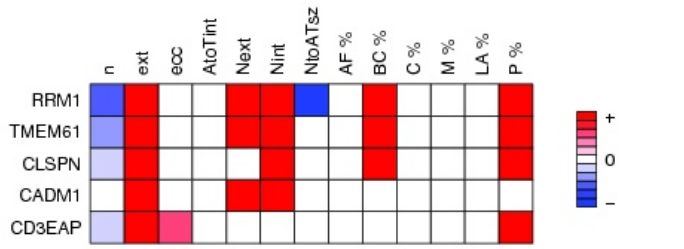
Phenotype landscape





- BL phenotype
- Bright nuclei
- Large nuclei
- Cells with protrusions
- Elongated cells
- Elong. cells with protrusions
- SM phenotype
- Small cells
- Low eccentricity cells
- High actin ratio cells
- Metaphase cells
- Other phenotype
- Actin fiber cells
- Big cells
- Large cells
- Lamellipodia cells
- Lamell. + high actin ratio cells
- Proliferating cells



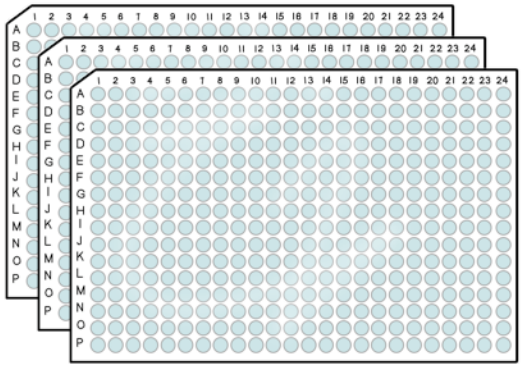


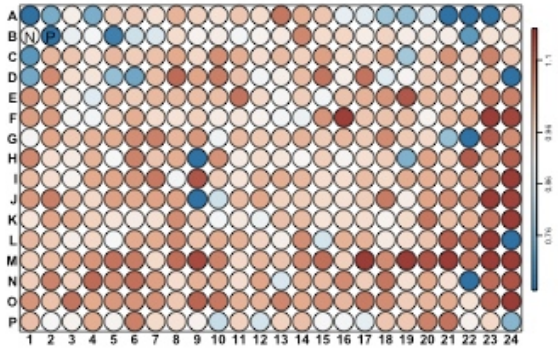
cellHTS2

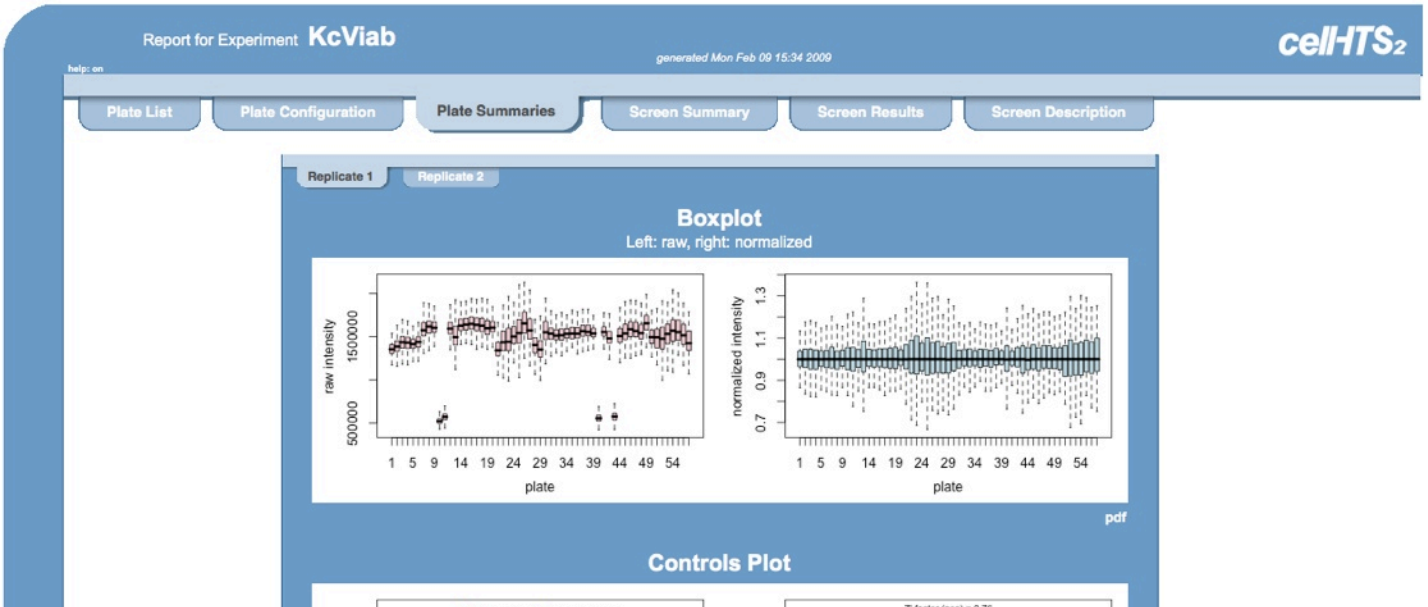
Analysis of high-throughput screens with low-order readout

Generation of analysis reports and scored phenotype lists

- Configuration
- Screenlog
- Annotation

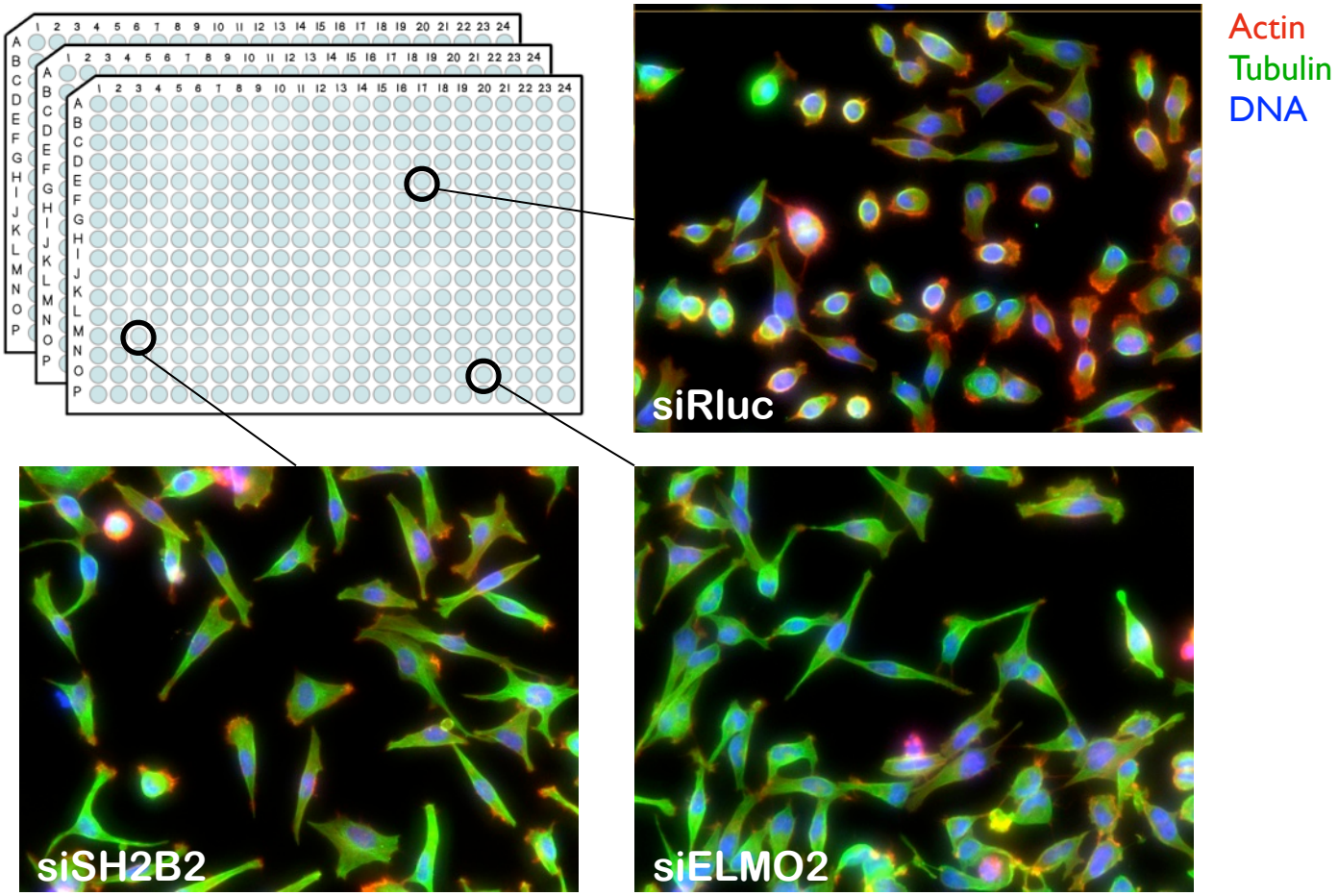




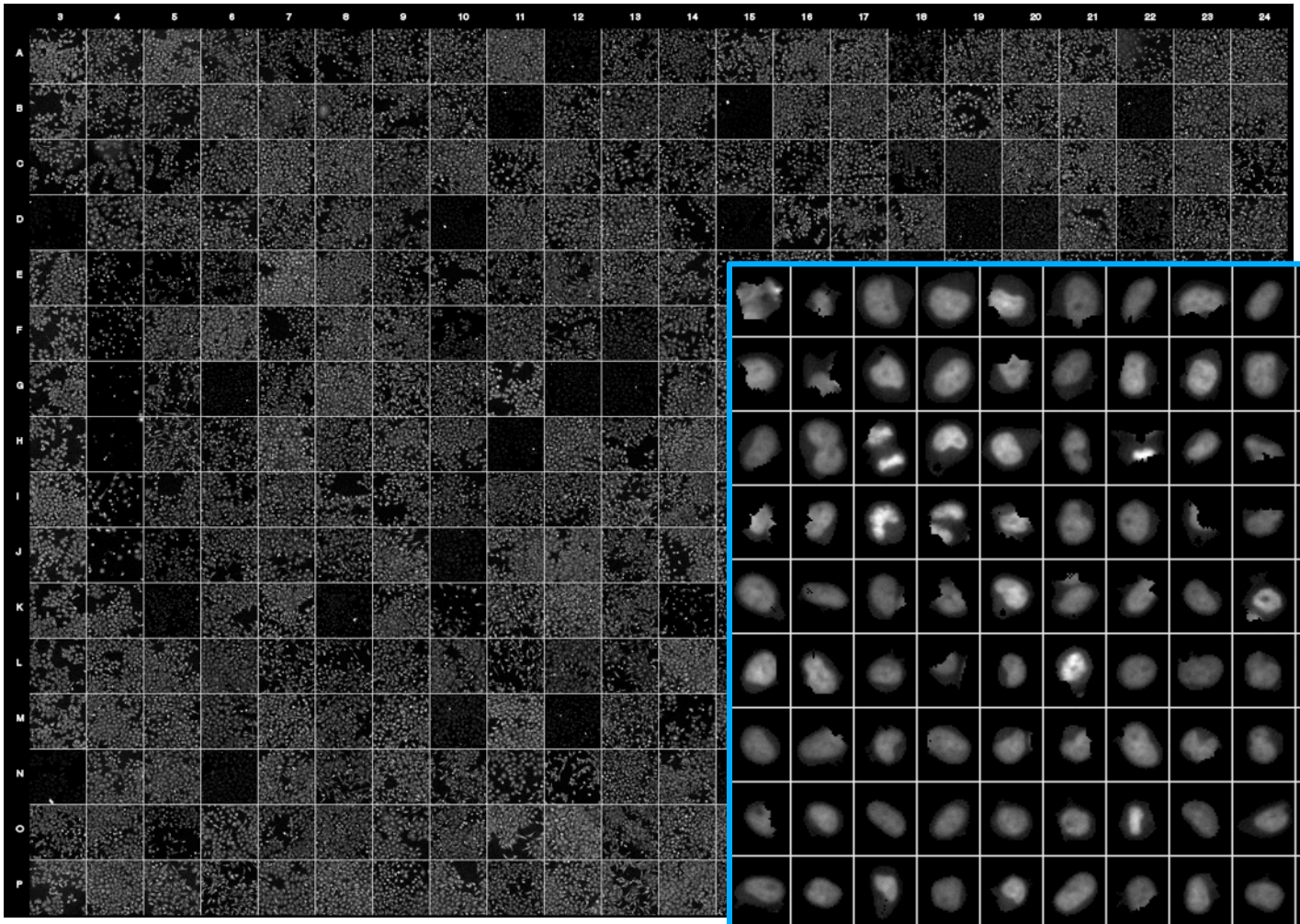


imageHTS

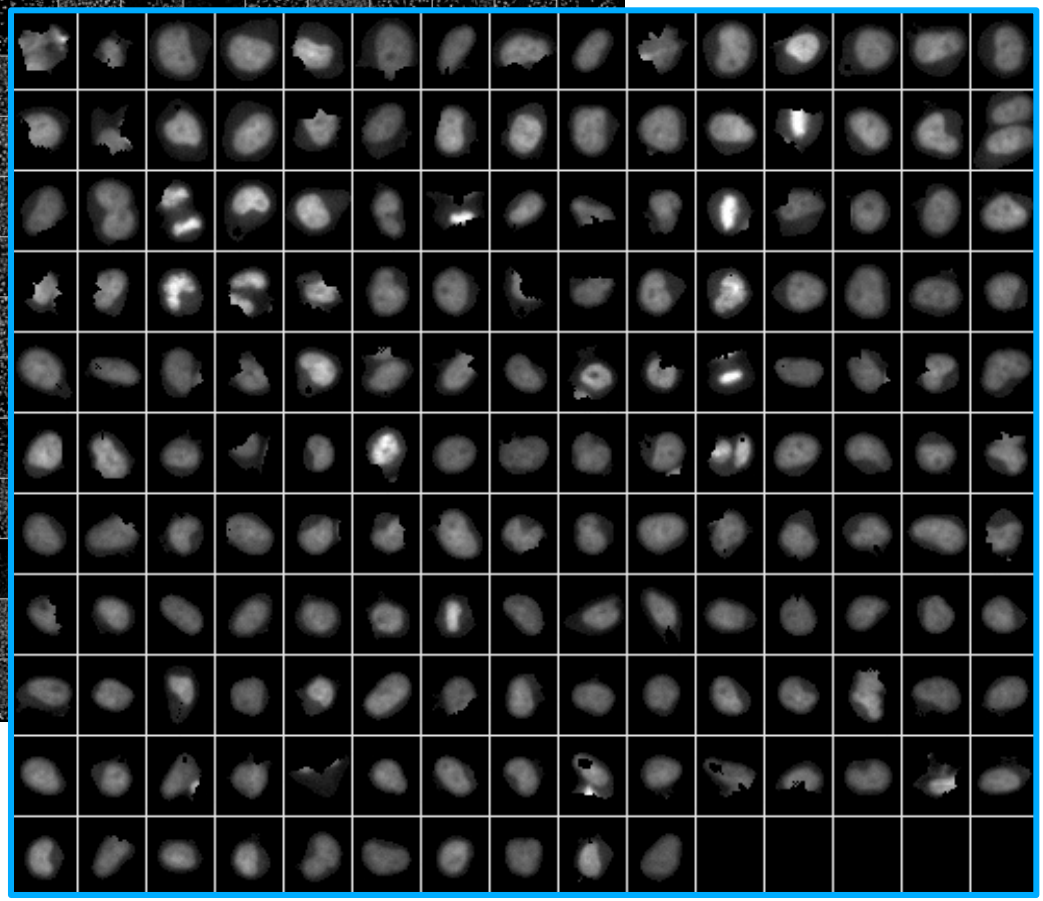
Analysis of high-throughput high-content assays



Quality metrics and plots



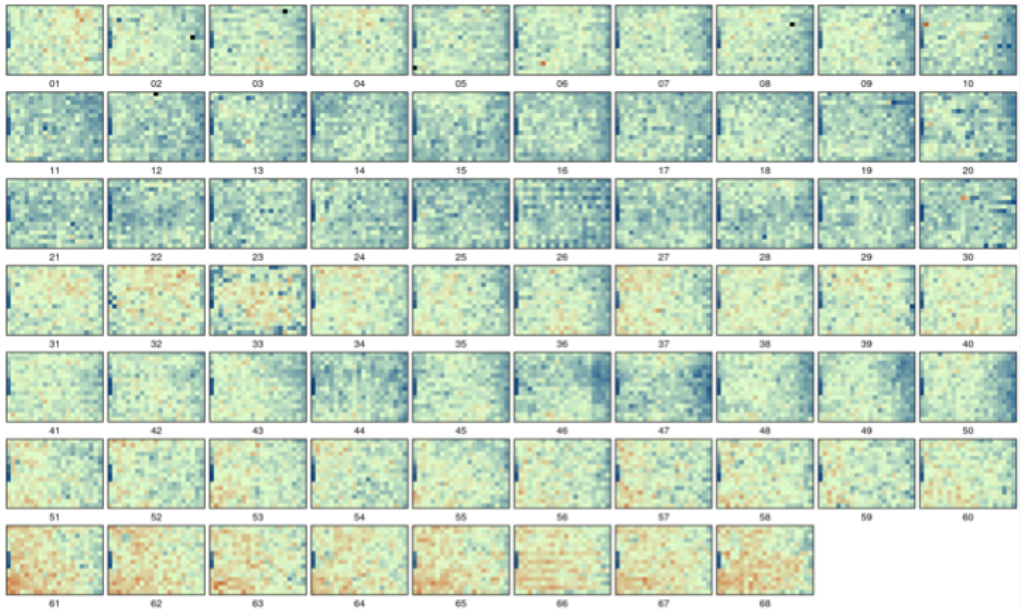
Thumbnail overview of one plate's images



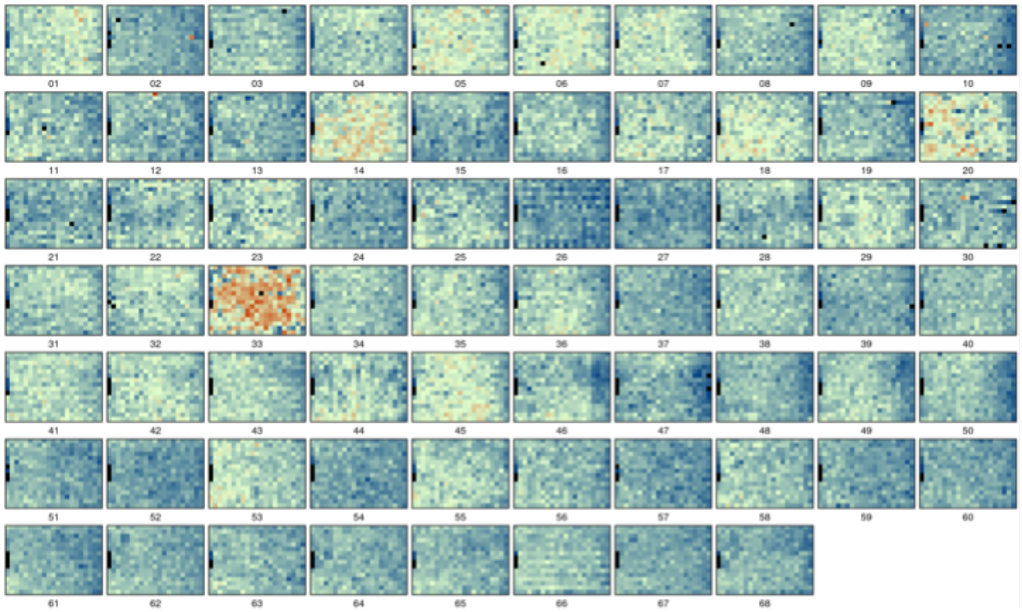
Gallery view of segmented objects of one well

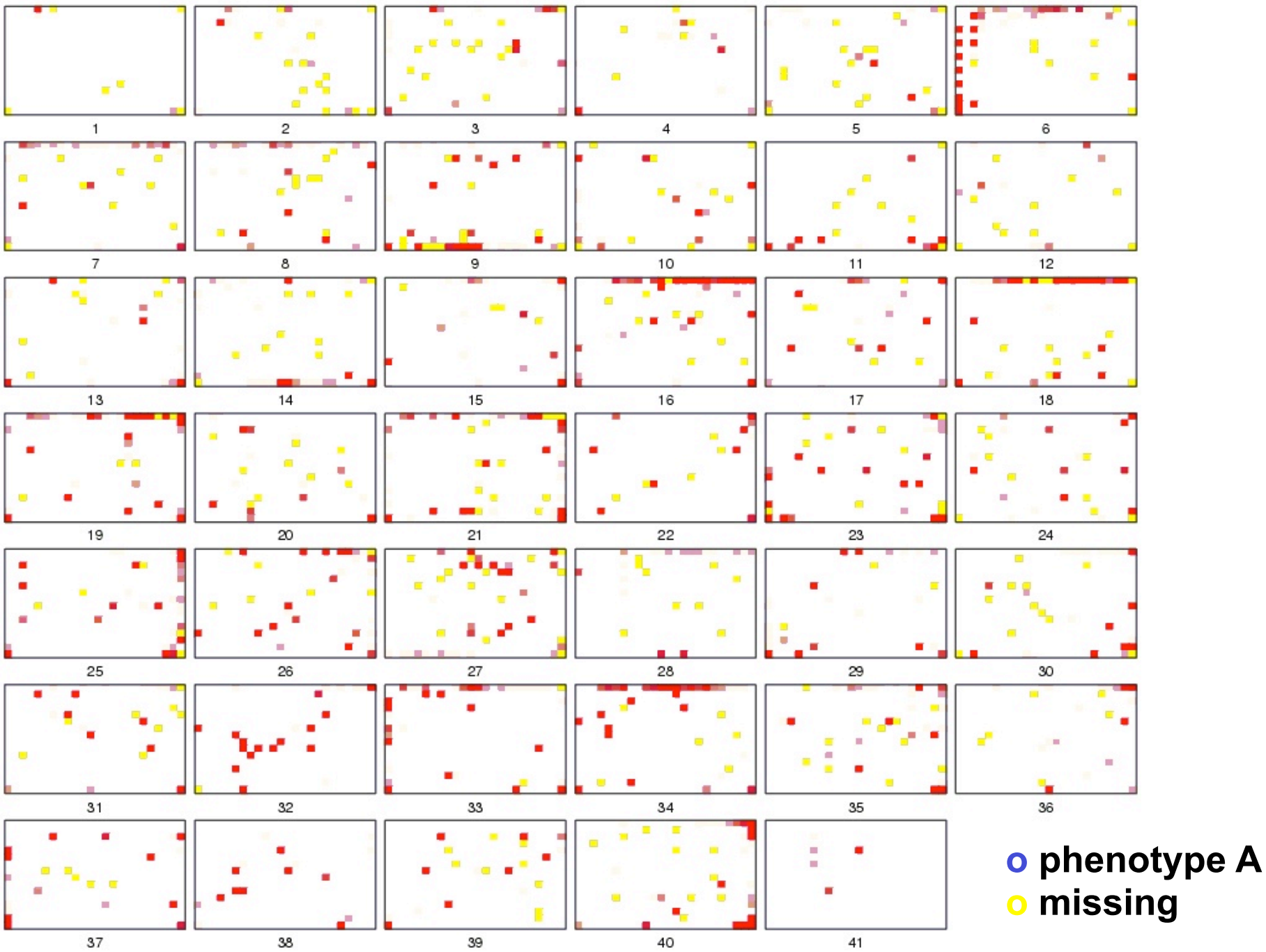
Long term drifts

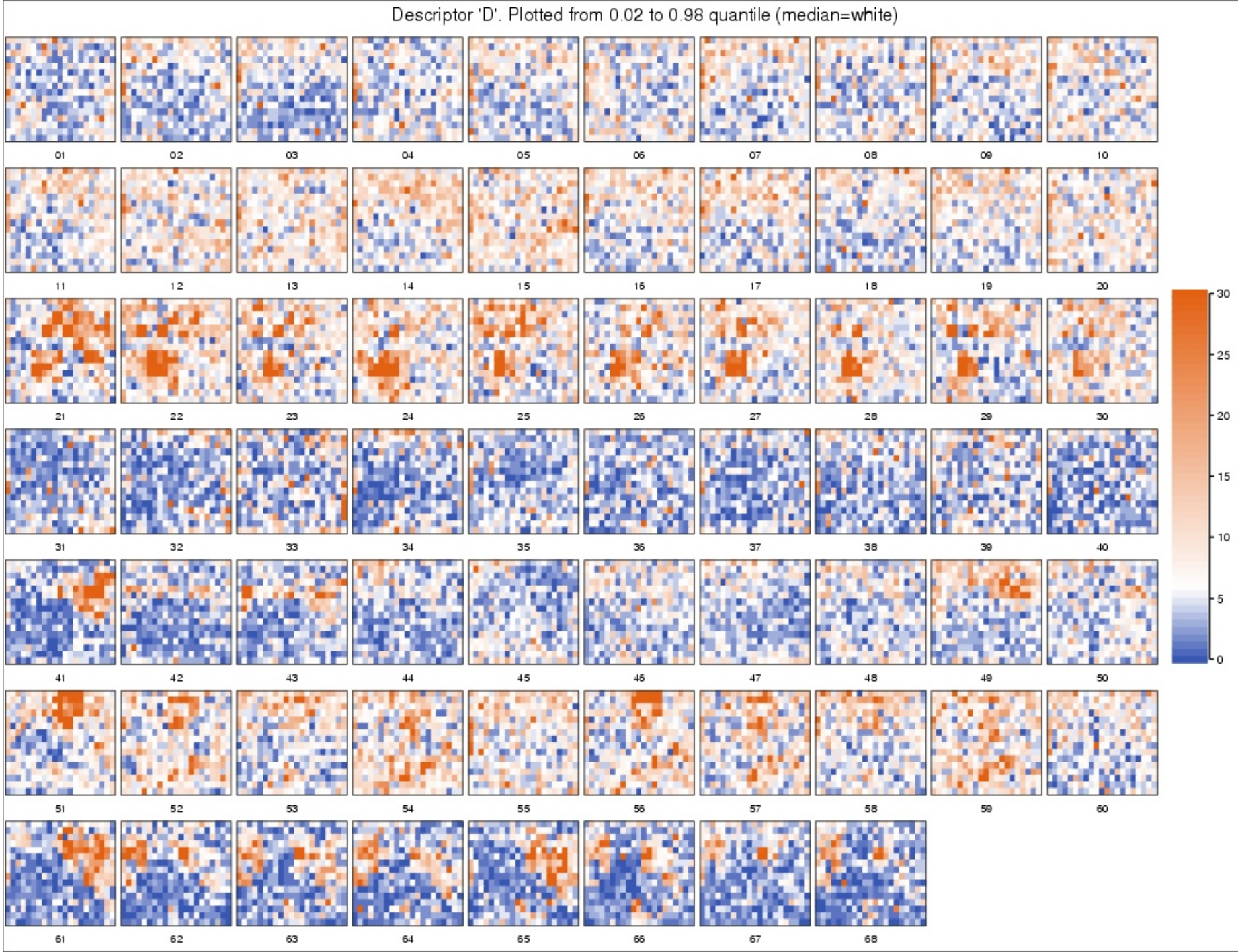
Number of cells



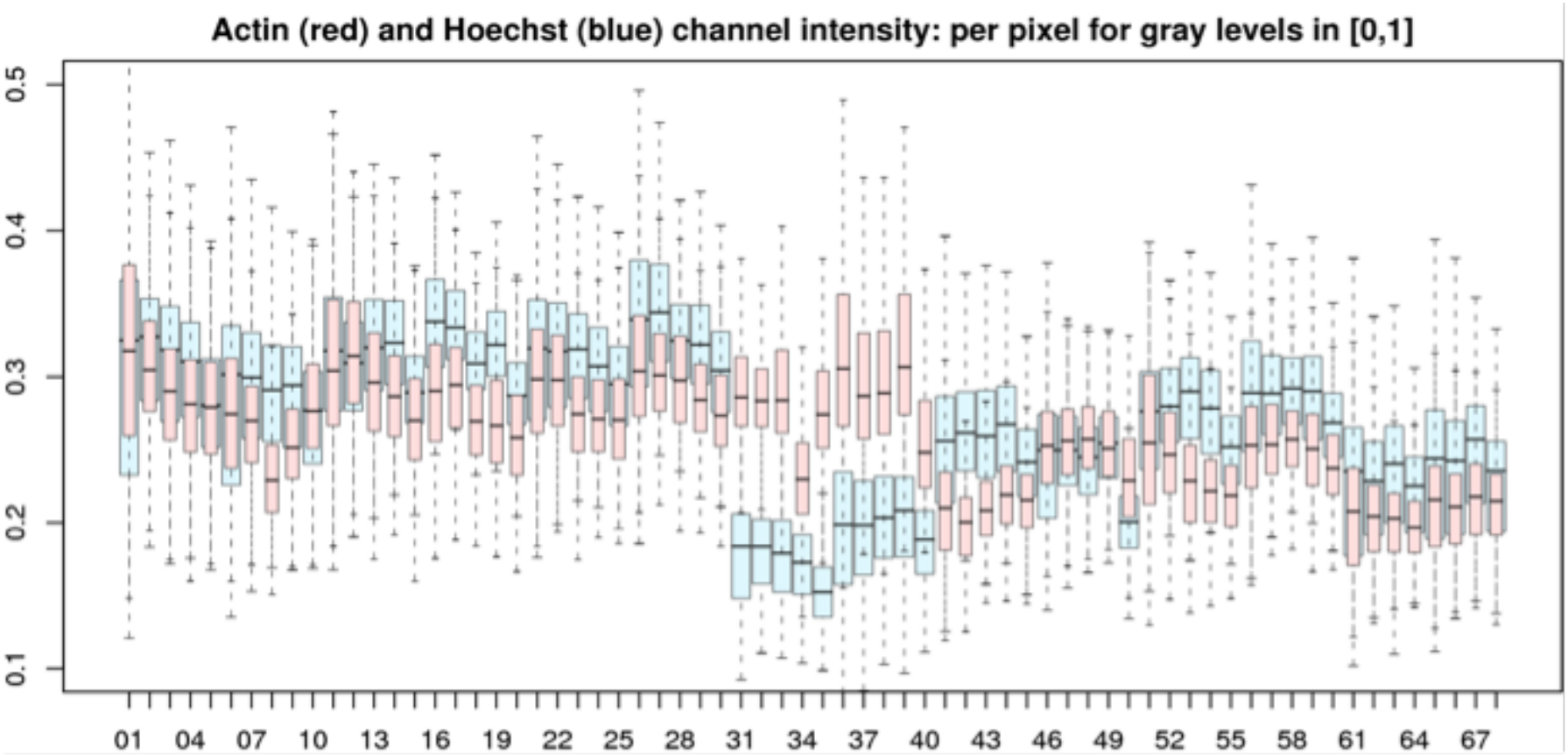
Number of cells / no. cells in negative controls in same plate



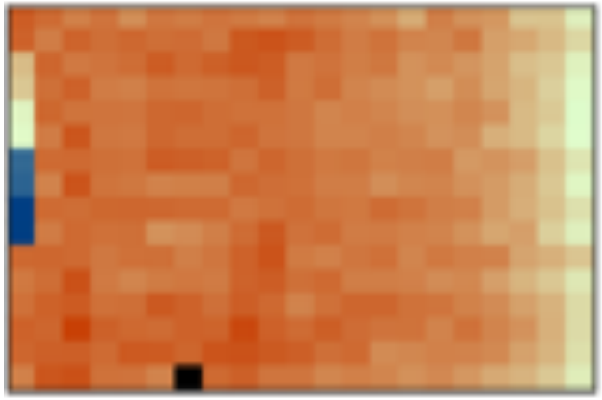

KcCellTiter033107–wh.txt (samples only)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
phenotype A
missing


Descriptor 'D'. Plotted from 0.02 to 0.98 quantile (median=white)
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
30
25
20
15
10
5
0

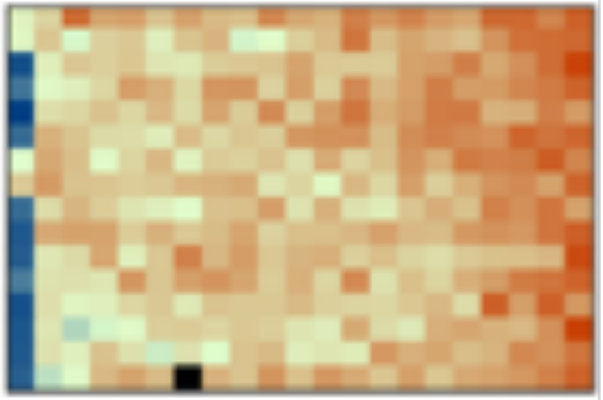
Batch effects



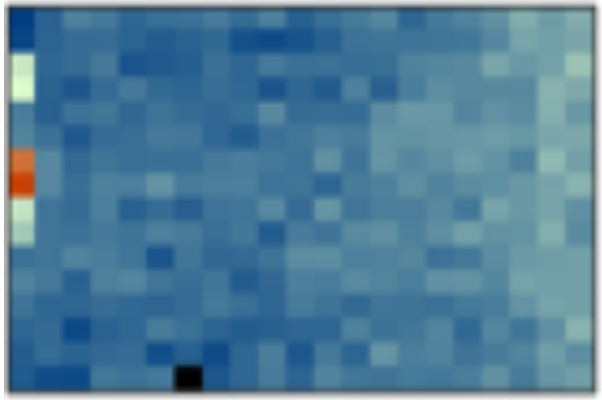
Within plate spatial trends (averaged over multiple plates)



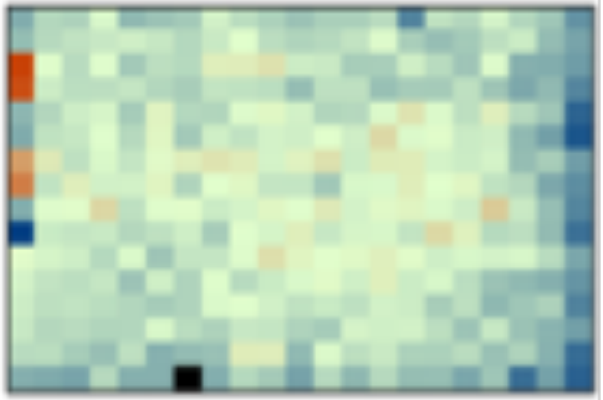
Number of cells



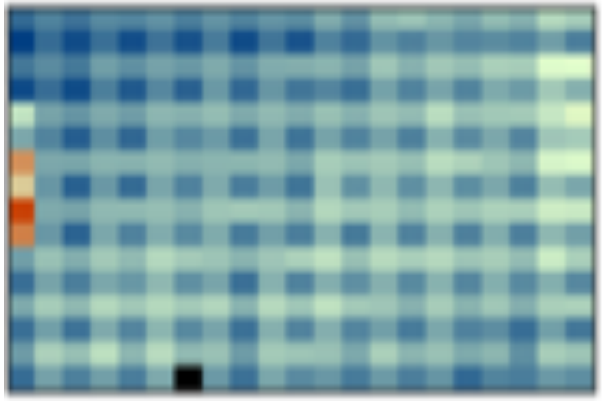
Eccentricity



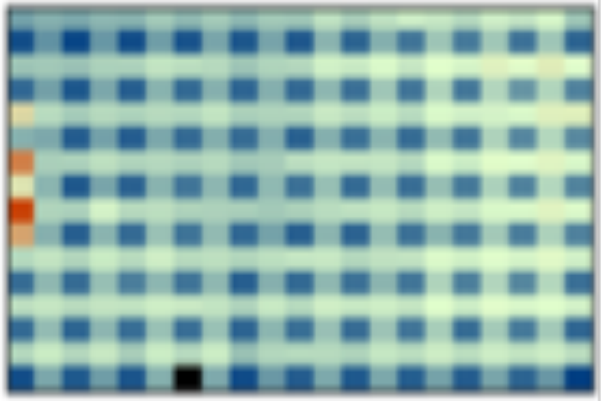
Cell size



Nuclei size

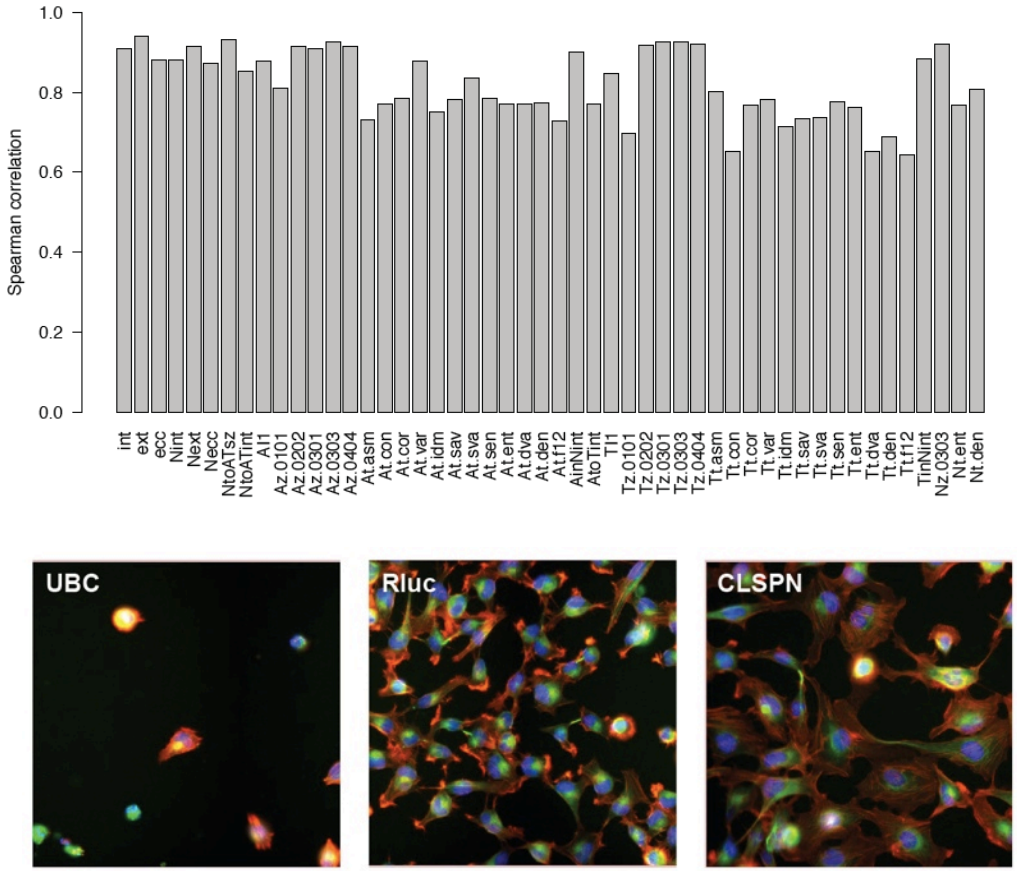


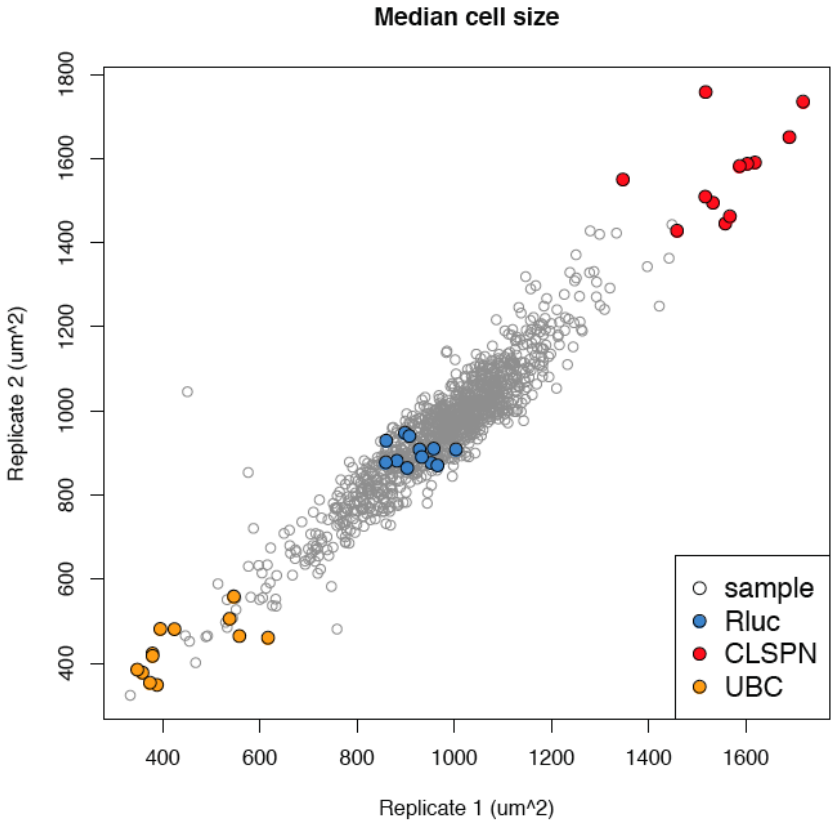
Actin intensity



Hoechst intensity

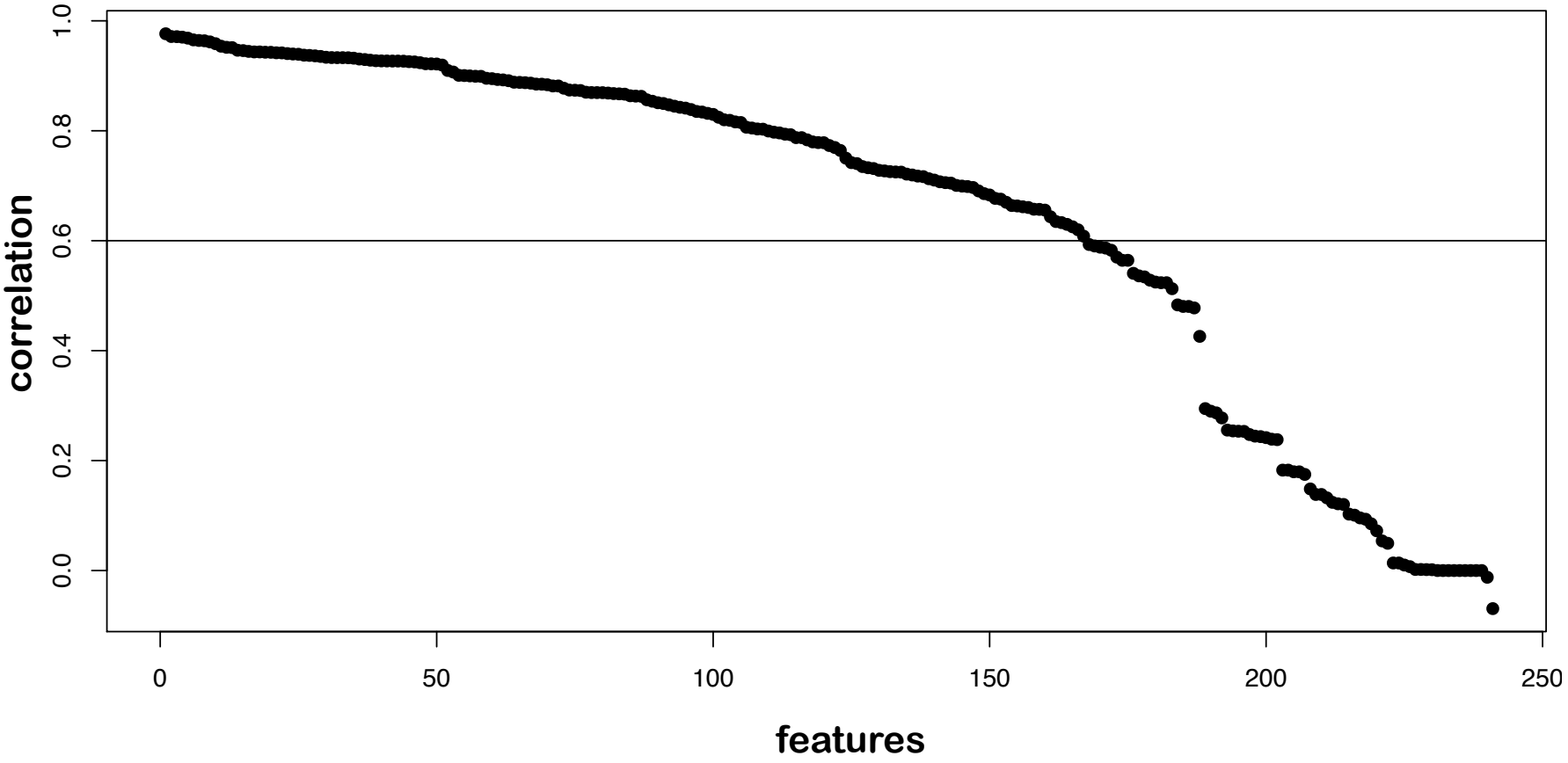
Quality metrics: reproducibility, controls





Fuchs, Pau *et al.*, Molecular Systems Biology 2010

Quality control of features

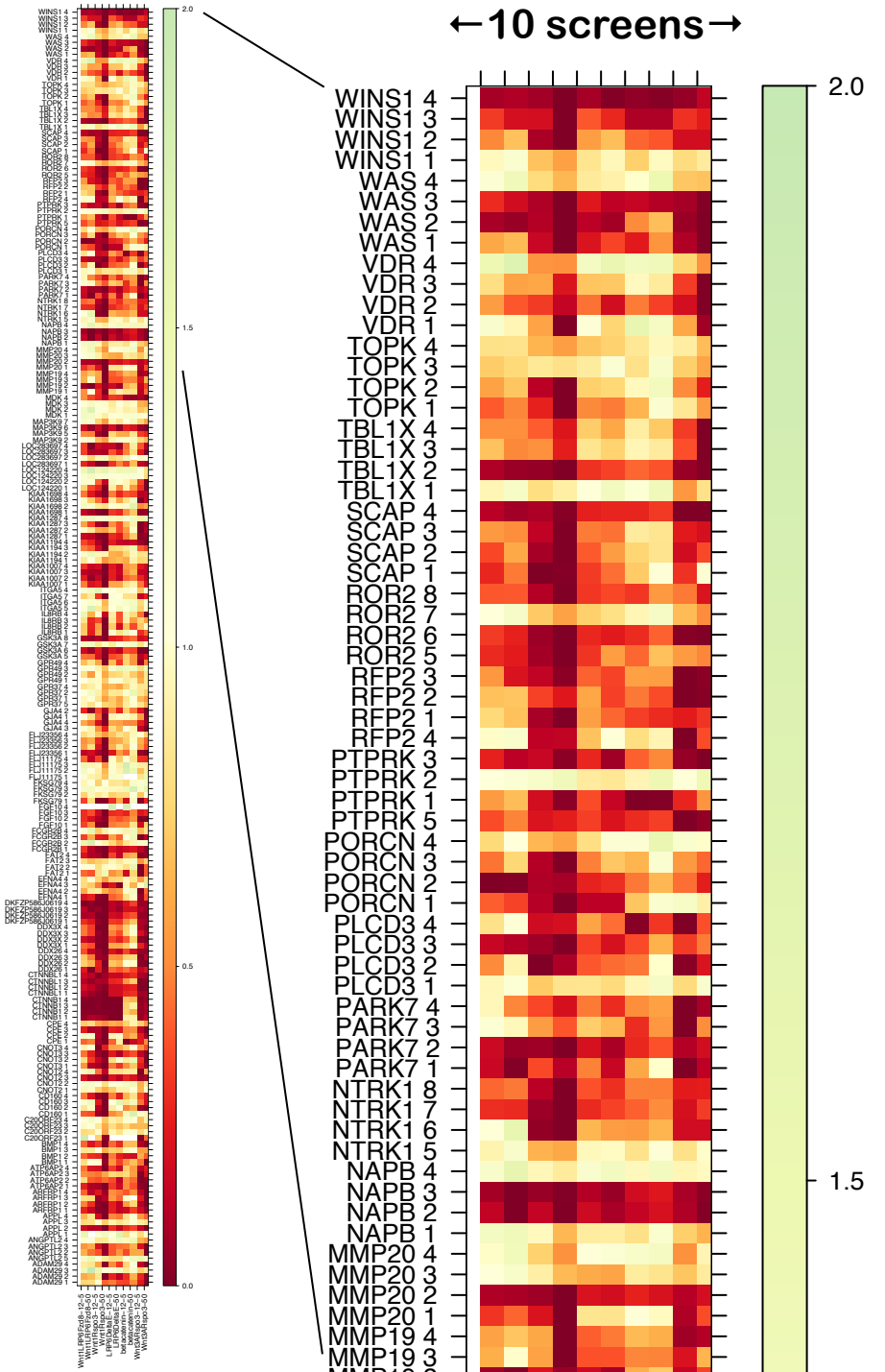


Quality criterium:
Correlation of interaction profiles between replicates and number missing values
162 features passed QC

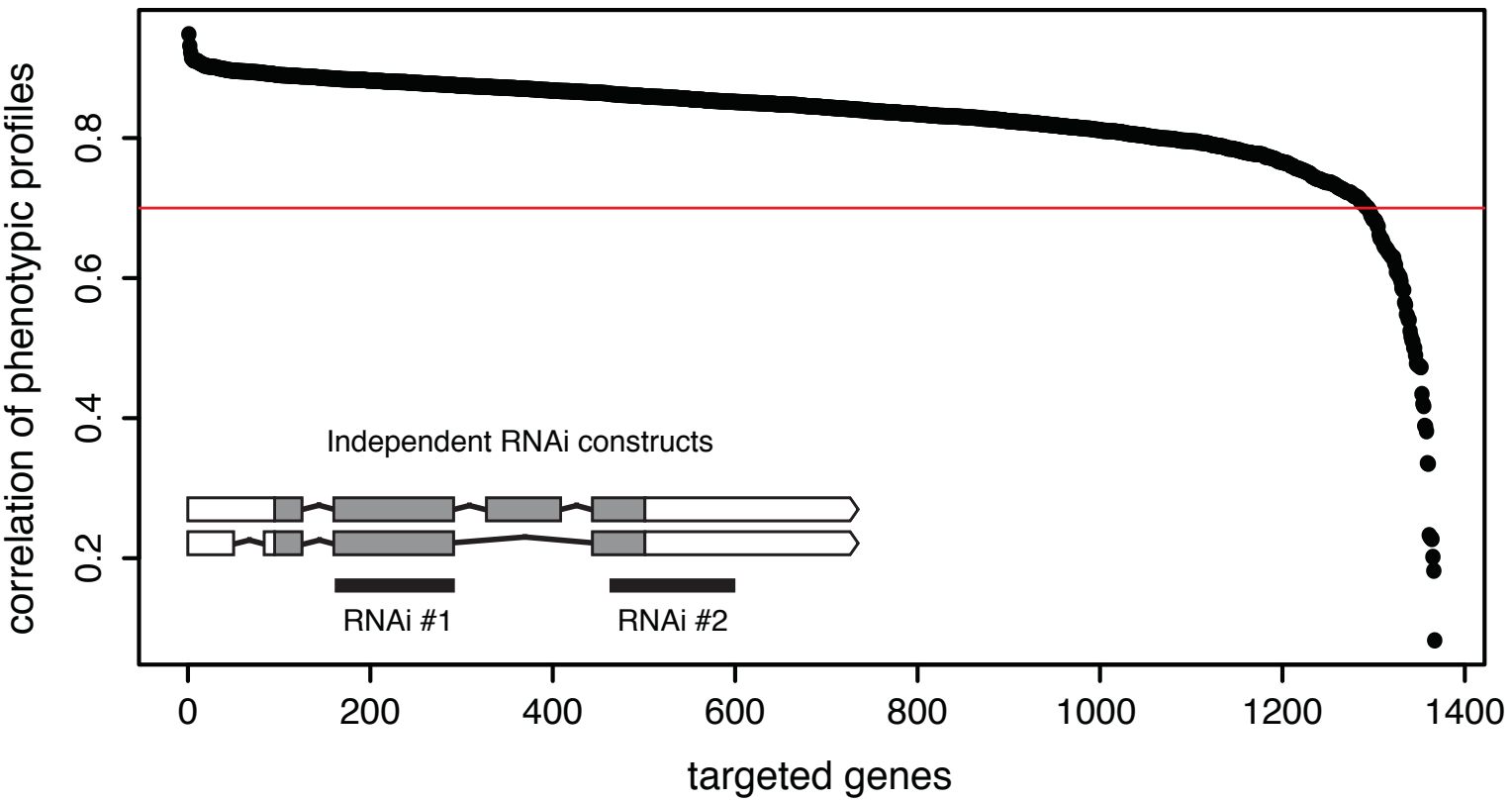
Concordance of siRNAs

Dharmacon library, 4 siRNAs per human target gene

10 screens: 5 conditions with 2 replicates each



Quality control of dsRNA designs



detection of possible off-target effects:
2 independent dsRNA designs per gene
quality criterium:
cor. of multi-phenotype interaction profile between designs
1293 genes passed QC

Simon Anders
Joseph Barry
Bernd Fischer
Julian Gehring
Bernd Klaus
Felix Klein
Andrzej Oleś
Małgorzata Oleś
Aleksandra Pekowska
Paul-Theodor Pyl
Alejandro Reyes
Maria Secrier

Gregoire Pau
Thomas Sandmann
Thomas Horn
Maximilian Billmann

Michael Boutros
Robert Gentleman
Jan Ellenberg
Martin Morgan

SEVENTH FRAMEWORK PROGRAMME
HFSP
DFG
EMBL