Portraying high-dimensional 'OMICs' data with individual resolution

© Hans Binder

Universität Leipzig

Interdisciplinary Centre for Bioinformatics

together with
Henry Wirth, Mario Fasold,
Lydia Hopp and Edith Willscher

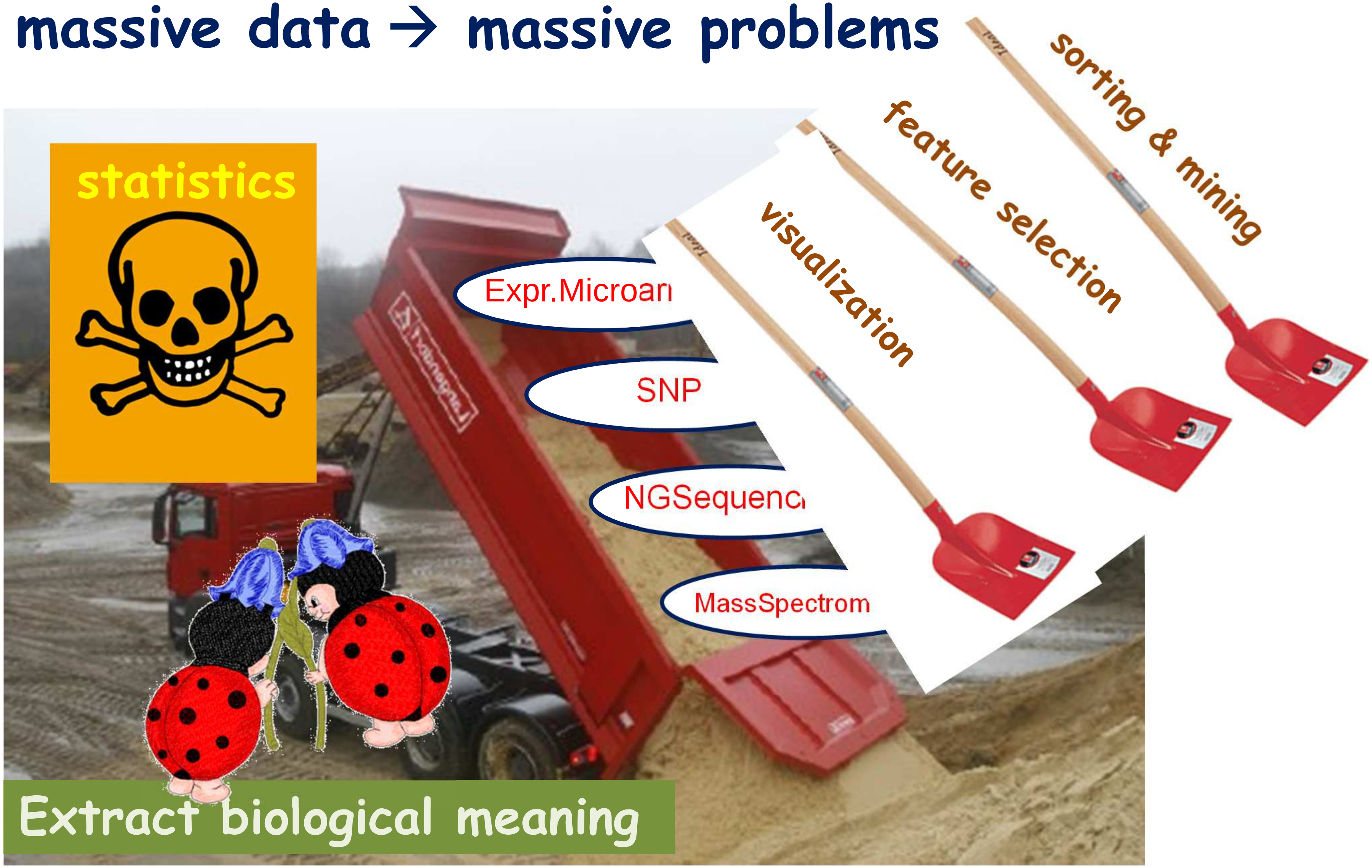
IZBI

UNIVERSITÄT LEIPZIG

CAMDA, Vienna July/2011

Motivation

massive data → massive problems



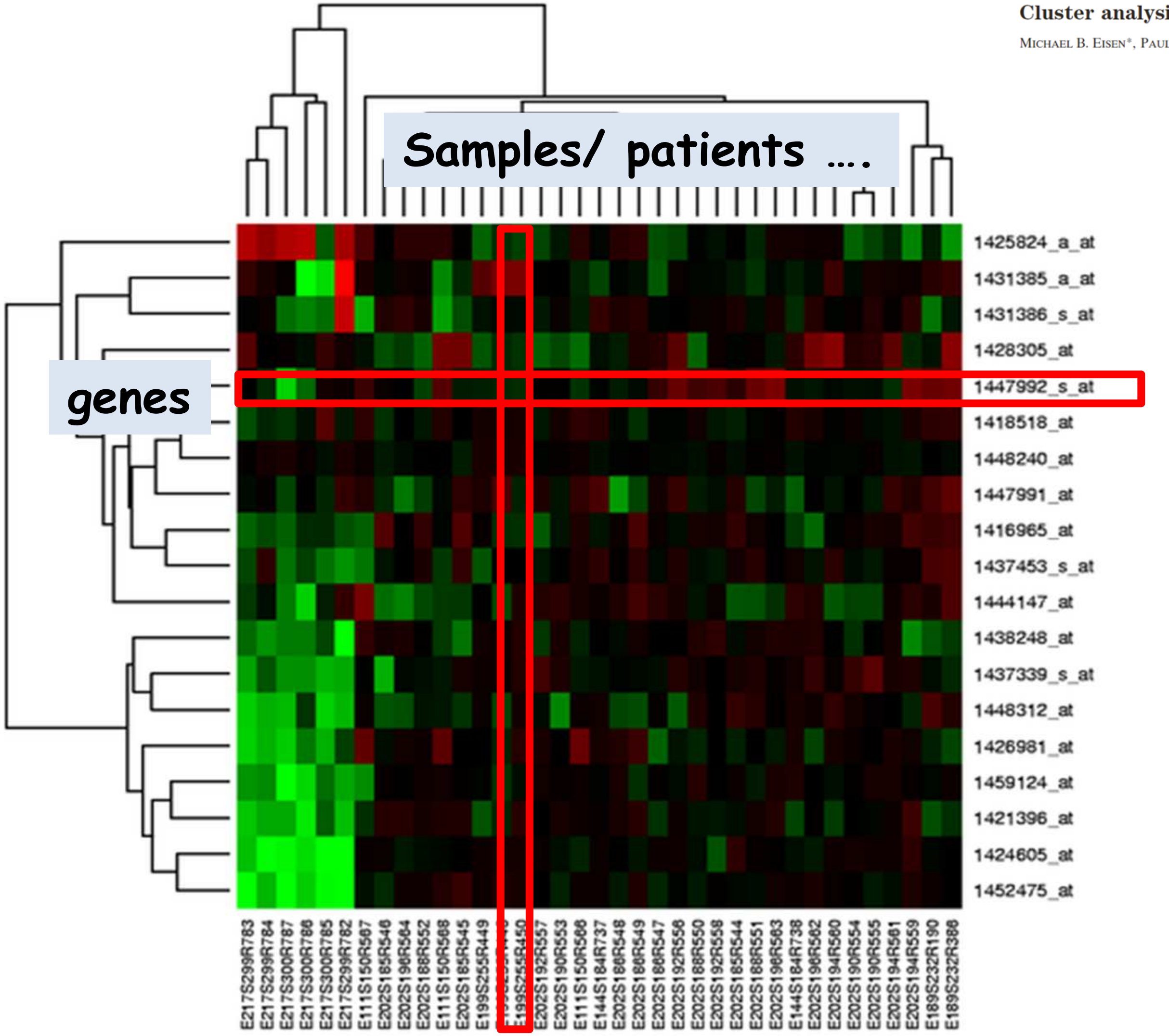
The ‚classical‘ approach

2way hierarchical clustering ‚heatmaps‘

Natl. Acad. Sci. USA
Vol. 95, pp. 14863–14868, December 1998
Genetics

Cluster analysis and display of genome-wide expression patterns

MICHAEL B. EISEN*, PAUL T. SPELLMAN*, PATRICK O. BROWN†, AND DAVID BOTSTEIN*‡



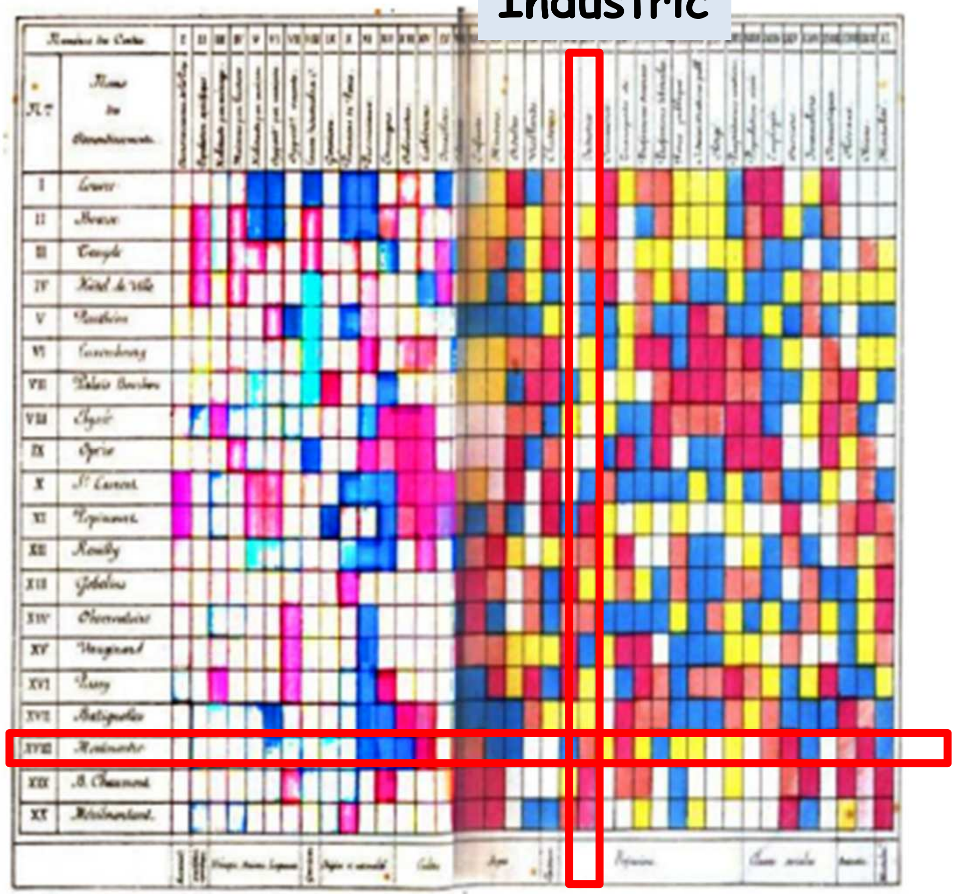
Sorting the data
(no compression):
1 gene = 1 point

An old problem

Loua, T. (1873), *Atlas statistique de la population de Paris*, Paris: J. Dejey.

Industric

20 districts of Paris



Montmartre

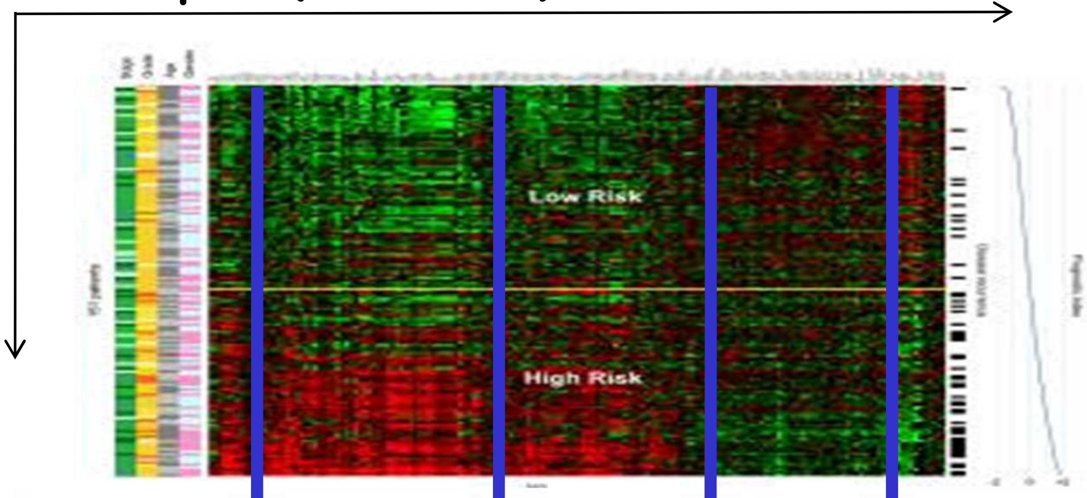
Color scale:
White: low
Yellow: medium-low
Blue: medium-high
Red: high

Features: national origin/ age/ profession/ social classes etc…

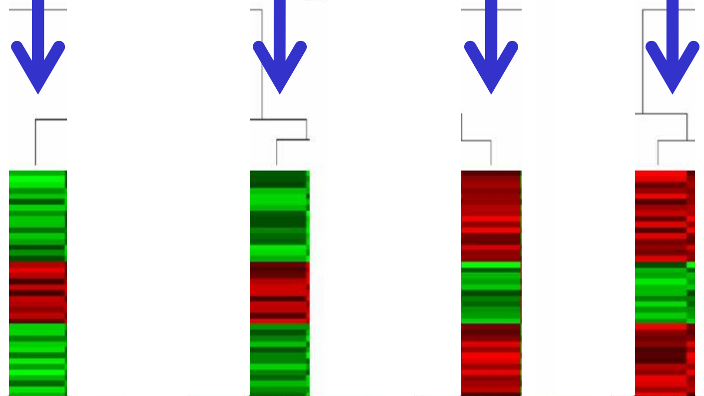
Figure 2: Shaded matrix display from Loua (1873). This was designed as a summary of 40 separate maps of Paris, showing the characteristics (national origin, professions, age, social classes, etc.) of 20 districts, using a color scale that ranged from white (low) through yellow and blue to red (high). A monochrome version can be found at http://www.math.yorku.ca/SCS/Gallery/images/loua1873-scalogram.jpg.

Samples (individuals)

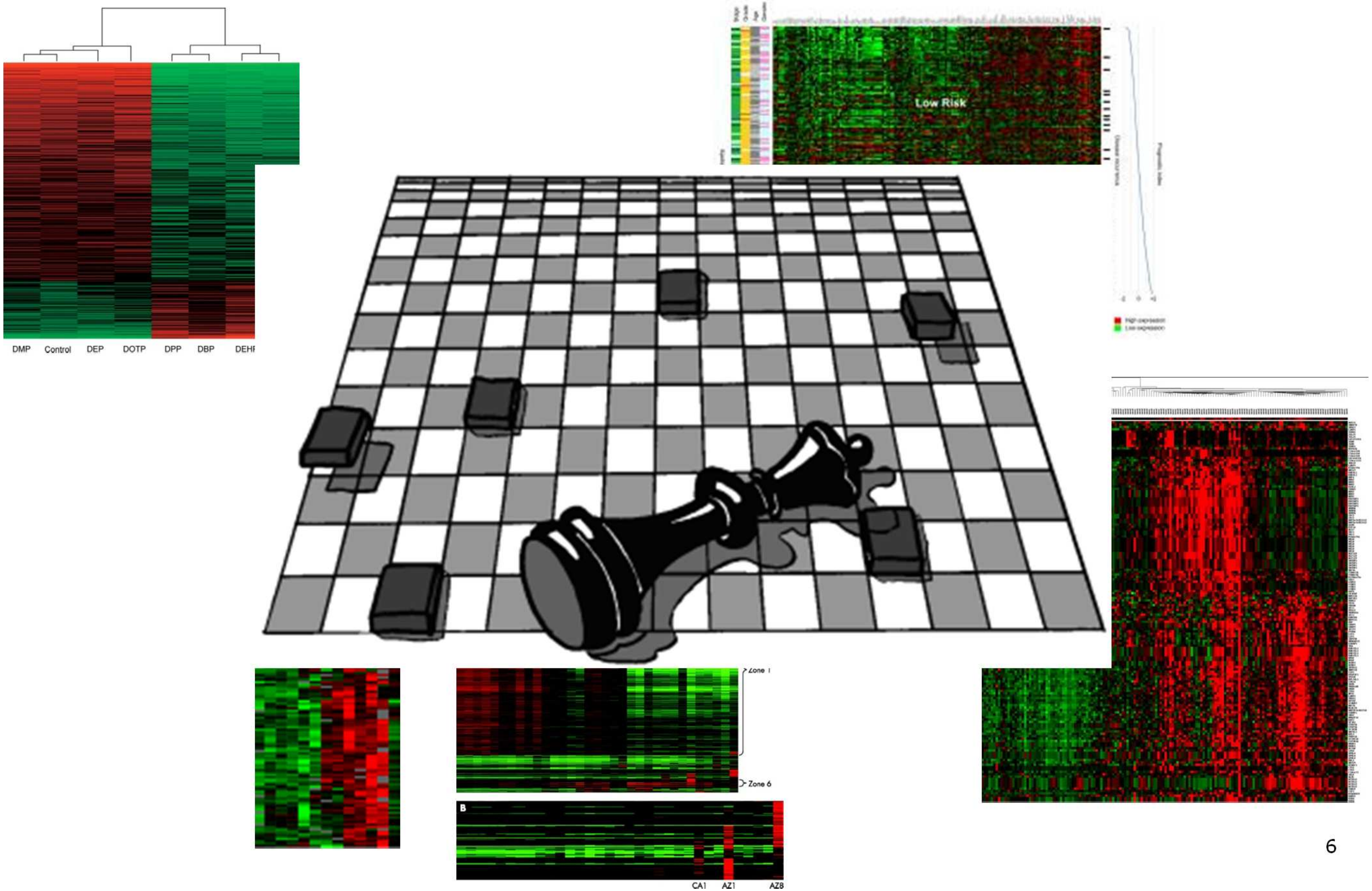
features



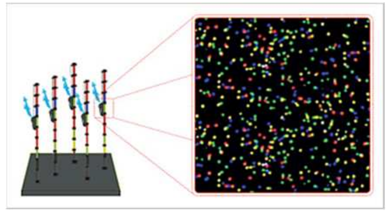
‚portrait' - vectors

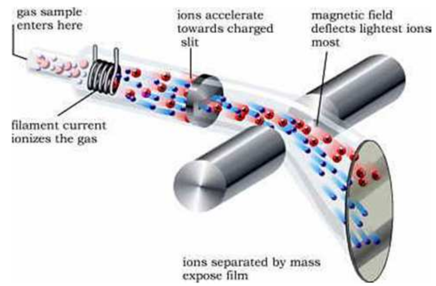


OMICs portraits
‚see the molecular faces'


Low Risk
DMP
Control
DEP
DOTP
DPP
DBP
Zone 1
Zone 6
B
CA1
AZ1
AZ8

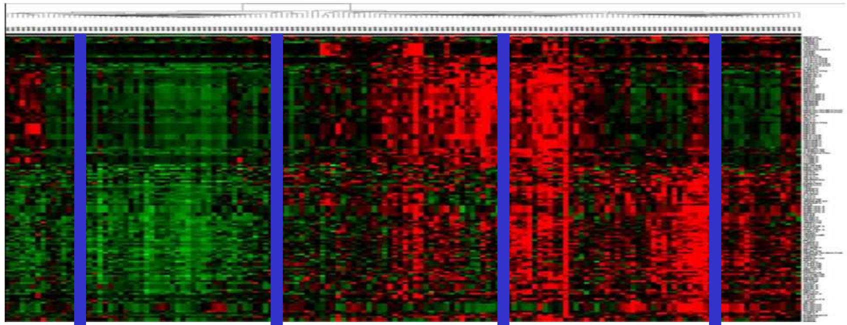
High throughput data
(e.g. OMICs)





Samples (individuals)

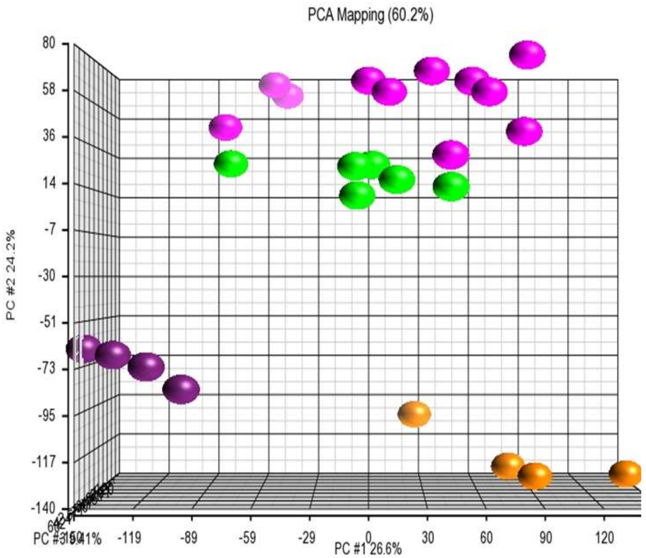
features



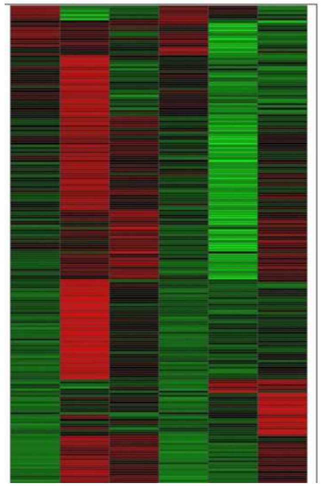
OMICs portraits
‚see the molecular faces'

The intermediate view: portraying the omic-faces

Intermediate: portraying high-dimensional Omics data



detailed gene centered view



nple

Requirements:

- **visual idendity** for each sample
- data **compression**…
- …without loss of **information**
- expressing **intrinsic features** of biological impact…
- … which can be treated as **new, complex objects** for next level analysis…

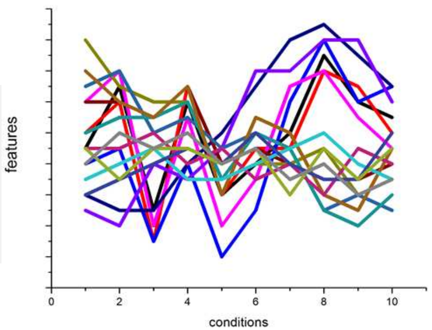
Sorting maschine

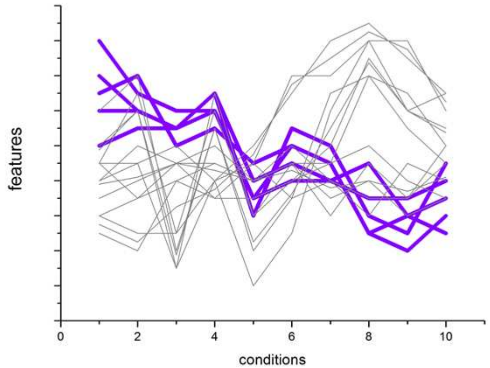
Data (letters)

sorting
compression

Transformed data
(boxes, towns)

values



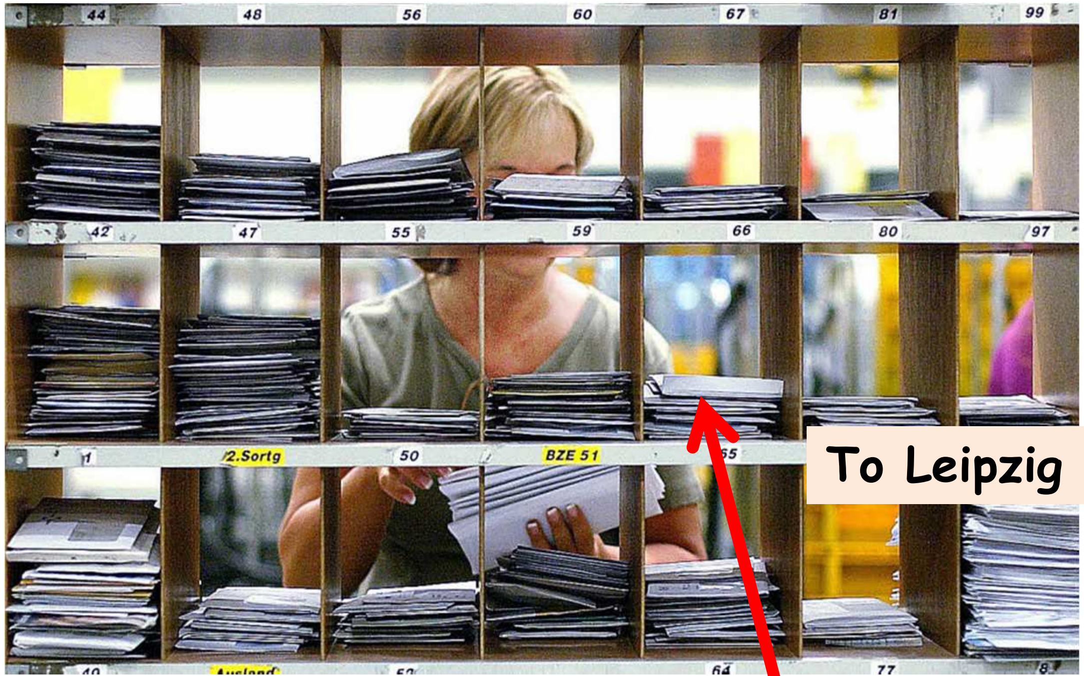


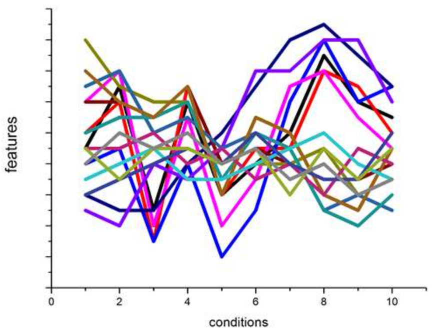
Cluster of similar
behaving features

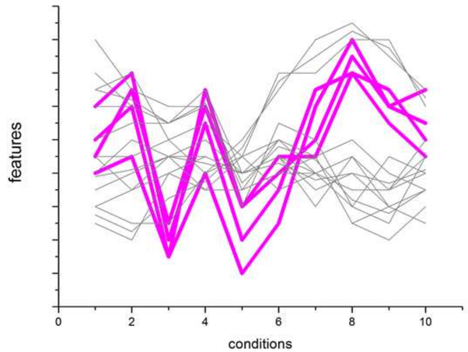
individuals/conditions

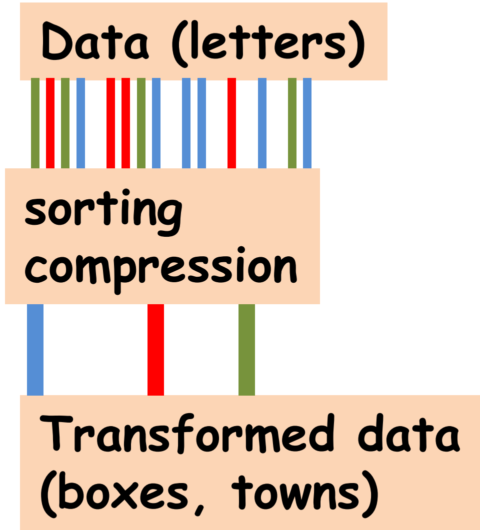
Sorting maschine

IZBI University Leipzig

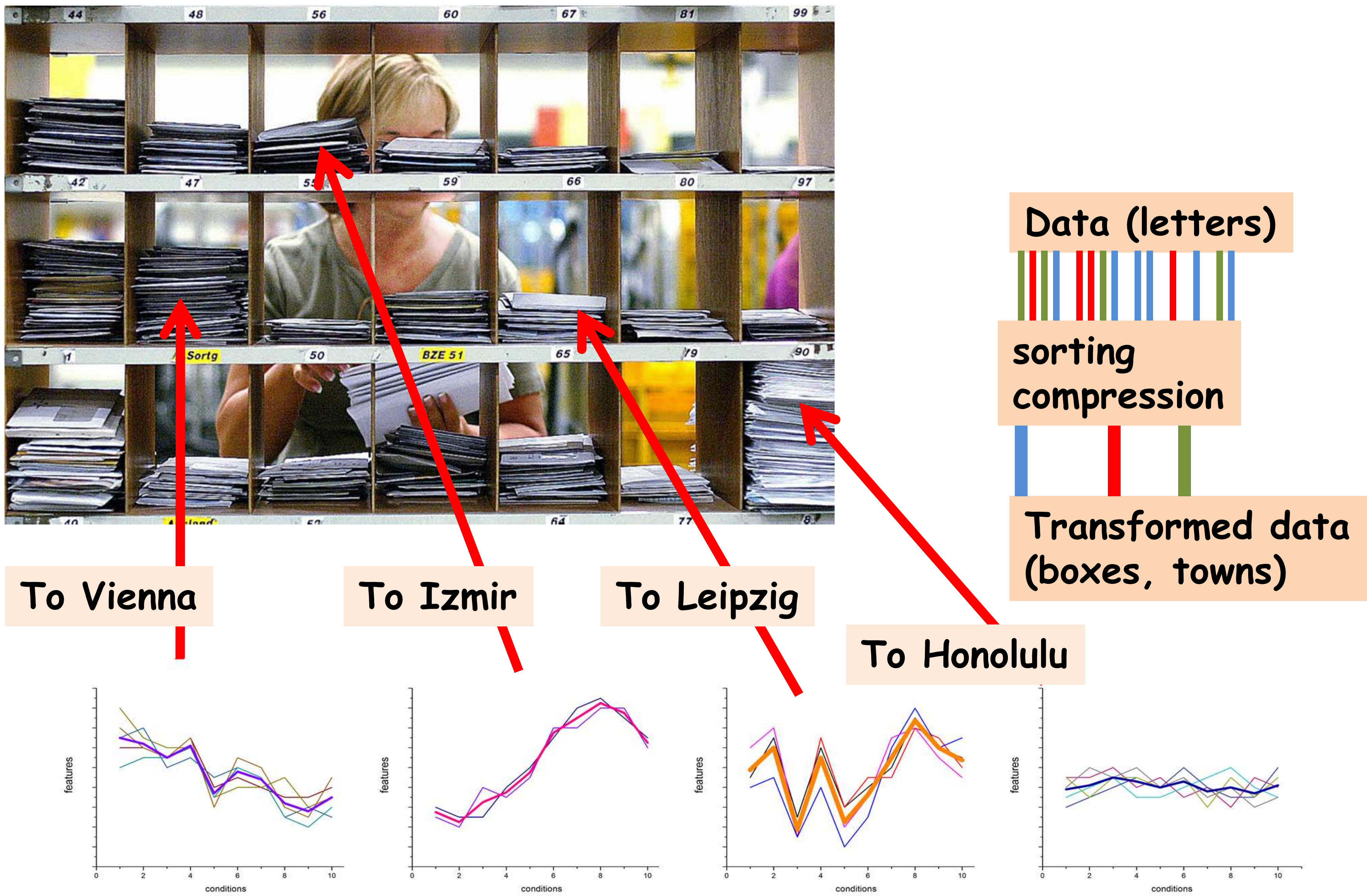








Sorting maschine



SOM machine learning: the sorting machine

Self-Organized Formation of Topologically Correct Feature Maps

Teuvo Kohonen

Department of Technical Physics, Helsinki University of Technology, Espoo, Finland

Biol. Cybern. 43, 59–69 (1982)

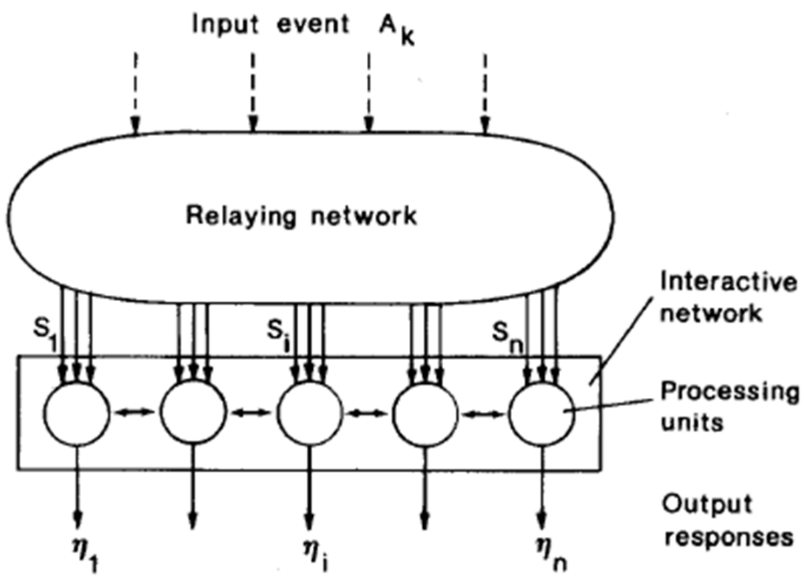
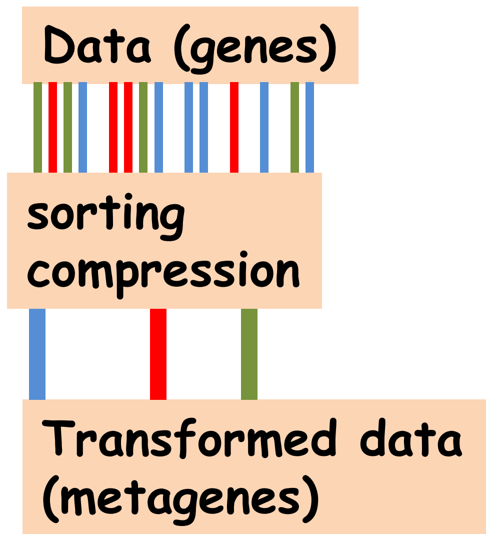


Fig. 1. Illustration of a system which implements an ordered mapping



Outline

1. Explain what SOM does !

2. How SOM can help to understand massive OMICs data

Examples (array expression data):

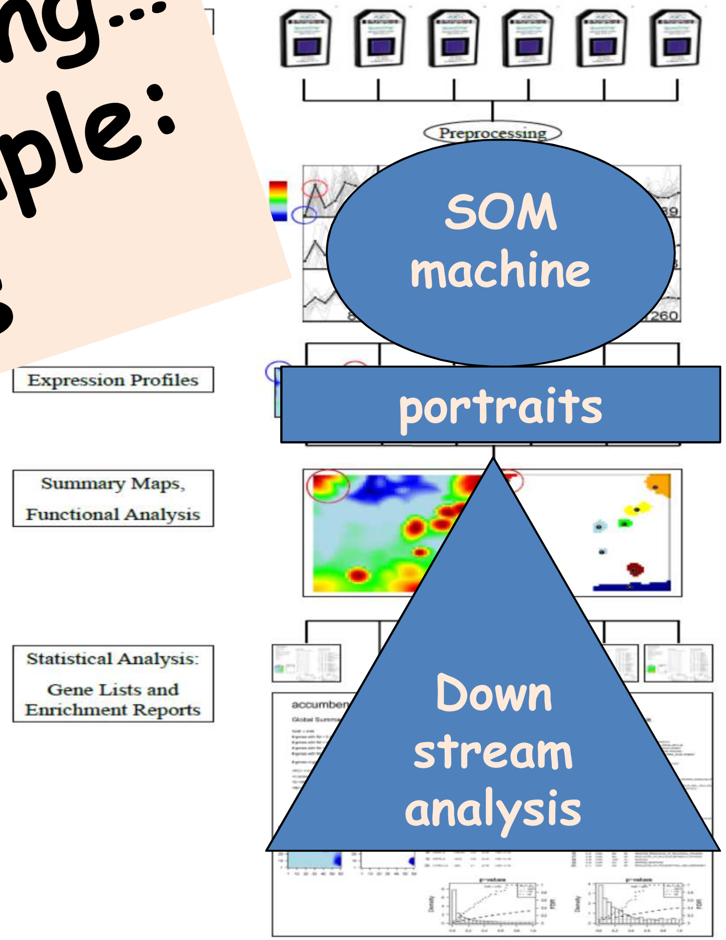
a) Human tissues
Well classified, diverse expression → teaching example

b) B-cell Lymphoma
Just another cancer → molecular cancer subtypes

c) Glioma Multiform
Its the CAMDA-‚must!' data set?
(we started after May 15th 2011)

Worked example: SOM atlas of human tissues

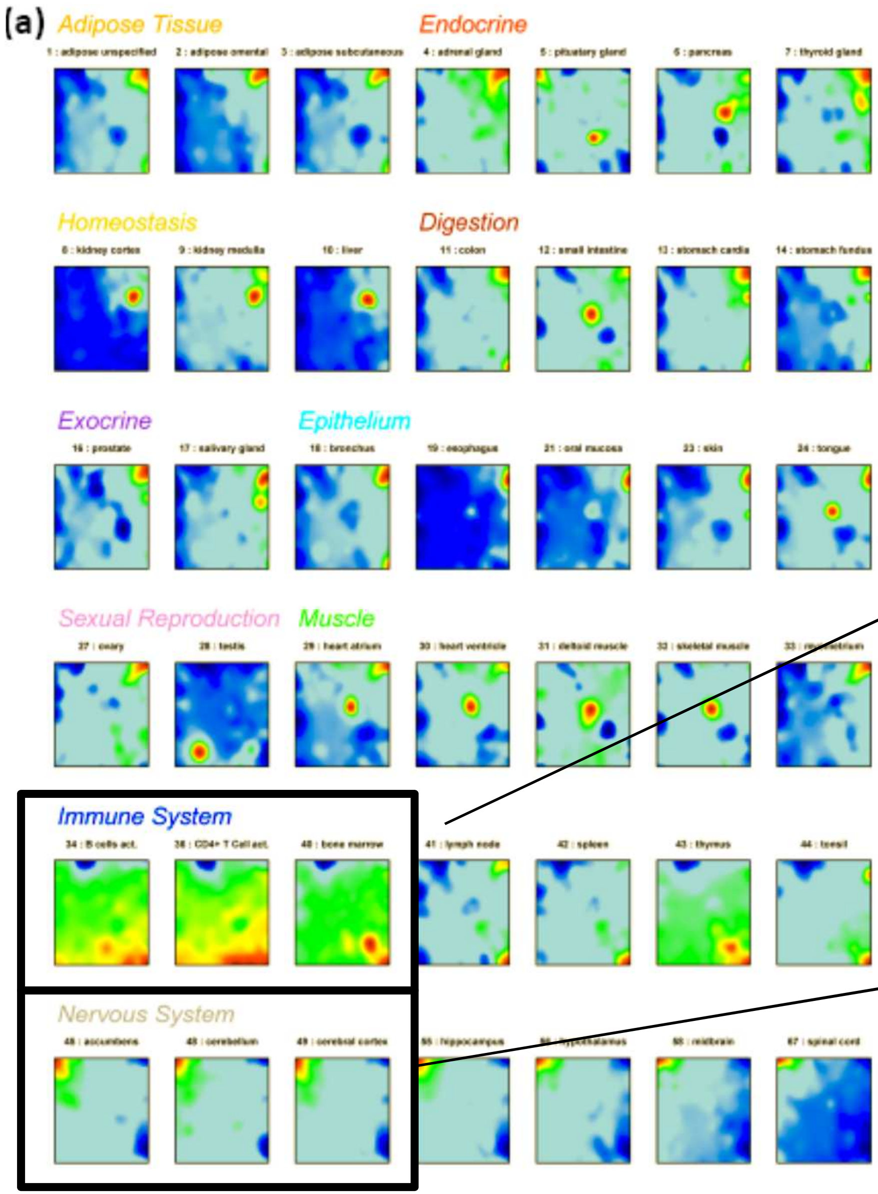
67 human tissues (Affy array data)
9 categories
- heterogeneous expression
- ... fined samples

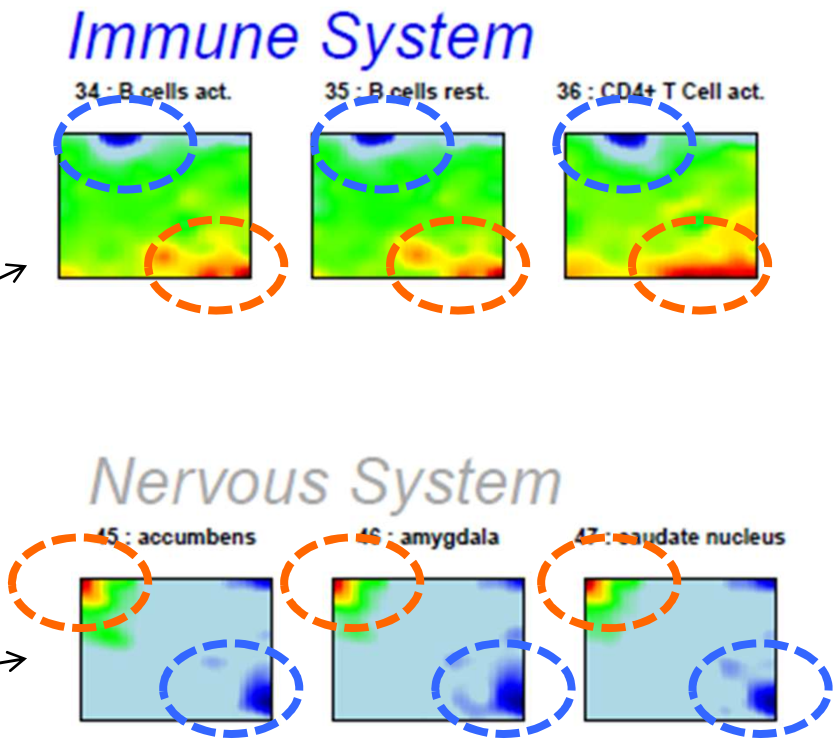


Learning by doing…
Teaching example:
Human tissues

Worked example: SOM atlas of human tissues

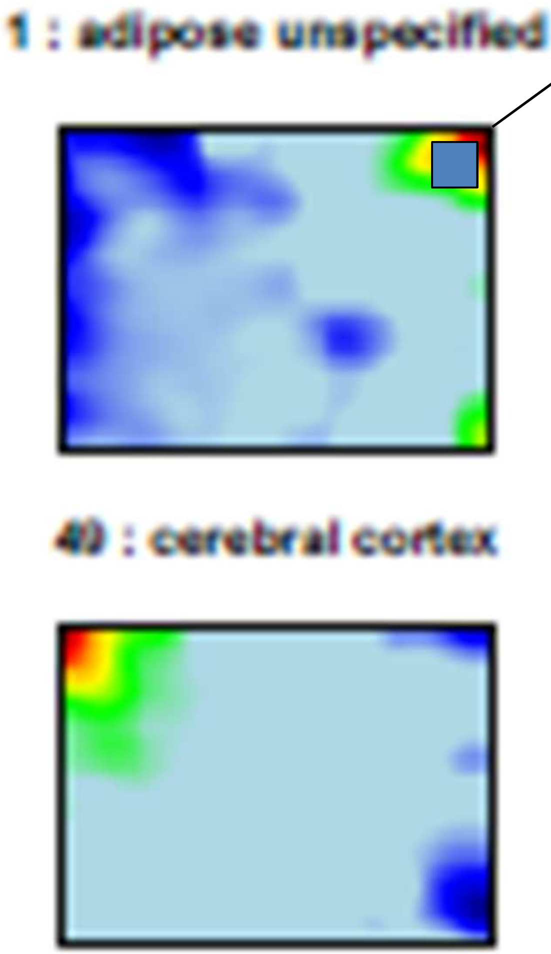
67 human tissues
9 categories
- **heterogeneous expression**
- **well defined samples**





Imaging of information

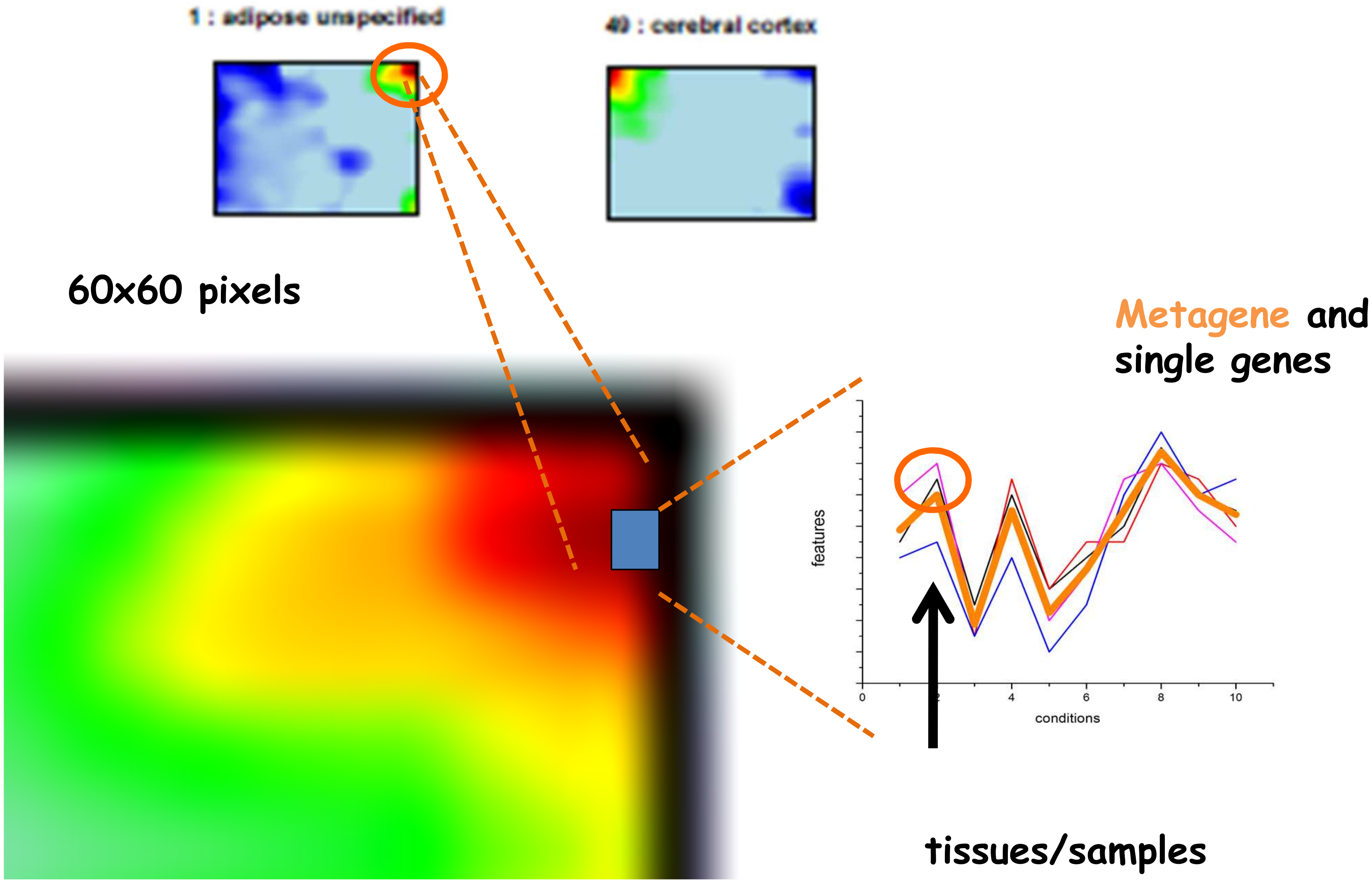
Image → pixels → information (brigthness, colour)



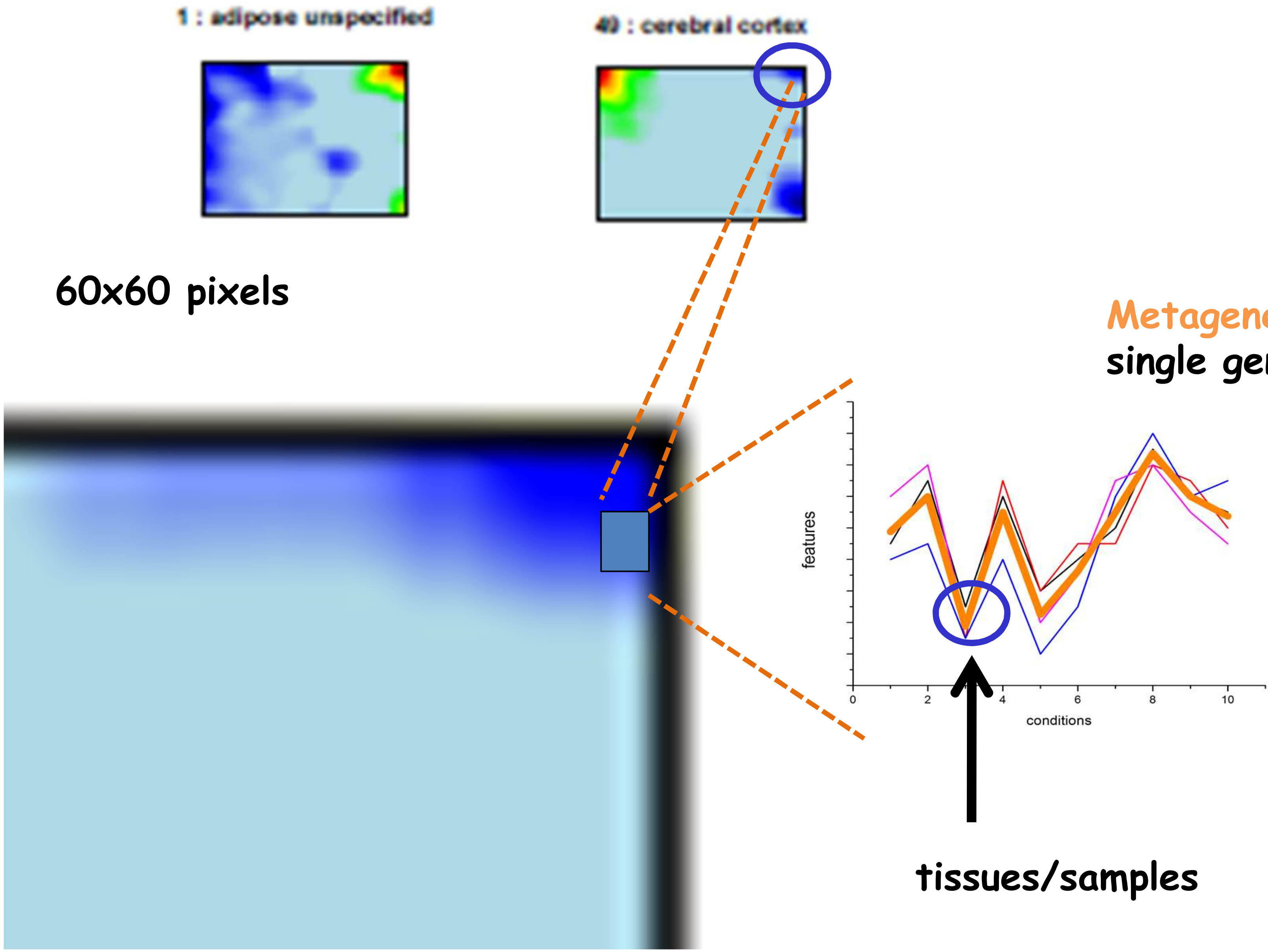
~60x60 pixels

~2000x1000 pixels

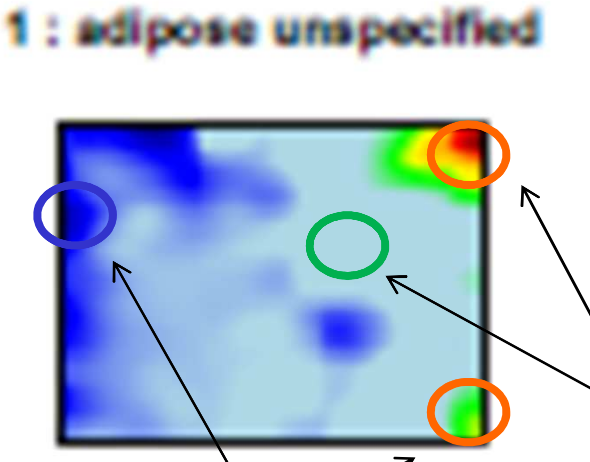
SOM image: clustering



SOM image: clustering

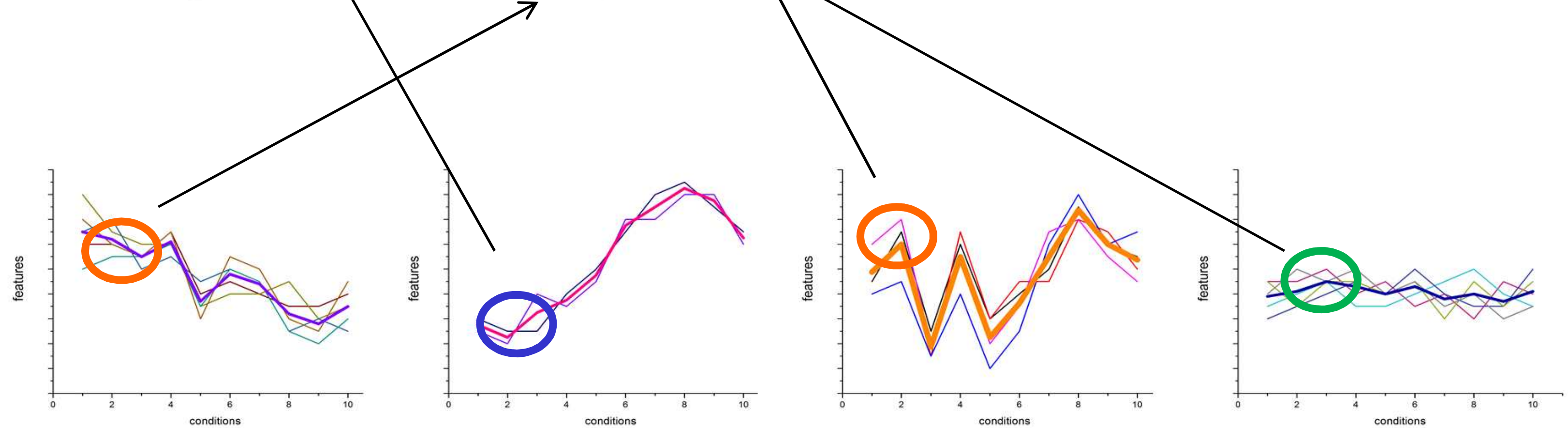


SOM image: clustering



Portrait:
distribution of all metagene expressions
in one sample

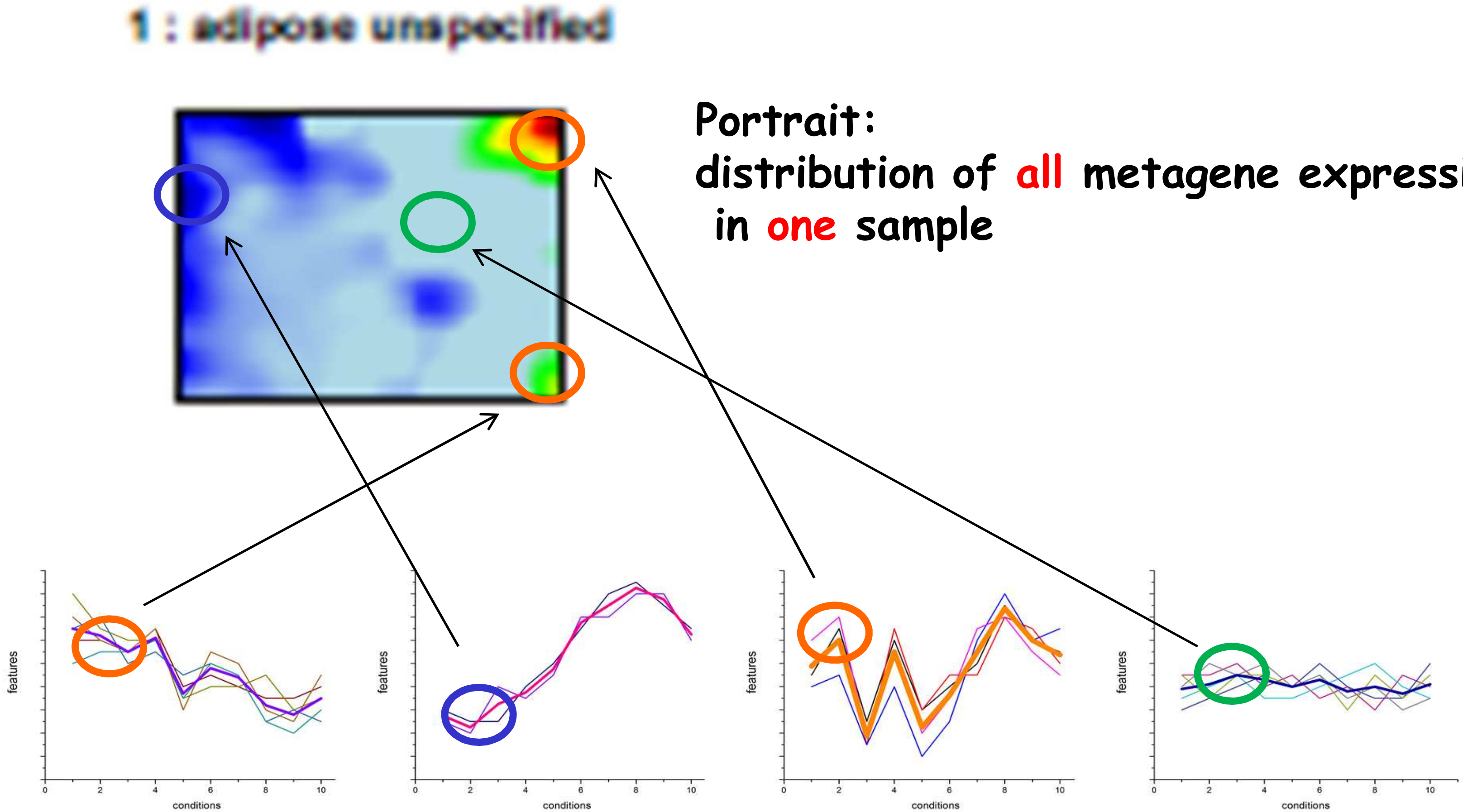
Receiving post offices
(letters come-in from Leipzig, Izmir…)



Metagene profile: distribution of one metagene in all samples

Sending post offices
(letters go-out to Leipzig, Novosibirsk…)

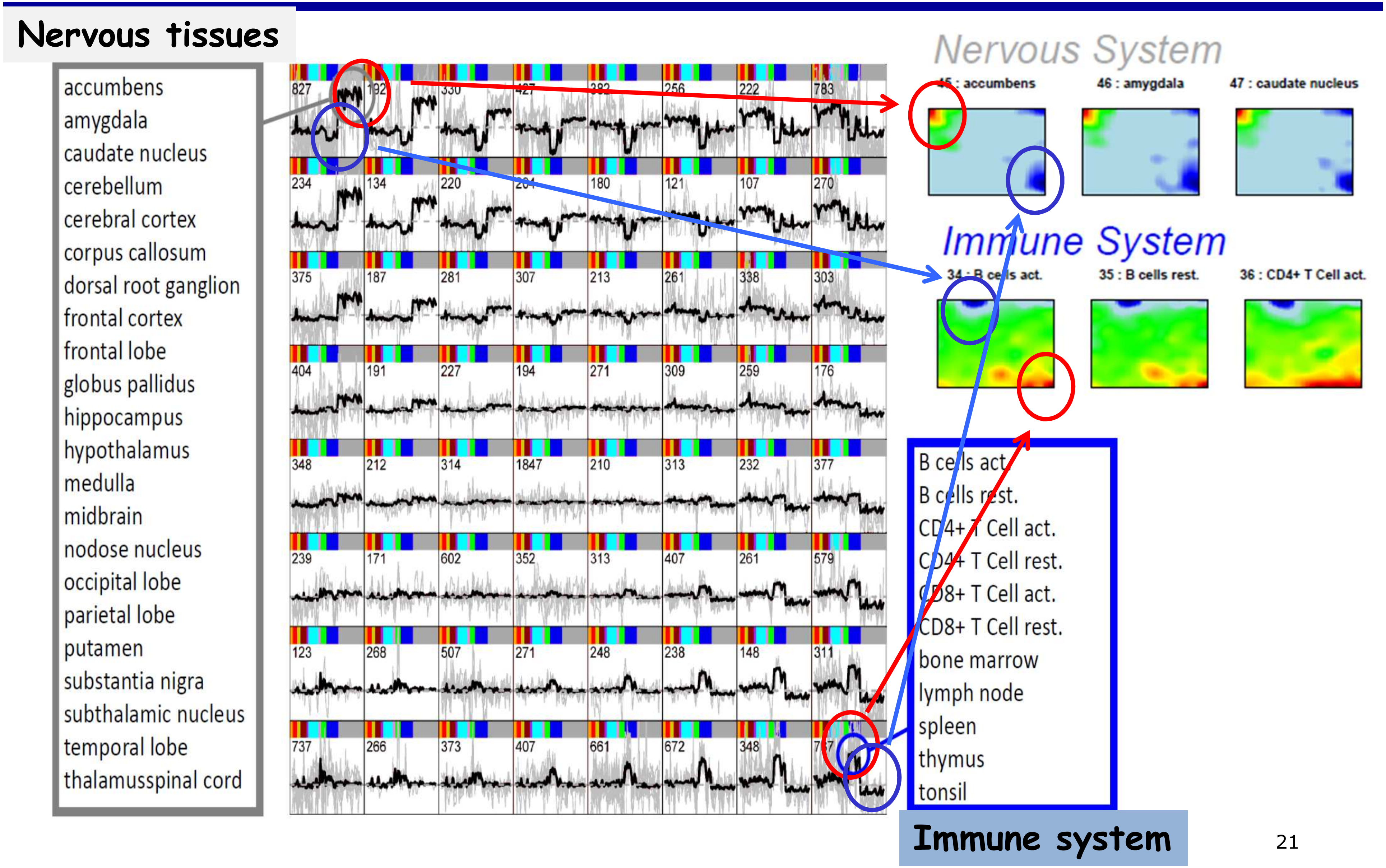
SOM image: clustering



Portrait:
distribution of all metagene expressions in one sample

Metagene profile: distribution of one metagene in all samples

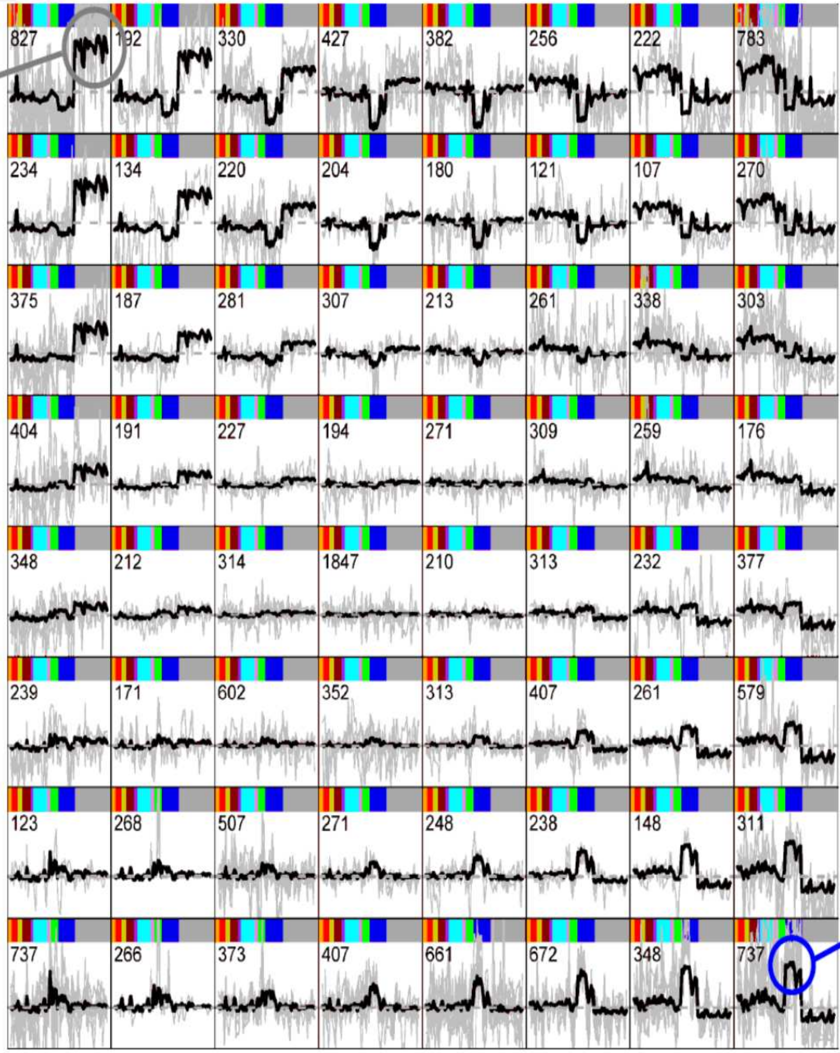
Profiling map: spots

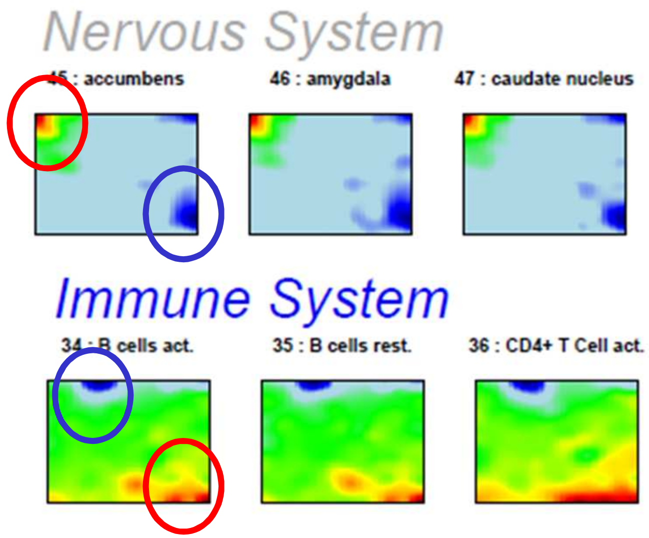


Profiling map: spots

Nervous tissues

- accumbens
- amygdala
- caudate nucleus
- cerebellum
- cerebral cortex
- corpus callosum
- dorsal root ganglion
- frontal cortex
- frontal lobe
- globus pallidus
- hippocampus
- hypothalamus
- medulla
- midbrain
- nodose nucleus
- occipital lobe
- parietal lobe
- putamen
- substantia nigra
- subthalamic nucleus
- temporal lobe
- thalamusspinal cord

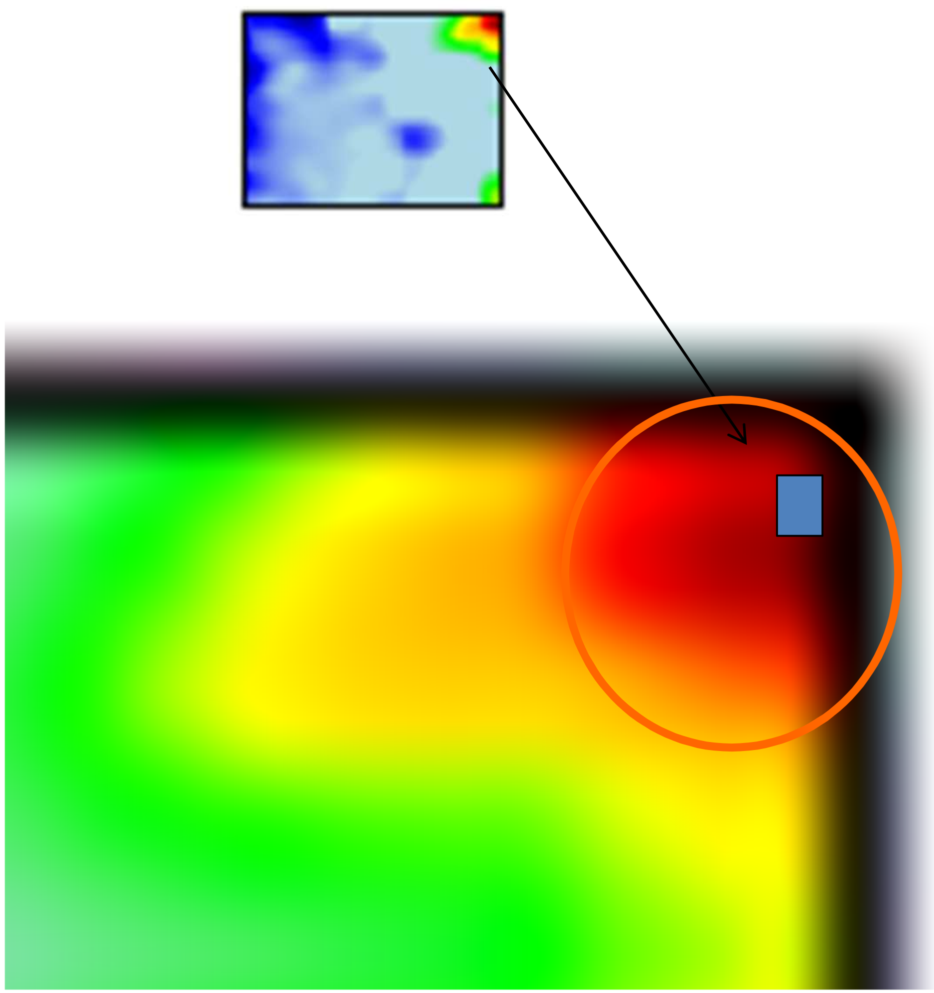




- B cells act.
- B cells rest.
- CD4+ T Cell act.
- CD4+ T Cell rest.
- CD8+ T Cell act.
- CD8+ T Cell rest.
- bone marrow
- lymph node
- spleen
- thymus
- tonsil

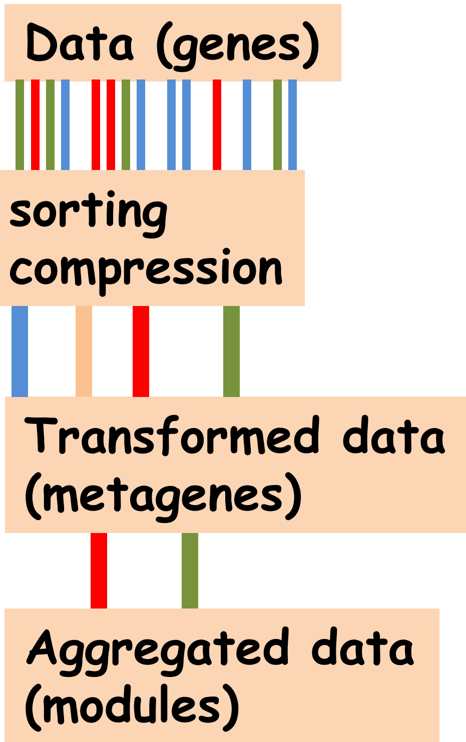
Immune system

SOM image: clustering

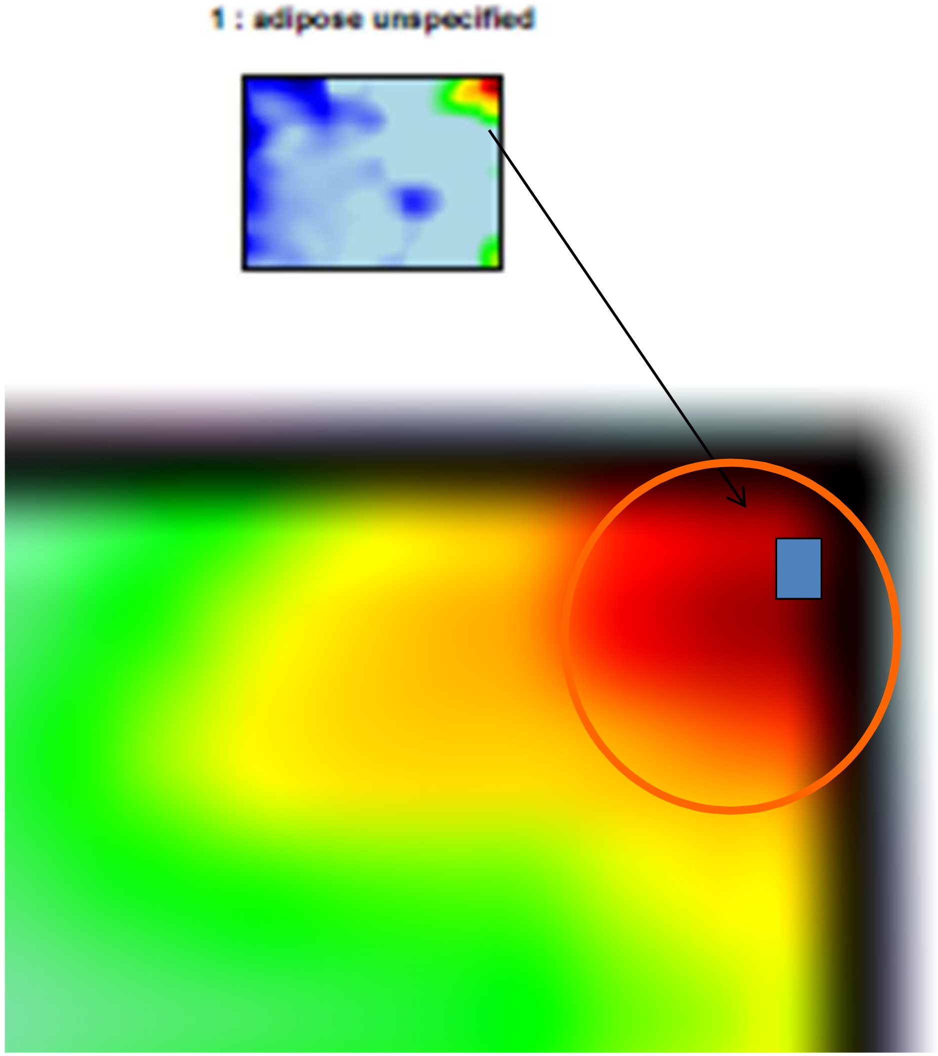


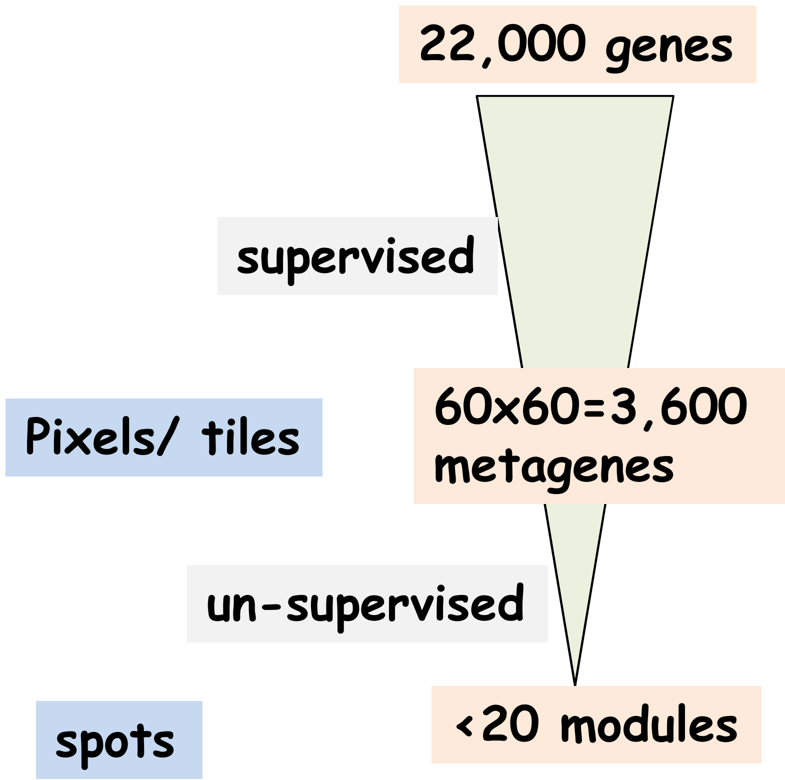
Pixels/ tiles

spots

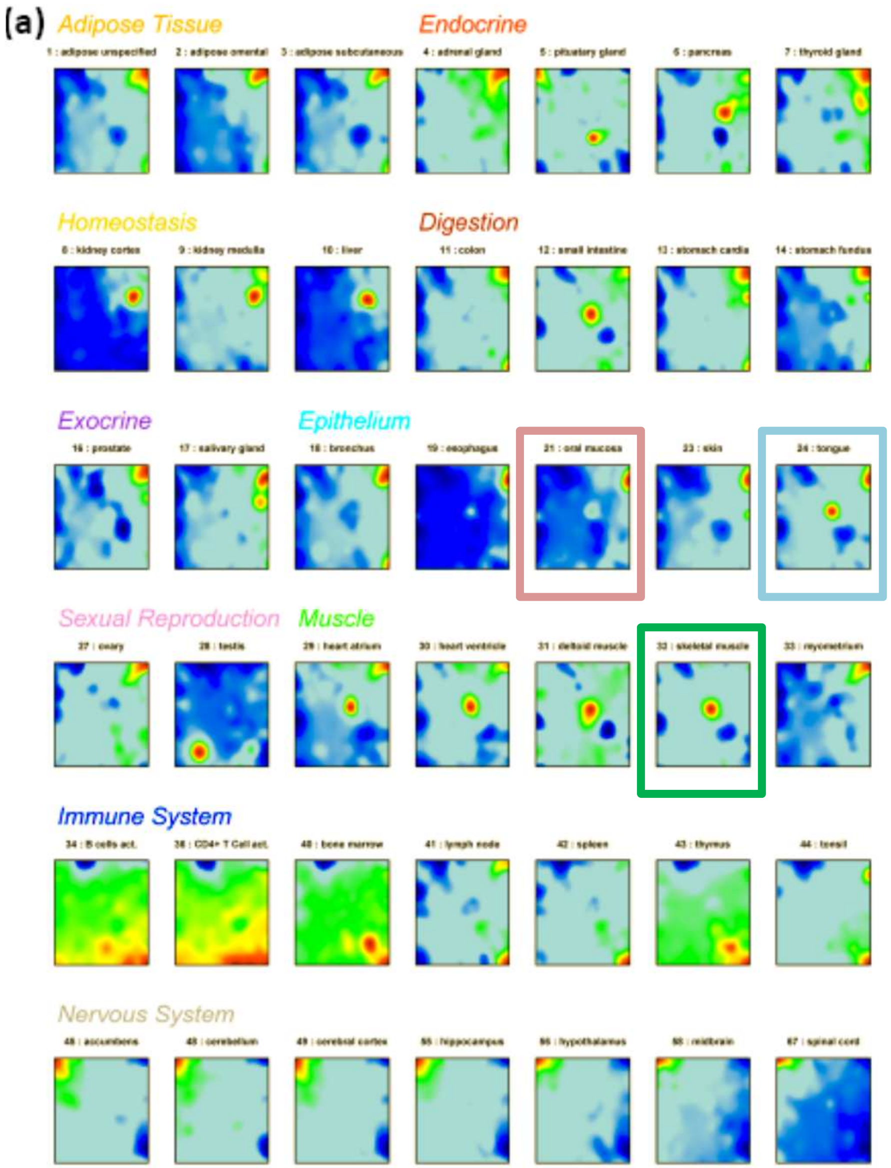


SOM image: spot clusters



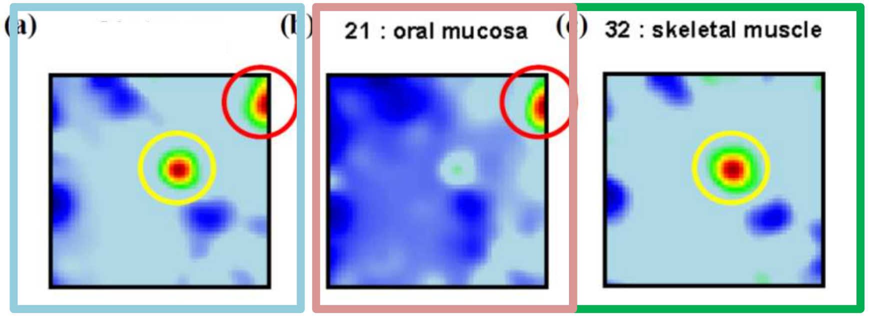


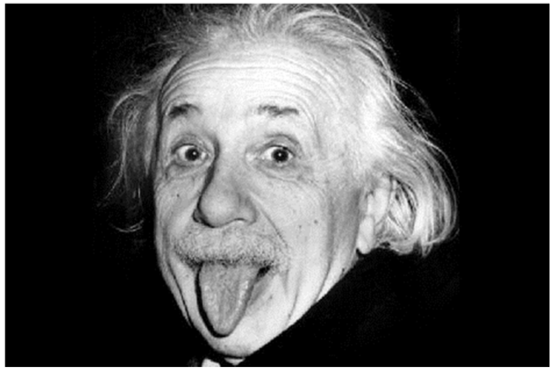
SOM: decomposition into parts



67 human tissues
9 categories
- heterogeneous expression
- well defined samples

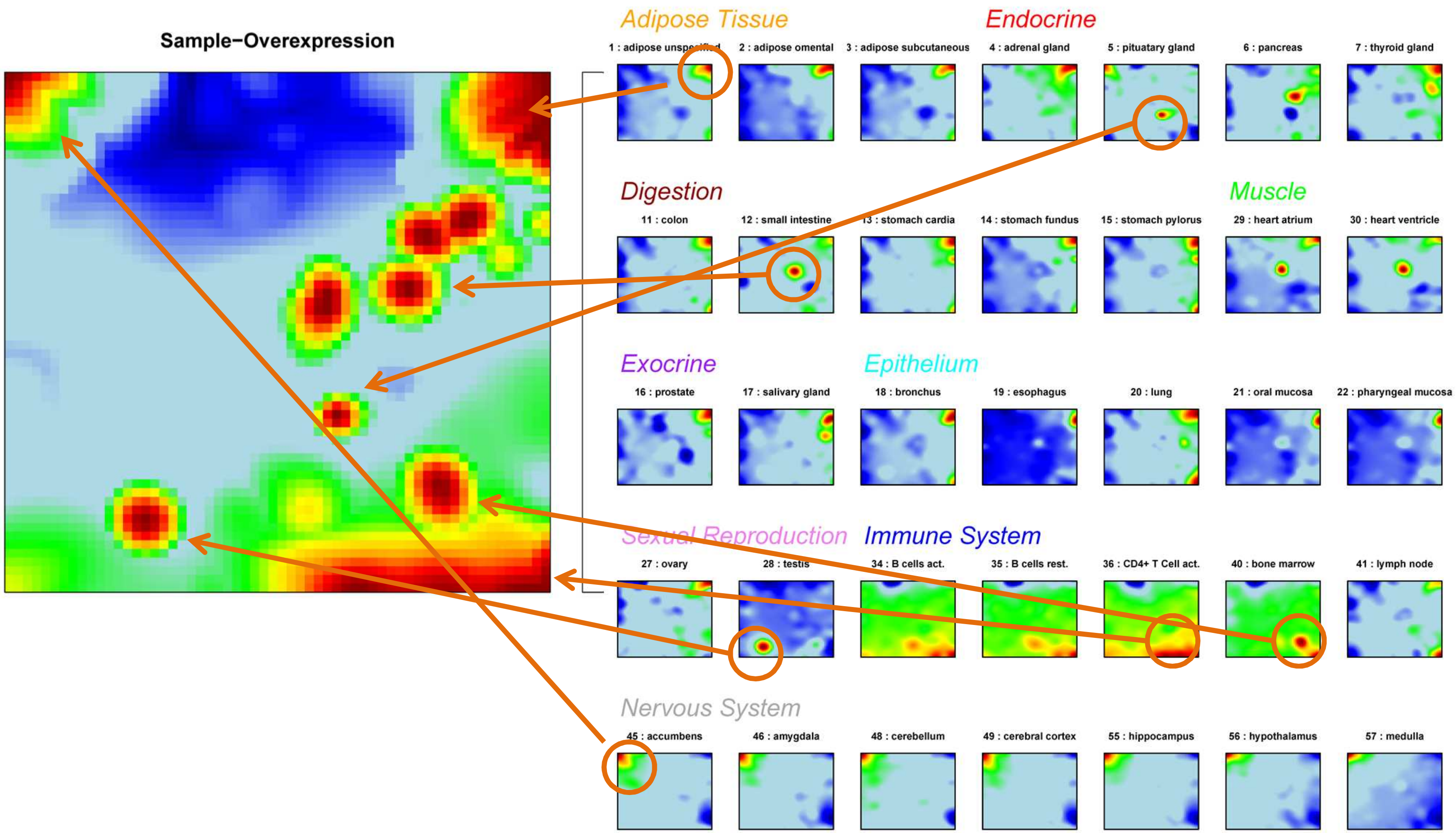
?



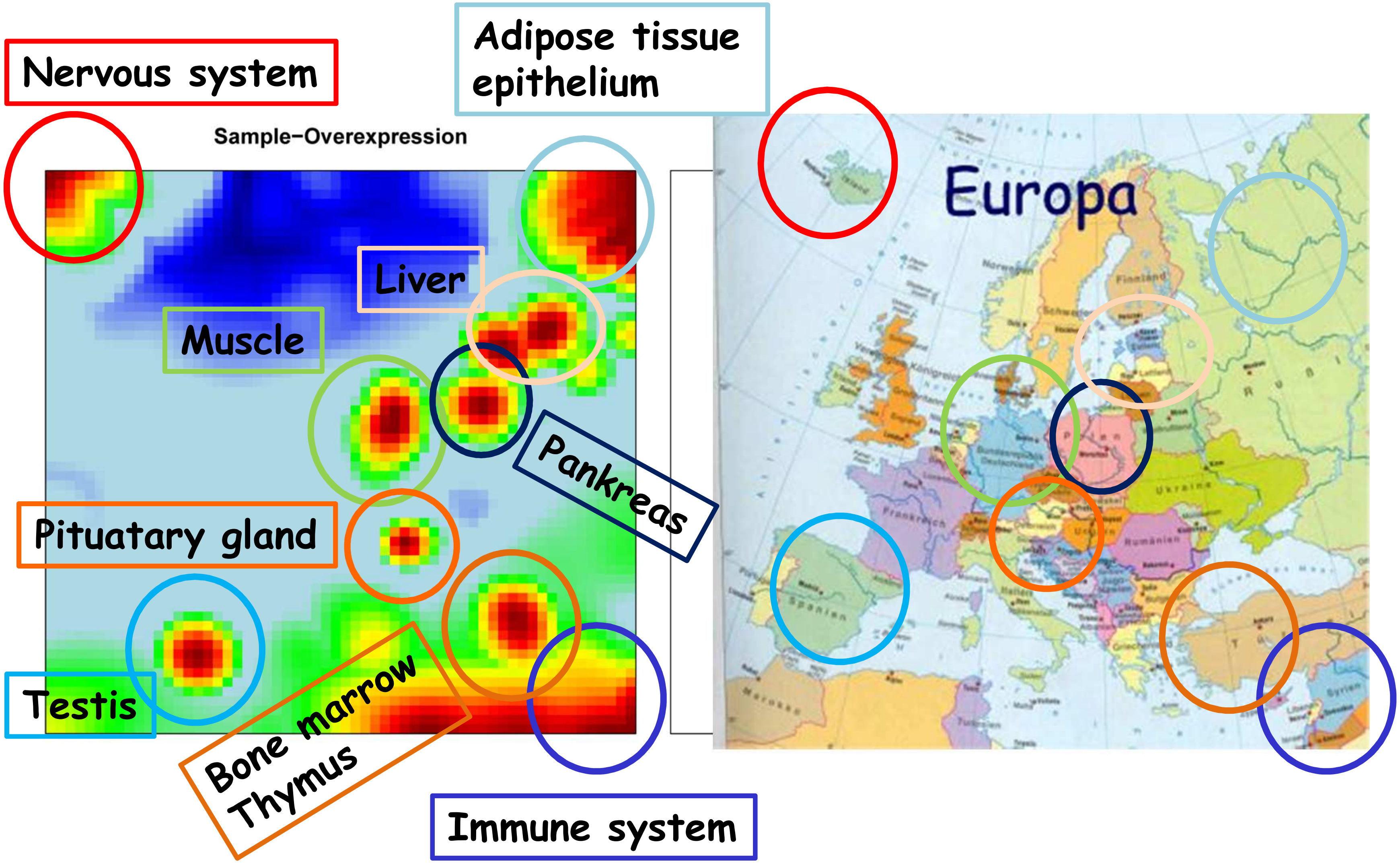


Spot summary

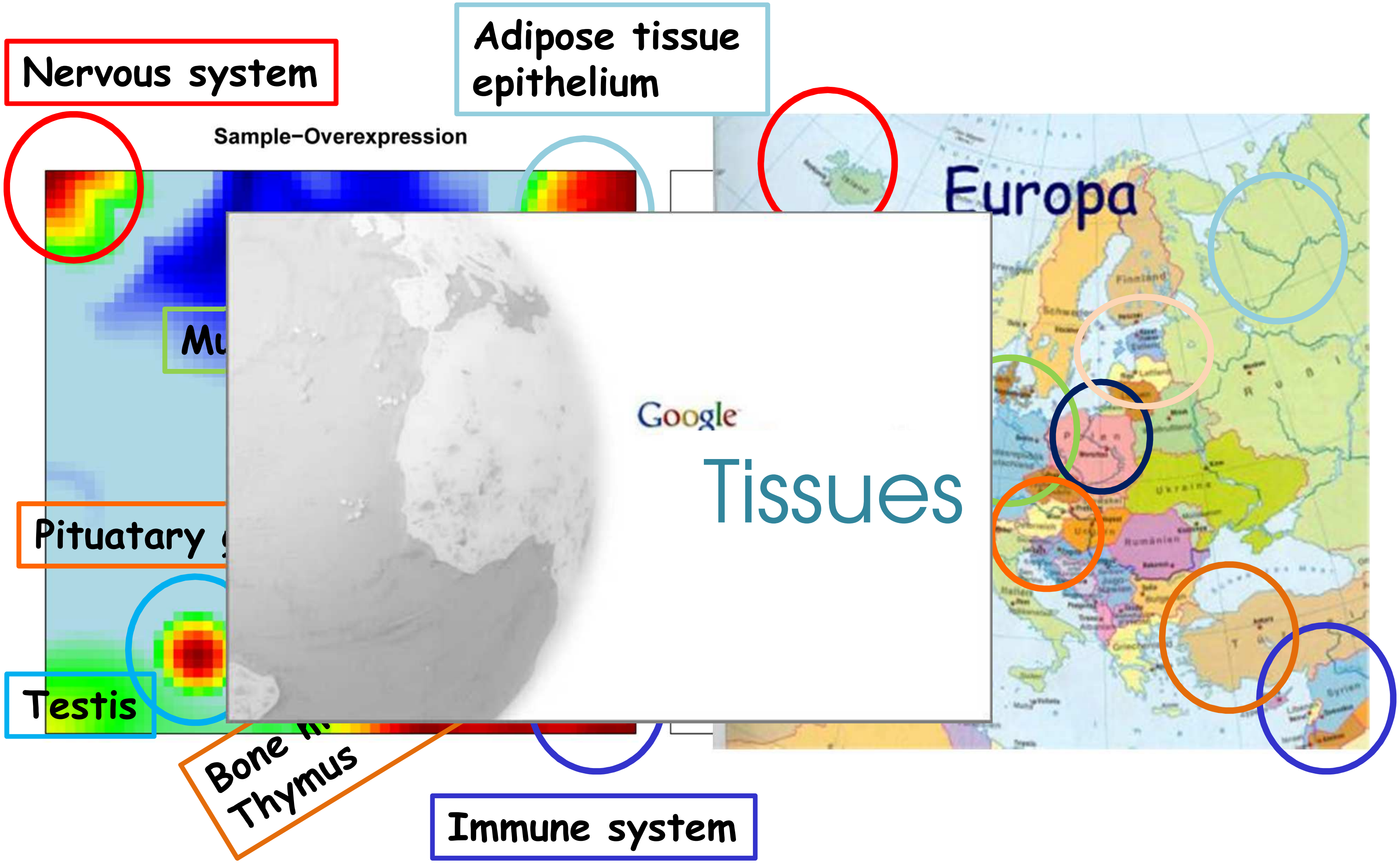
Overexpression summary map



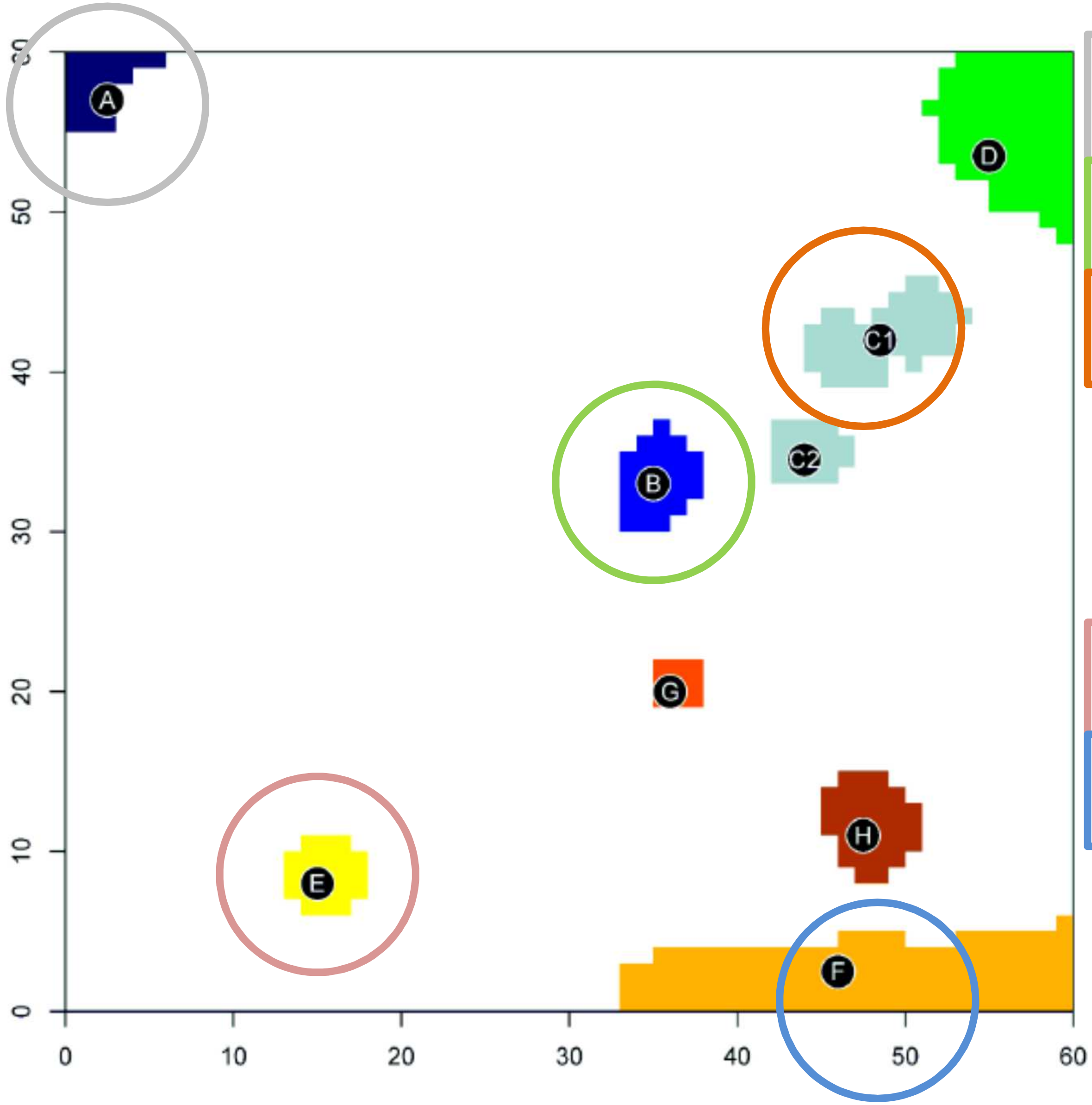
Expression landscape of human tissues



SOM: Feature map



GO-Gene set overrepresentation in the metagene spots



A	**Nervous system samples** *Nervous system development* *Synaptic transmission*
B	**Muscle samples** *Structural constituent of muscle* *System process*
C1	**Liver, kidney** *Carboxylic acid metabolic process* *Organic acid metabolic process*
C2	**Pancreas** *Carboxypeptidase activity* *Carboxylesterase activity*
D	**Adipose tissue, Epithelium samples** *Tissue development* *Ectoderm development*
E	**Testes** *Sexual reproduction* *Reproduction*
F	**Immune system samples** *Immune system process* *Immune response*
G	**Pituatary gland** *Hormone activity* *DNA fragmentation during apoptosis*
H	**Bone marrow, Thymus** *Cell cycle process* *Mitotic cell cycle*

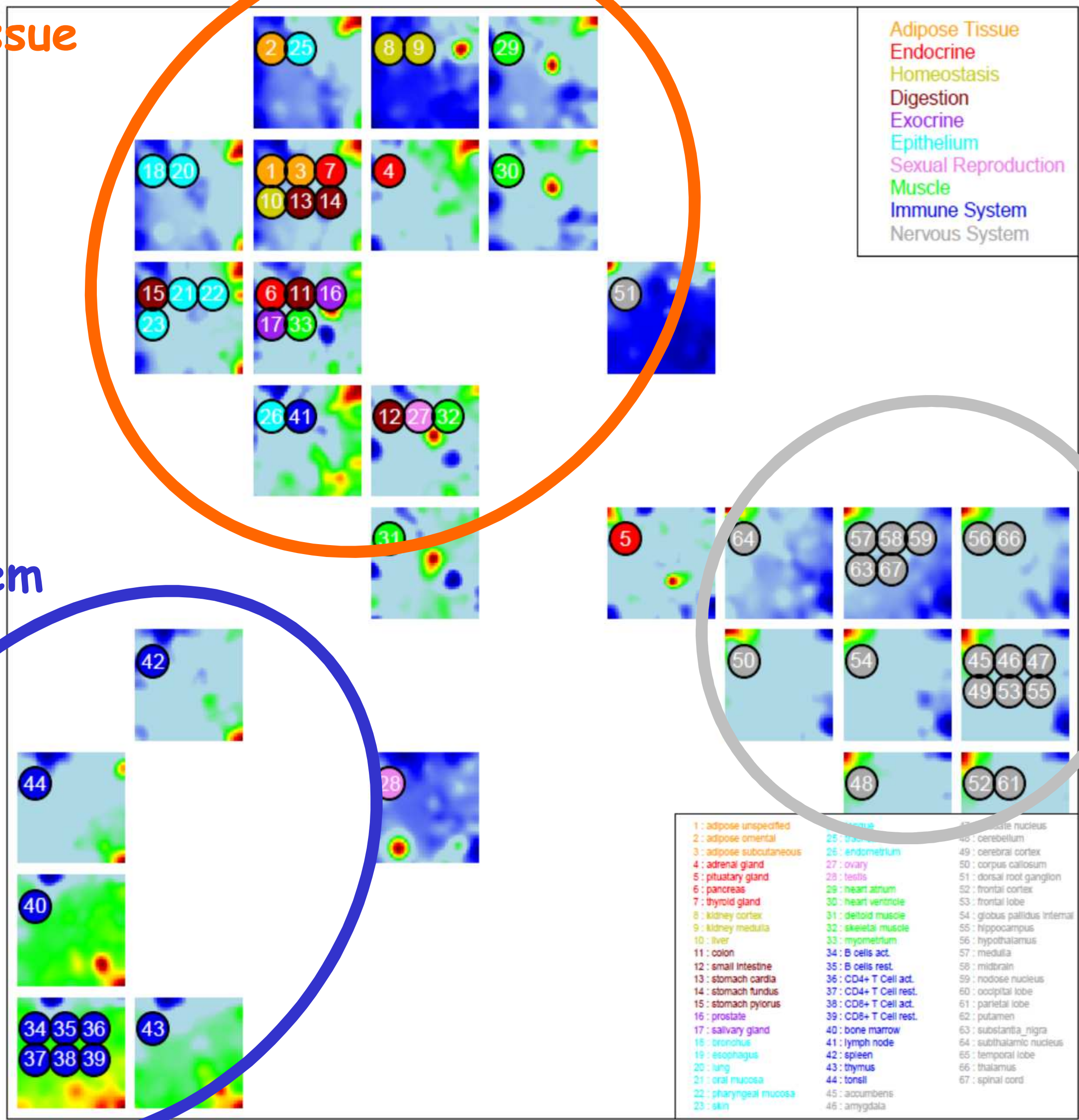
1500 gene sets tested (http://www.broadinstitute.org/gsea/i

Similarities between the tissues: 2nd level SOM

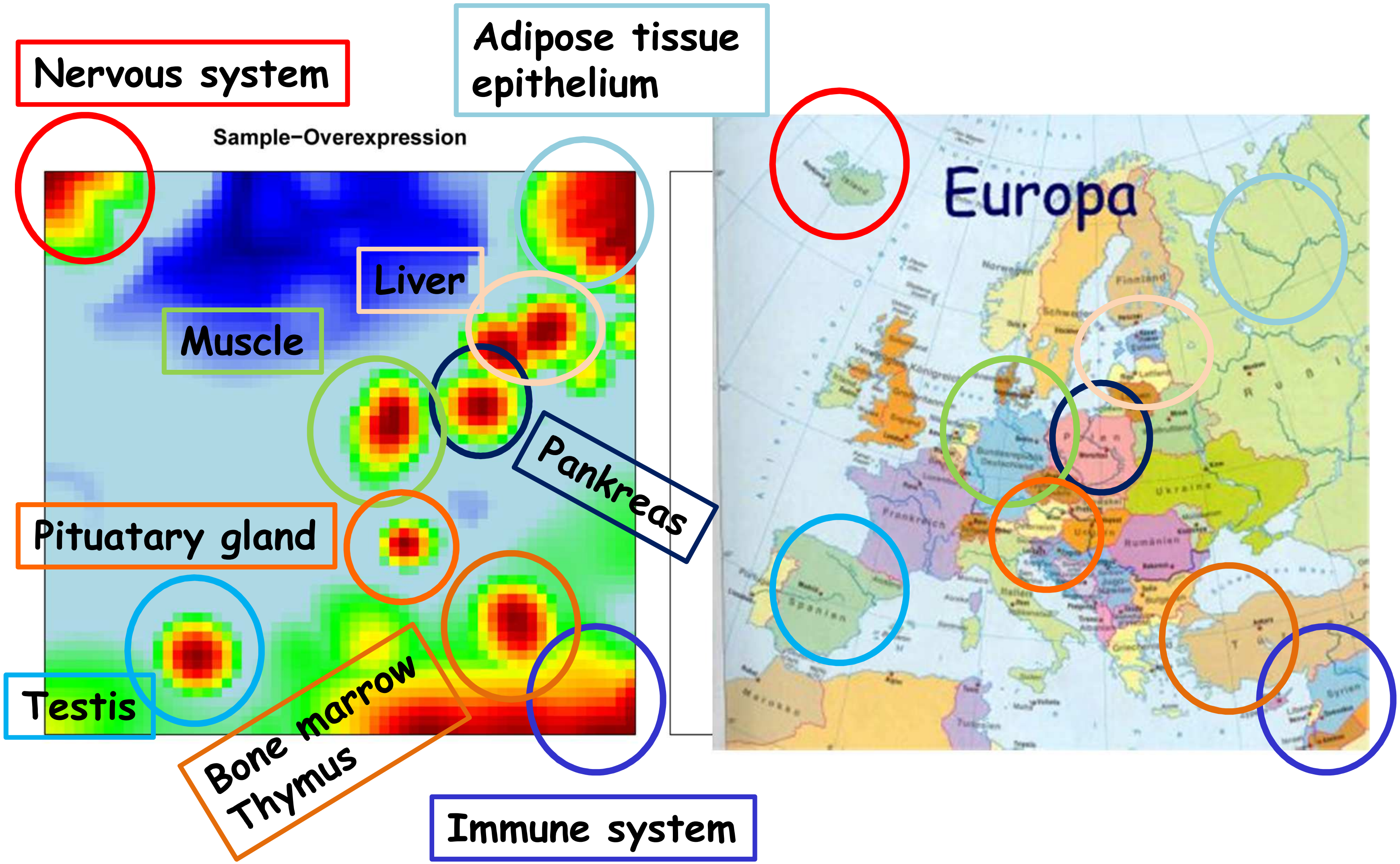
diverse tissue

immune system

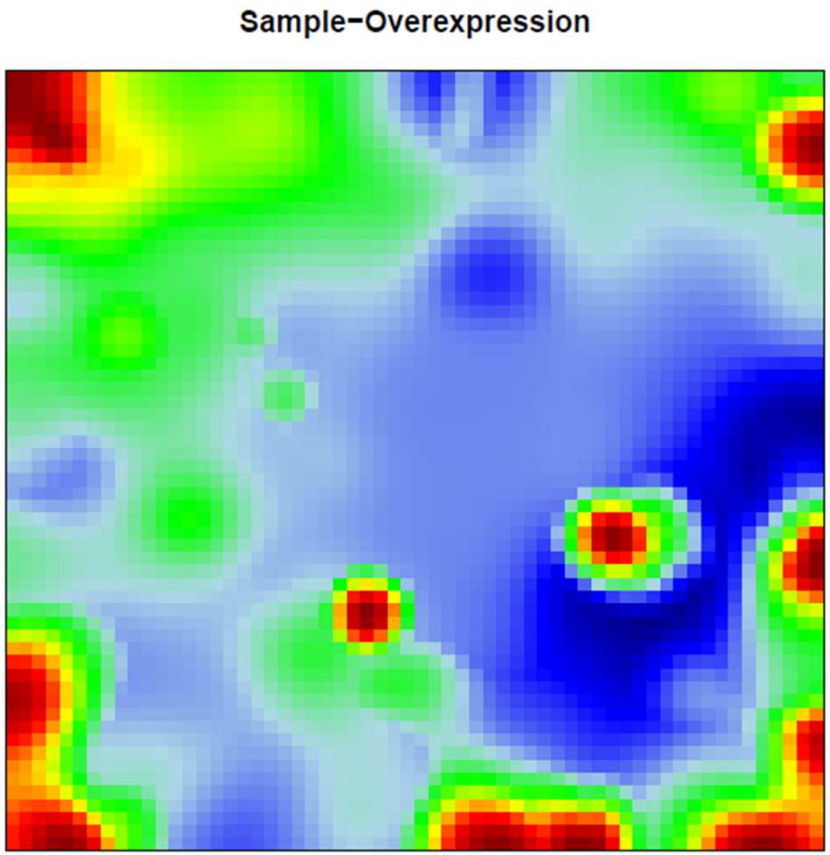
nervous system

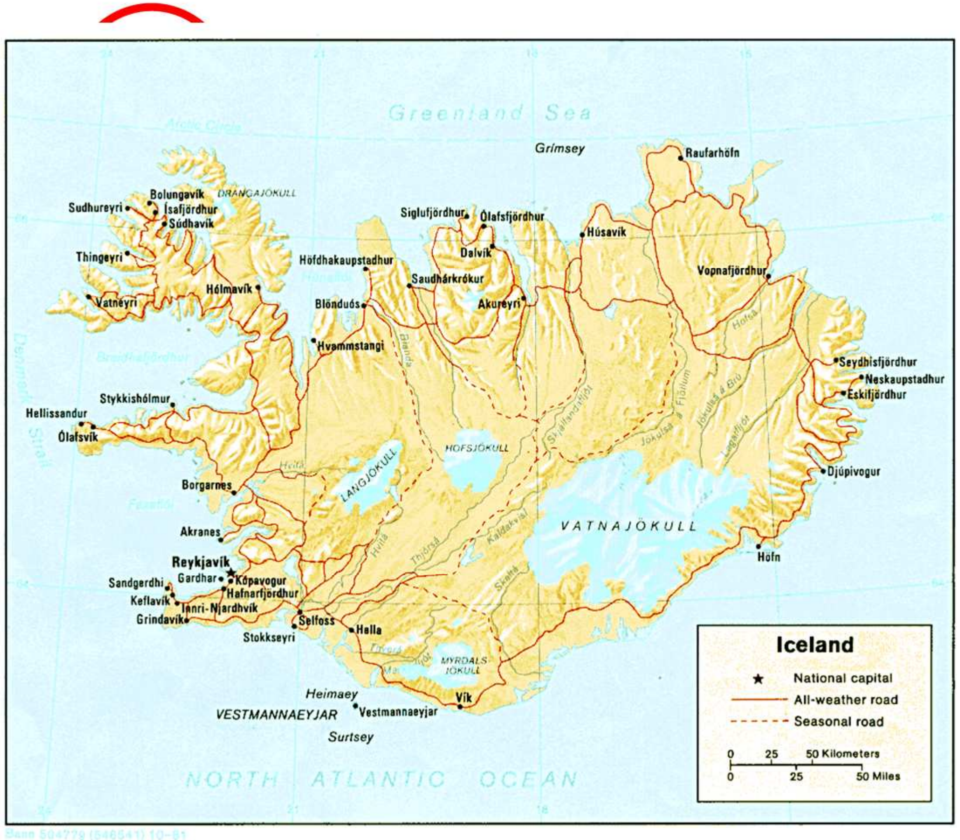


Tissue map


Nervous system
Adipose tissue
epithelium
Sample–Overexpression
Liver
Muscle
Pankreas
Pituatary gland
Testis
Bone marrow
Thymus
Immune system
Europa

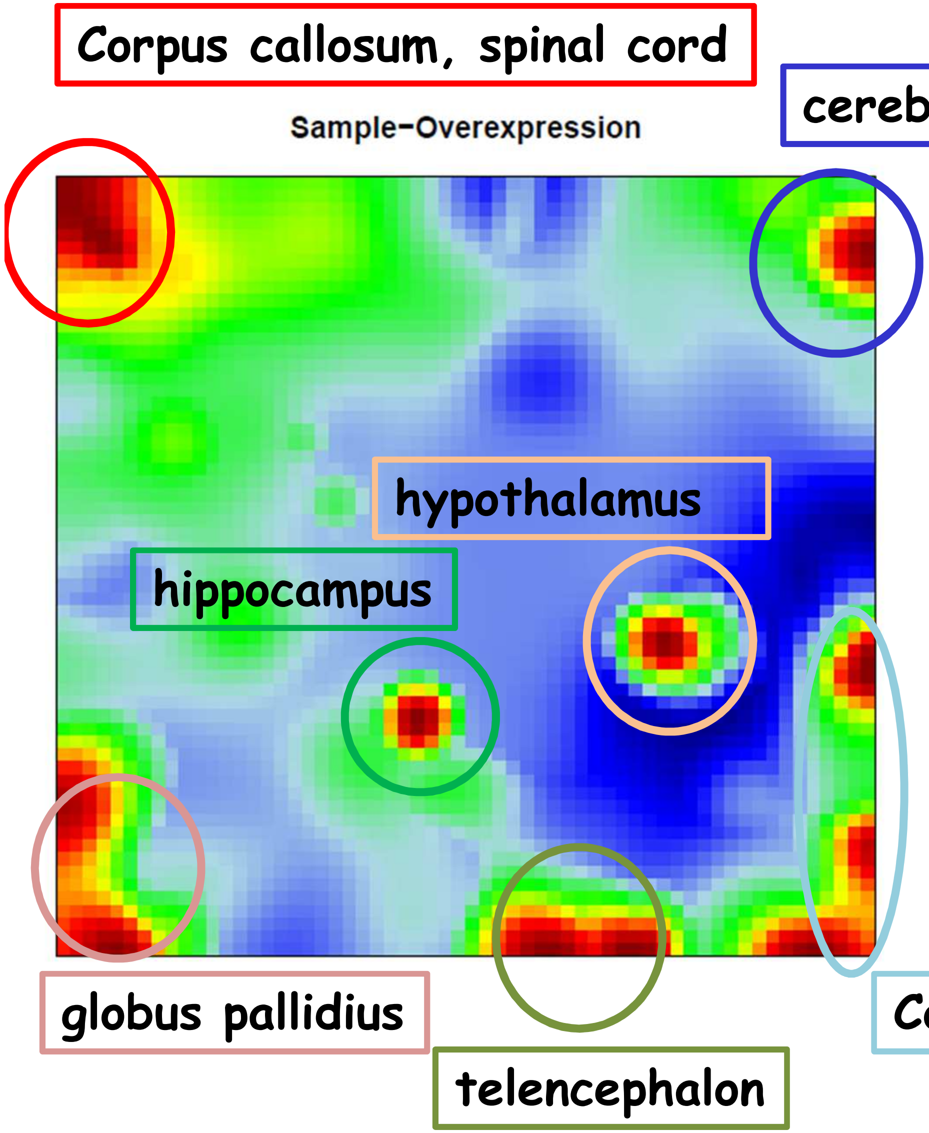
Zoom in: nervous tissues

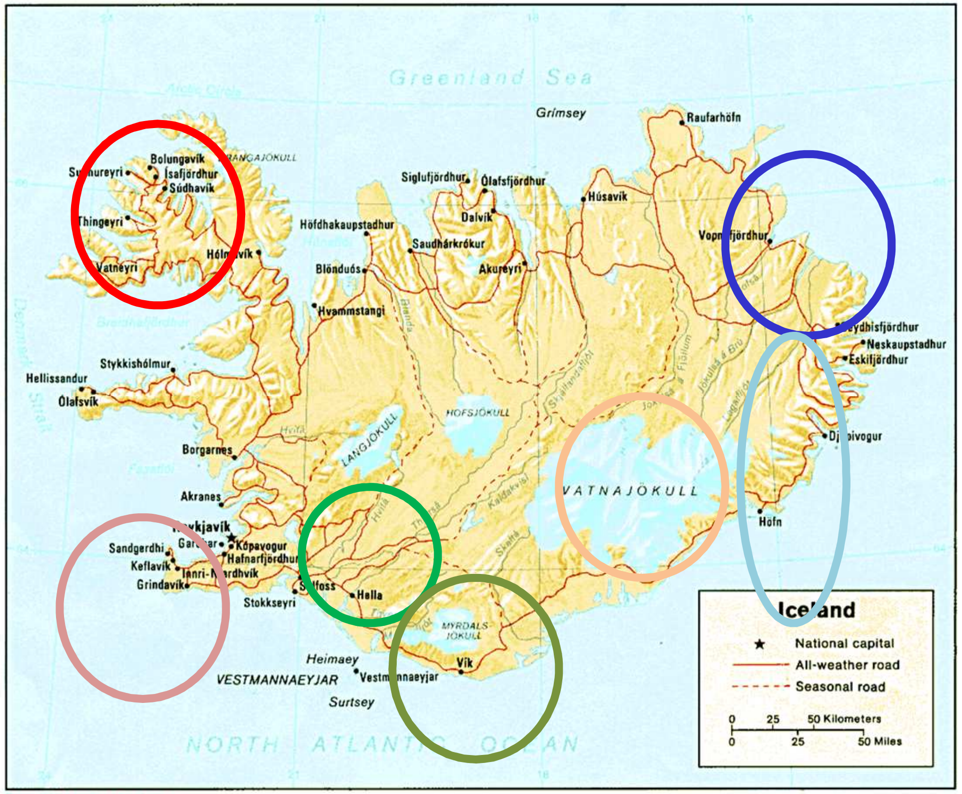




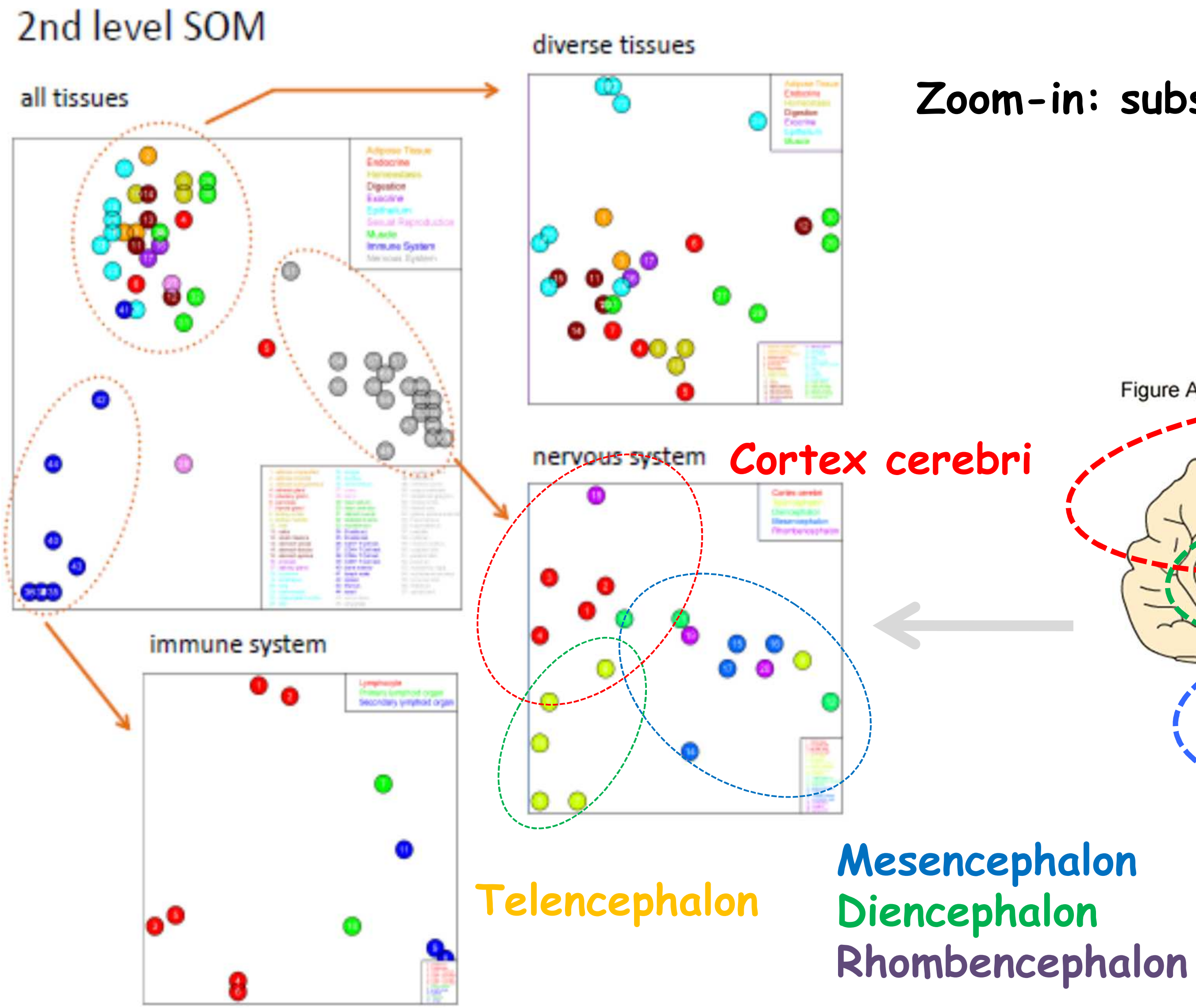
…train a new SOM with a subset of tissues (e.g. nervous tissues)

Map of nervous tissues

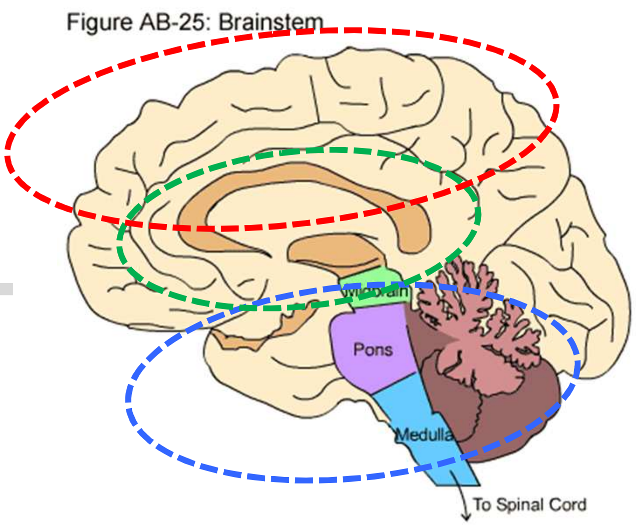




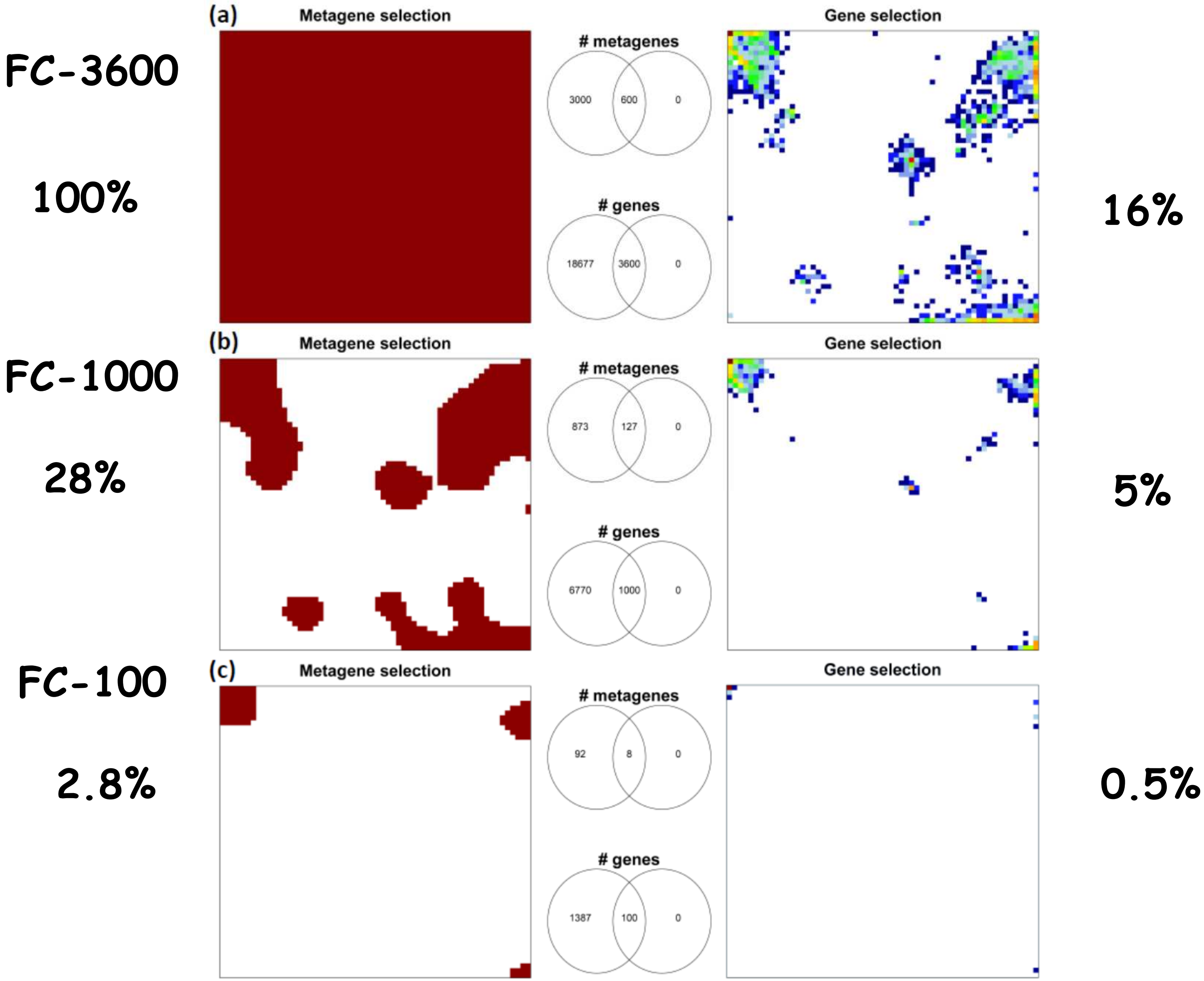
Zoom-in: Landscaping using 2nd level SOM



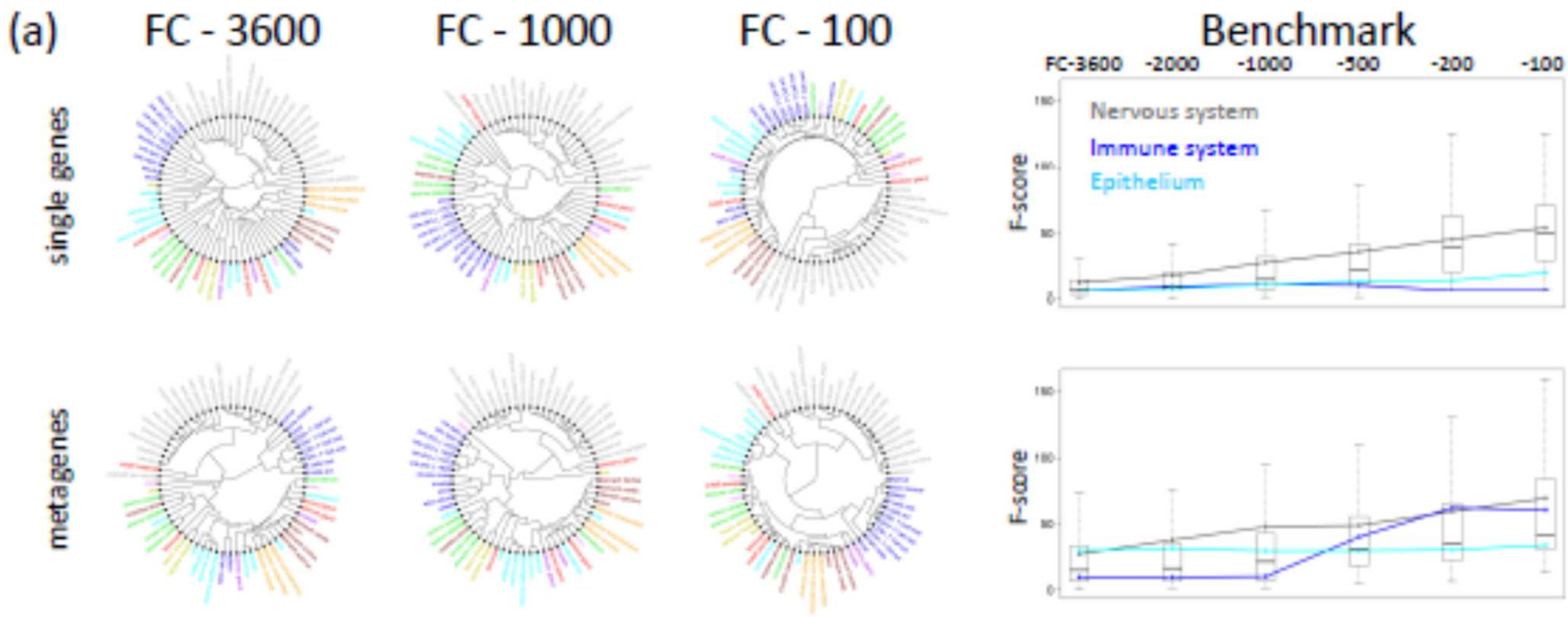
Zoom-in: subsets of all tissues



Metagene-vs-single gene analysis: filtering

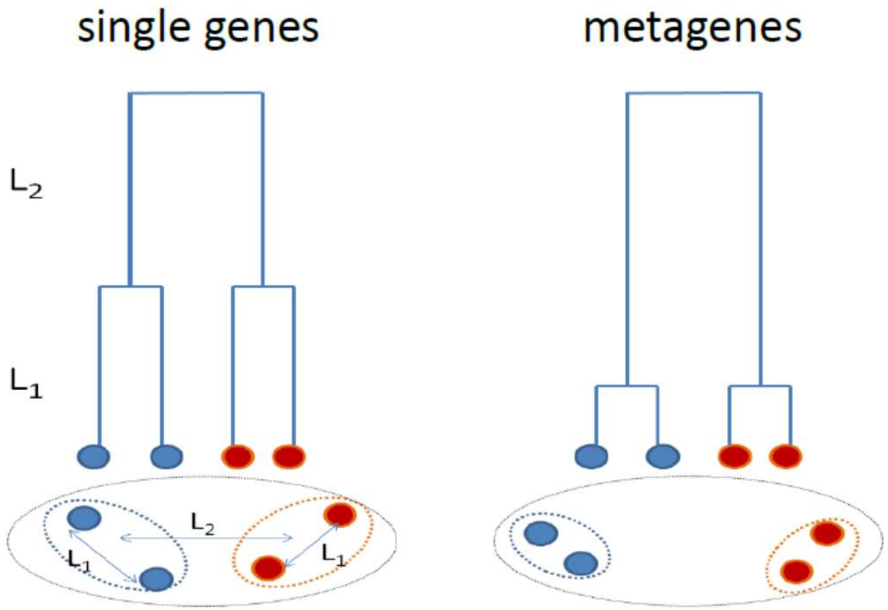


Metagene-vs-single genes: MG provide more compact cluster

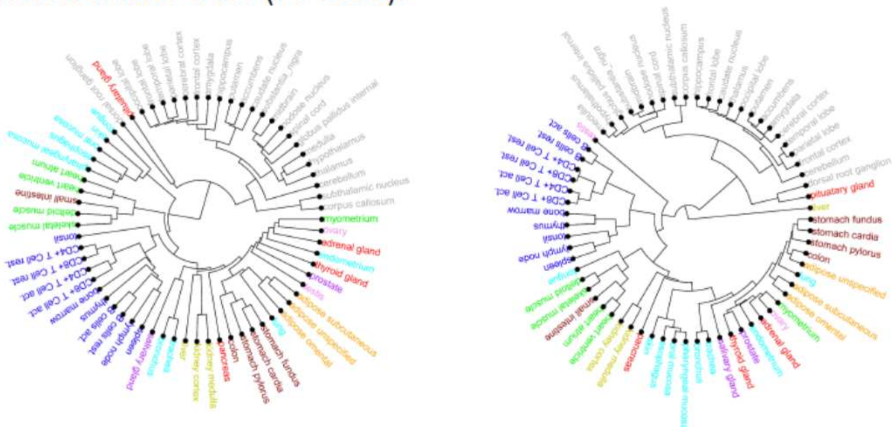


F-score:
Inter – to – intra cluster ratio of distances

Metagene-vs-single gene analysis: clusters are more compact



Radial cluster trees (FC-1000):

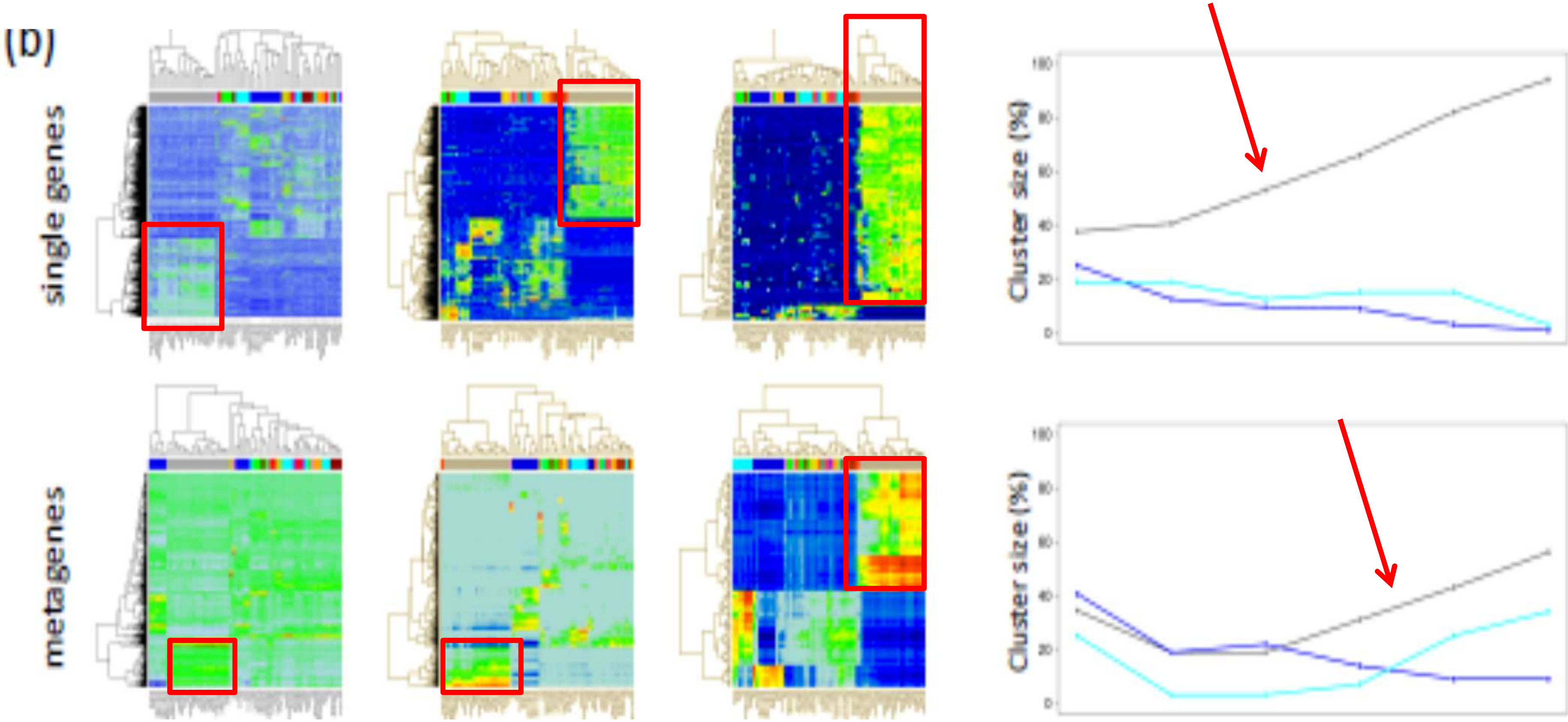


…because metagenes are representative and less noisy

Metagene-vs-single genes: MGs are more representative

FC-3600 FC-1000 FC-100

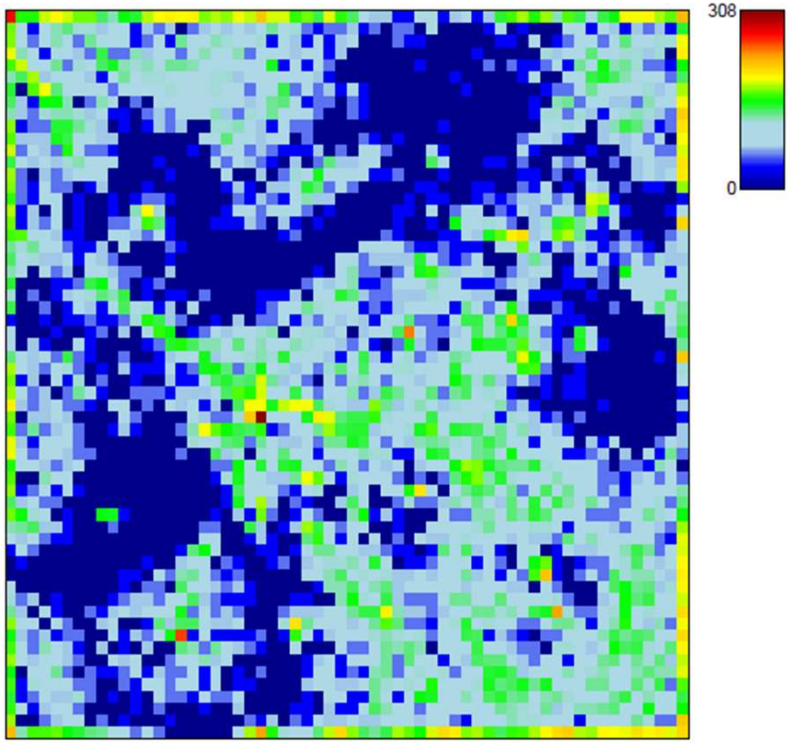
%genes from nervous tissues



FC-3600 FC-1000 FC-100

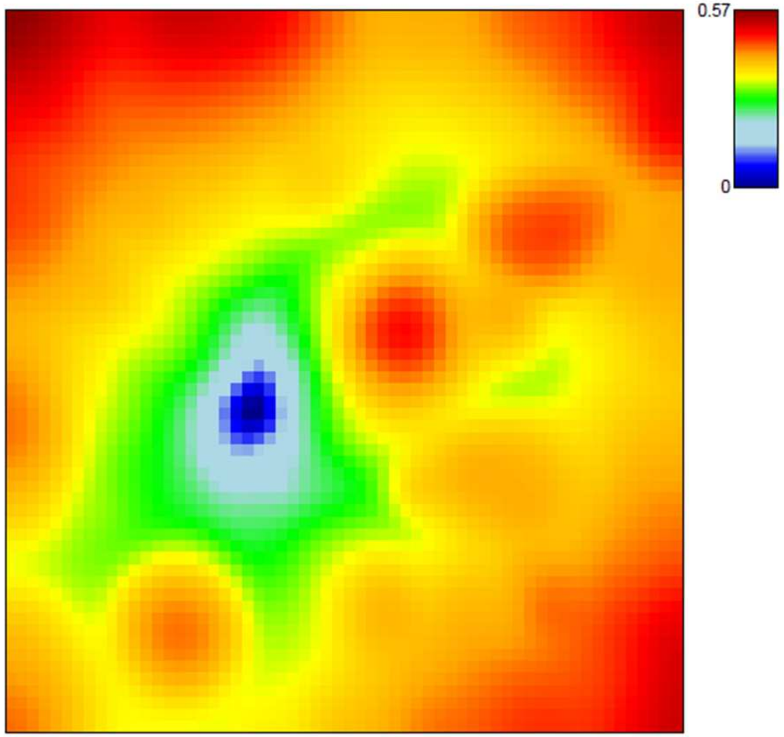
…because metagenes down-weight redundant information

Population Map



log (# genes in metagene)

Metagene Variance Map

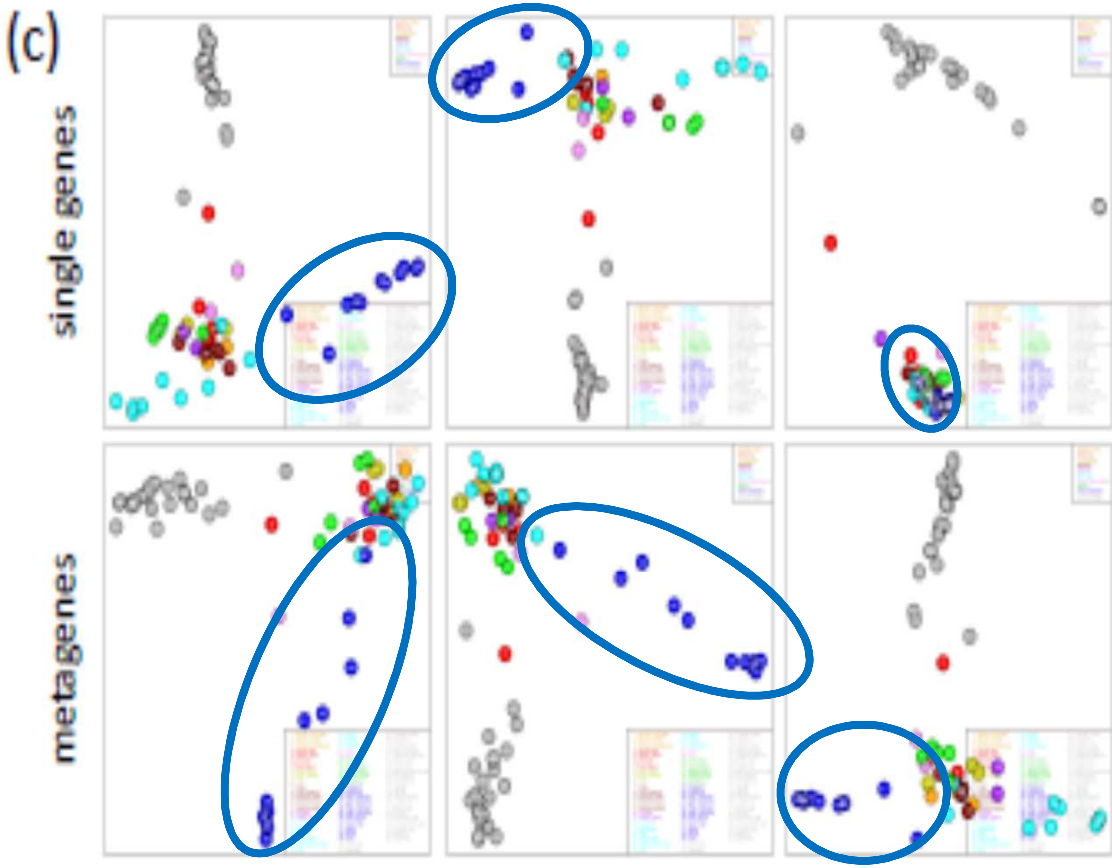


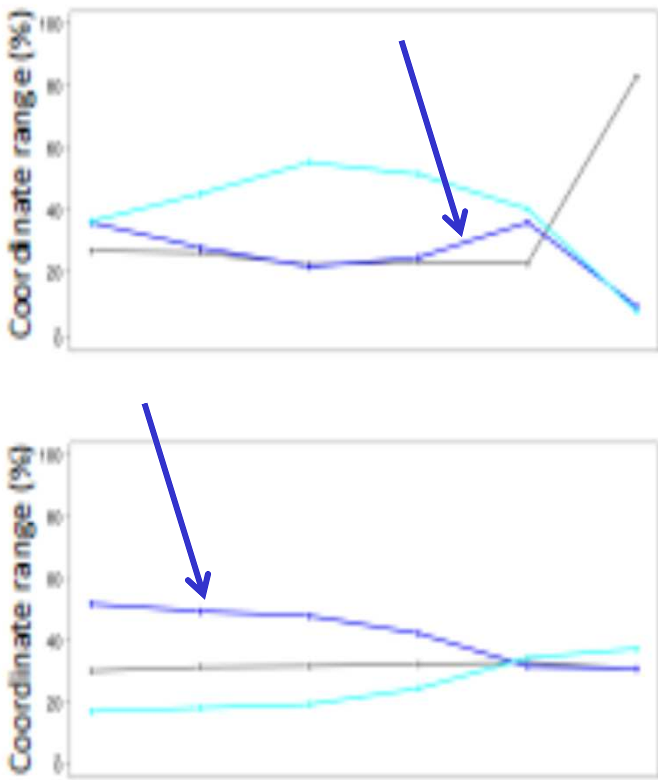
log (metagene variance)

Metagene-vs-single genes: MGs provide better resolution

FC-3600 FC-1000 FC-100

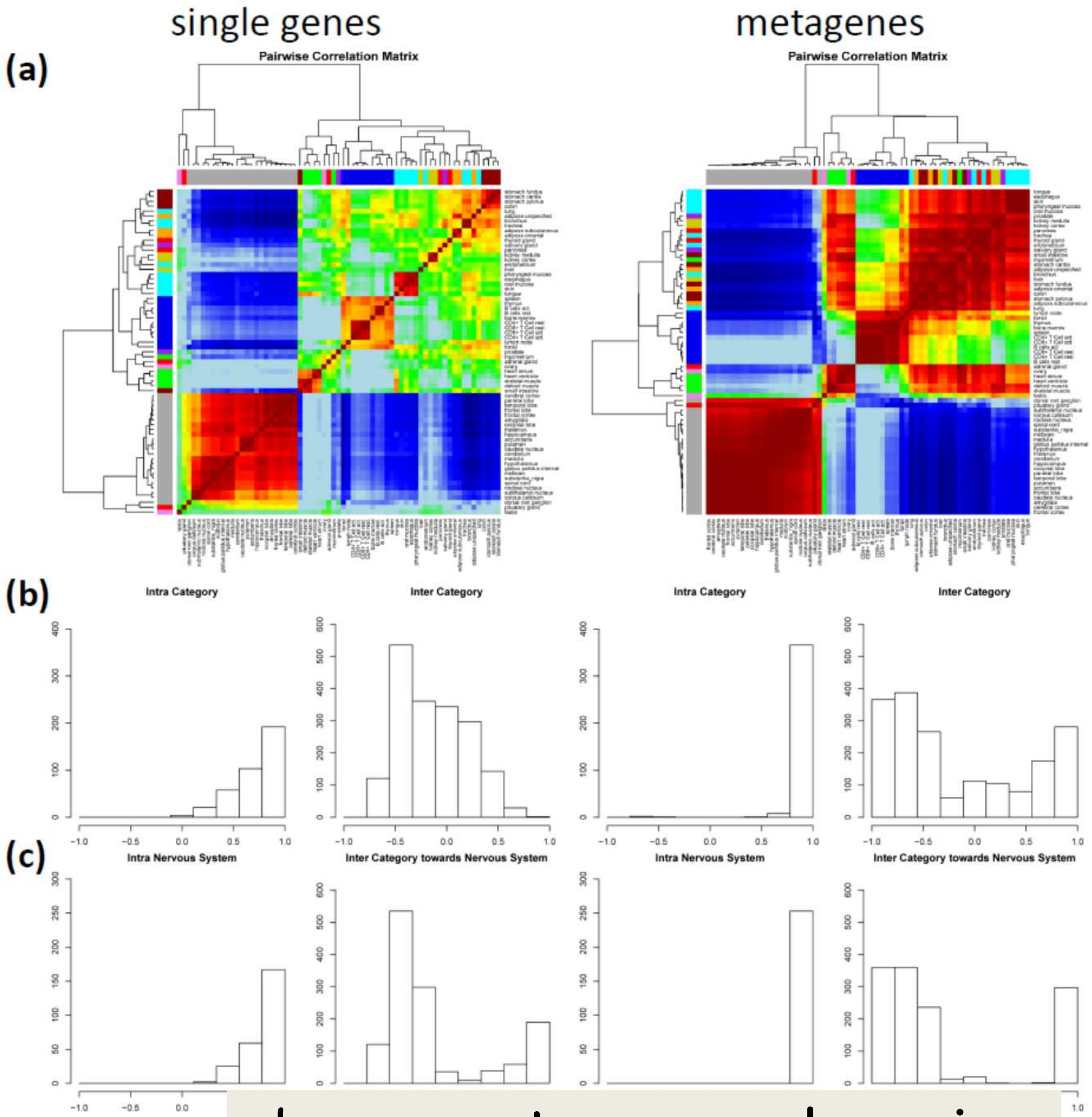
%cluster size of immune system tissues





FC-3600 FC-1000 FC-100

Metagene-vs-single genes: MG provide better correlations



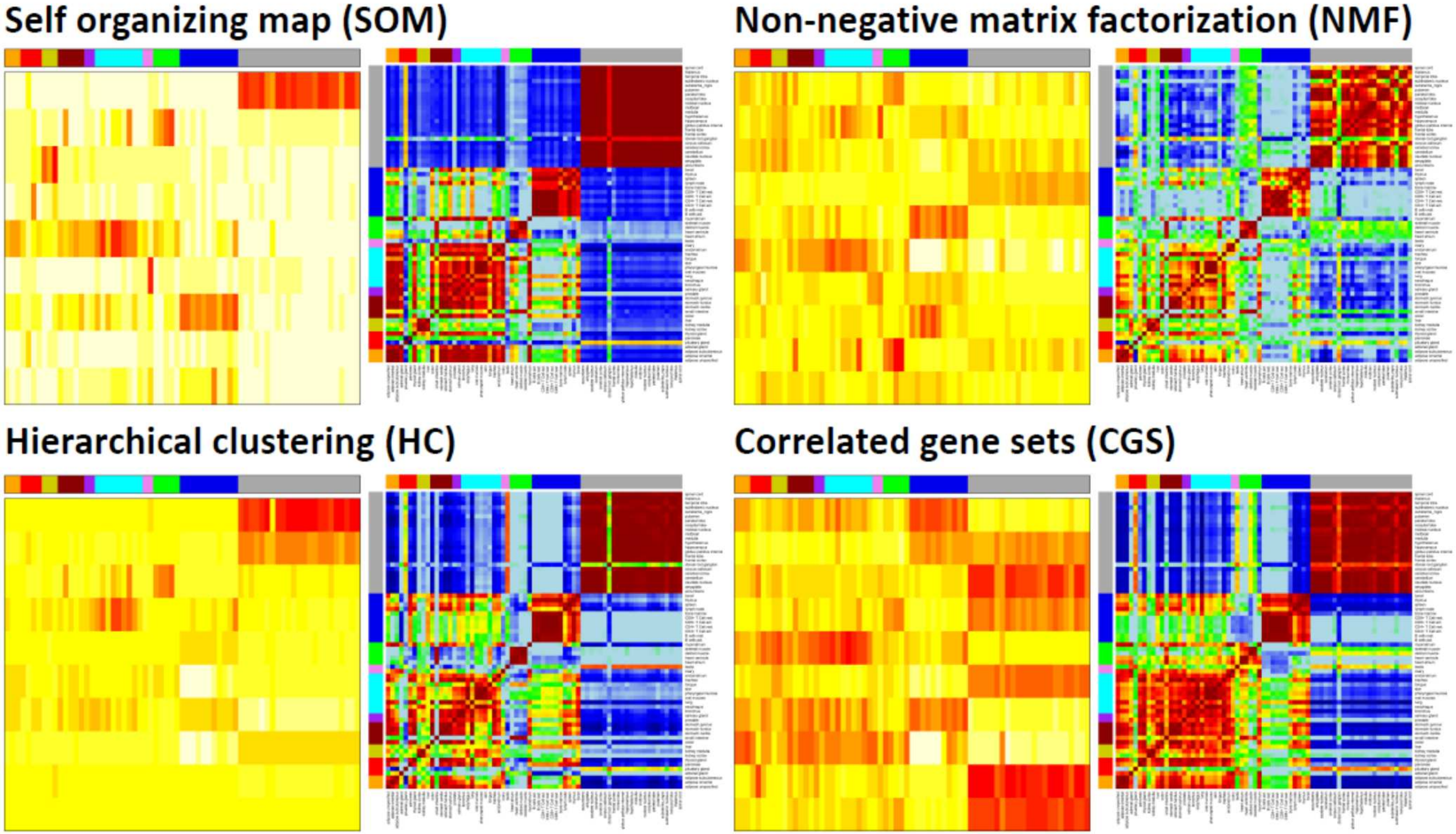
Intra-category

Inter-category

Only-nervous t.

…because metagenes are less noisy

Comparison of clustering methods: SOM identifies class-related features

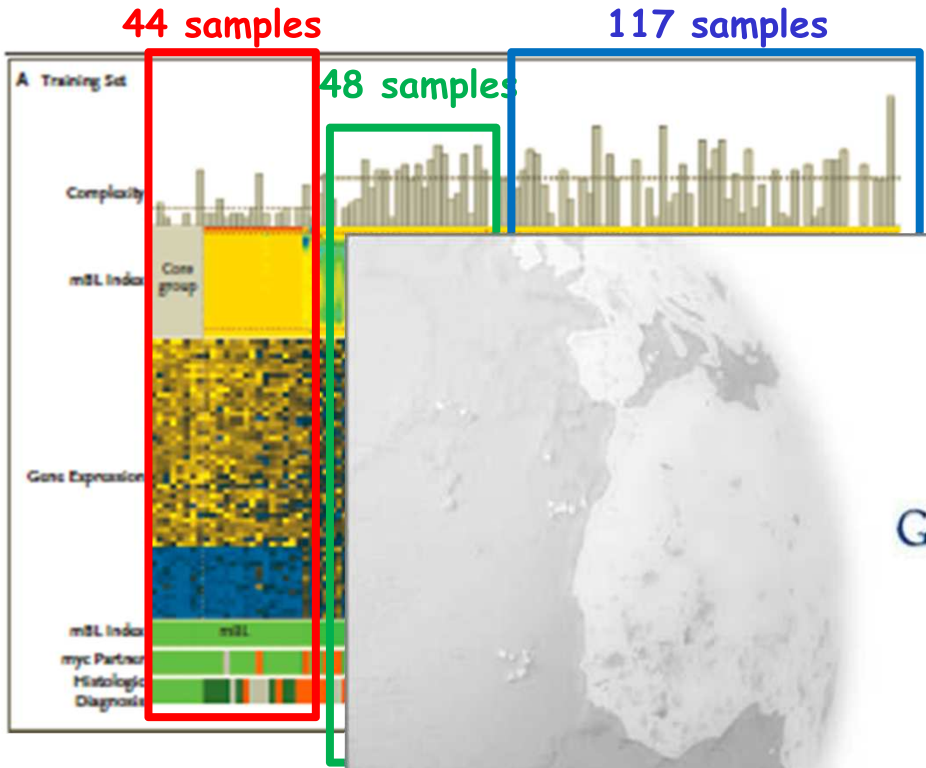


…because SOM uses distance similarity AND flexible projection into visualization space

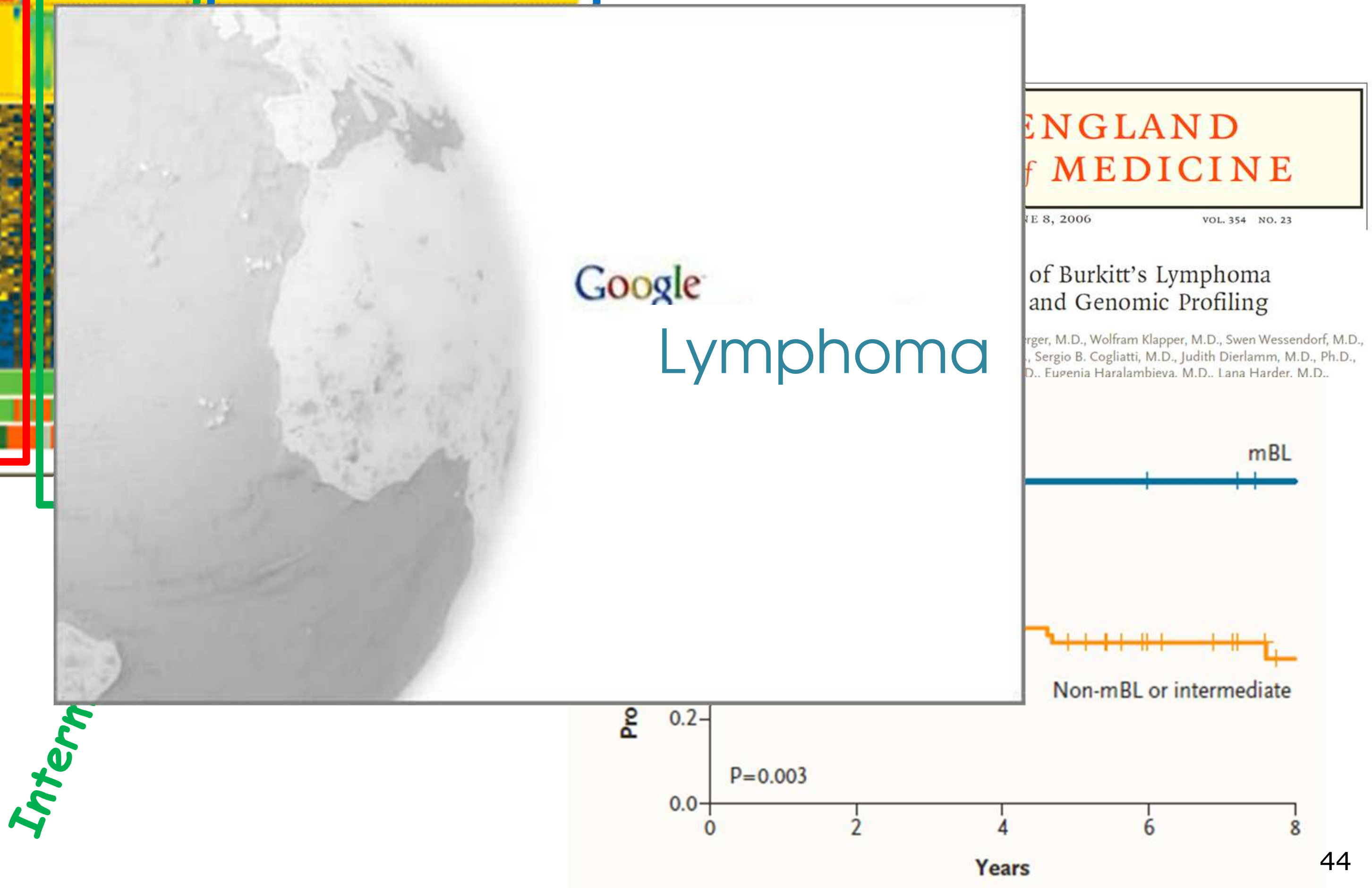
SOM: Basal properties

1. Portrays each sample
2. Dimension reduction (meta-features, prototypes) without loss of information (all single features are still present in the clusters)
3. Highly intuitive (spot pattern)
4. Interpretable (concepts of function, GSEA)
5. Scalable (zoom-in)
6. Gene centered analysis (GSEA)
7. Sample centered analysis (similarity analysis)
8. Metagenes are usually better than single genes: more representative and less noisy

Starting point: DLBCL-subtypes

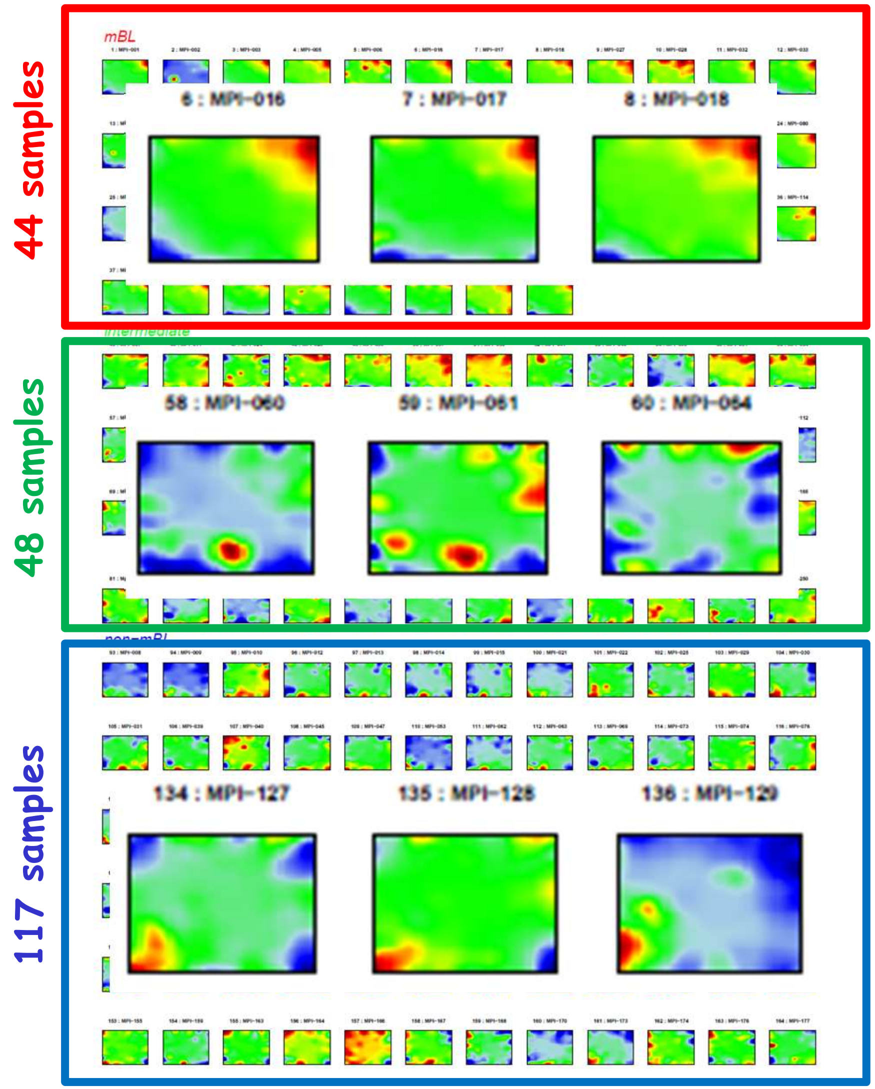


Different ‘molecular’ diseases of Diffuse large B-cell lymphoma (DLBCL)



molecular Burkitt's lymphoma (mBL)

Disentangling lymphoma subtypes



Log FC scale

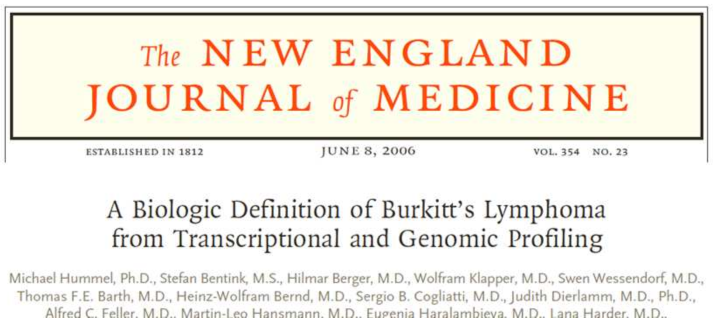
Different 'molecular' diseases

molecular Burkitt's lymphoma (mBL)

Diffuse large B-cell lymphoma (DLBCL)

Intermediate

Non-molecular BL: n-mBL



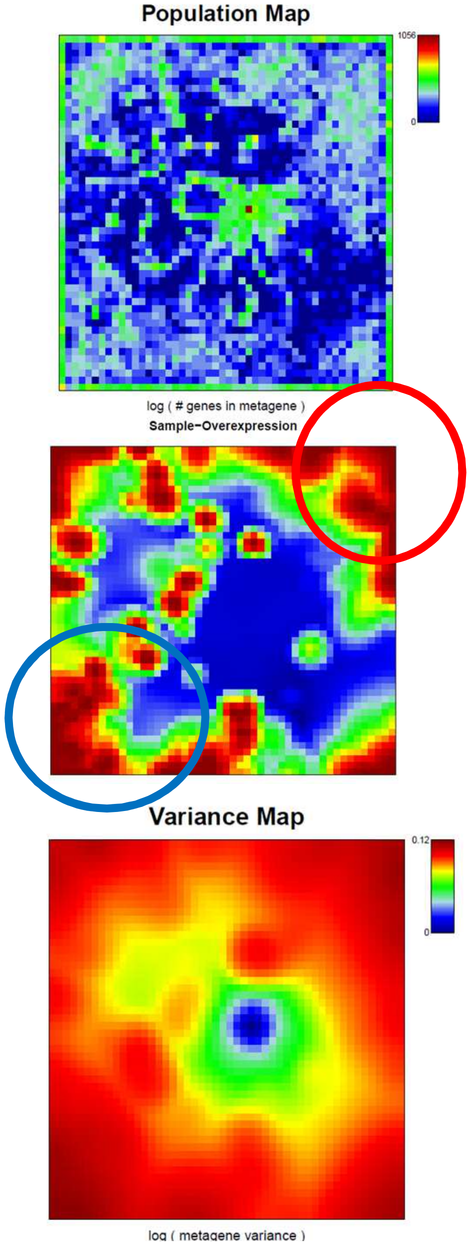
The NEW ENGLAND JOURNAL of MEDICINE

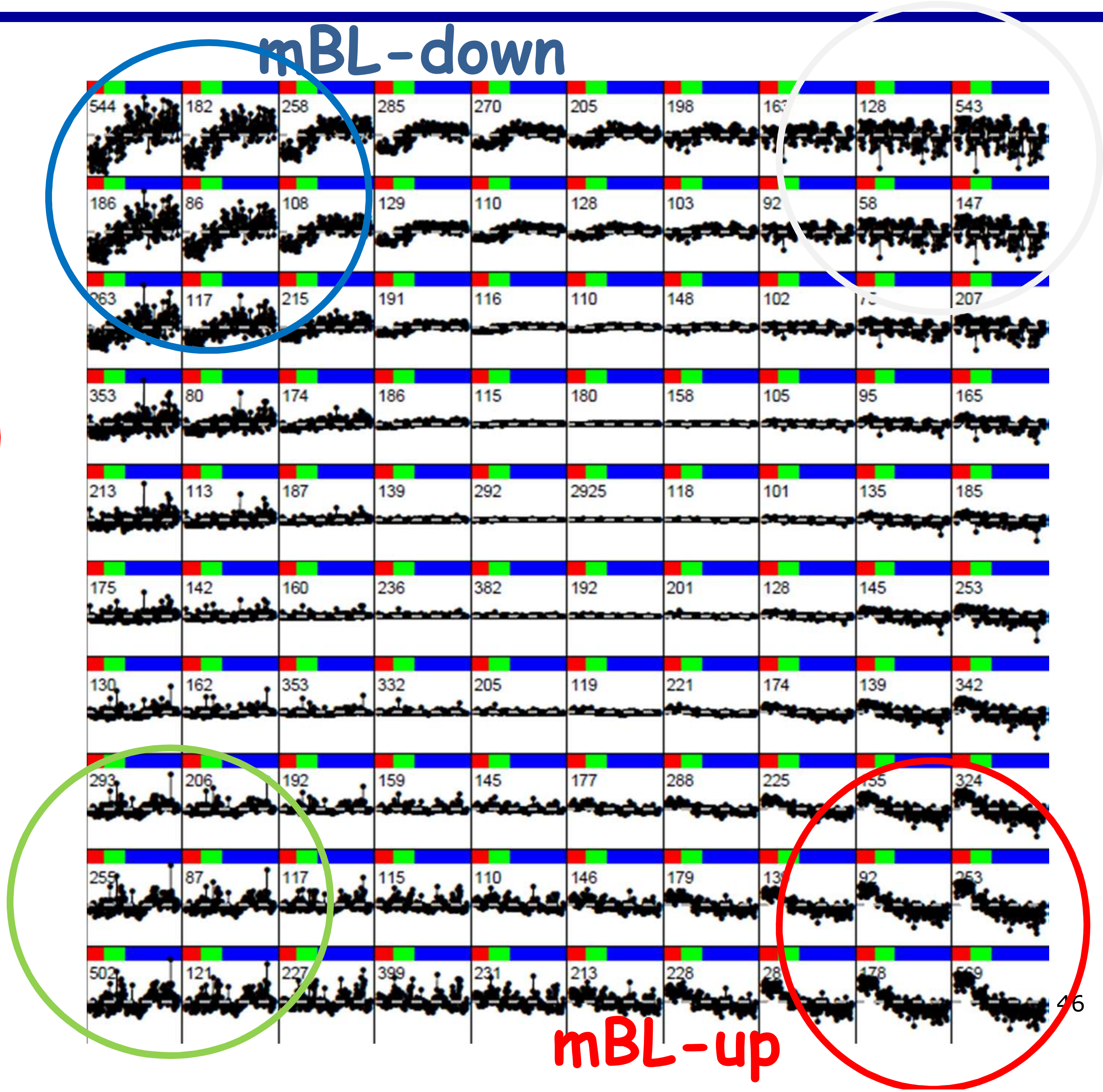
ESTABLISHED IN 1812 | JUNE 8, 2006 | VOL. 354 NO. 23

A Biologic Definition of Burkitt's Lymphoma from Transcriptional and Genomic Profiling

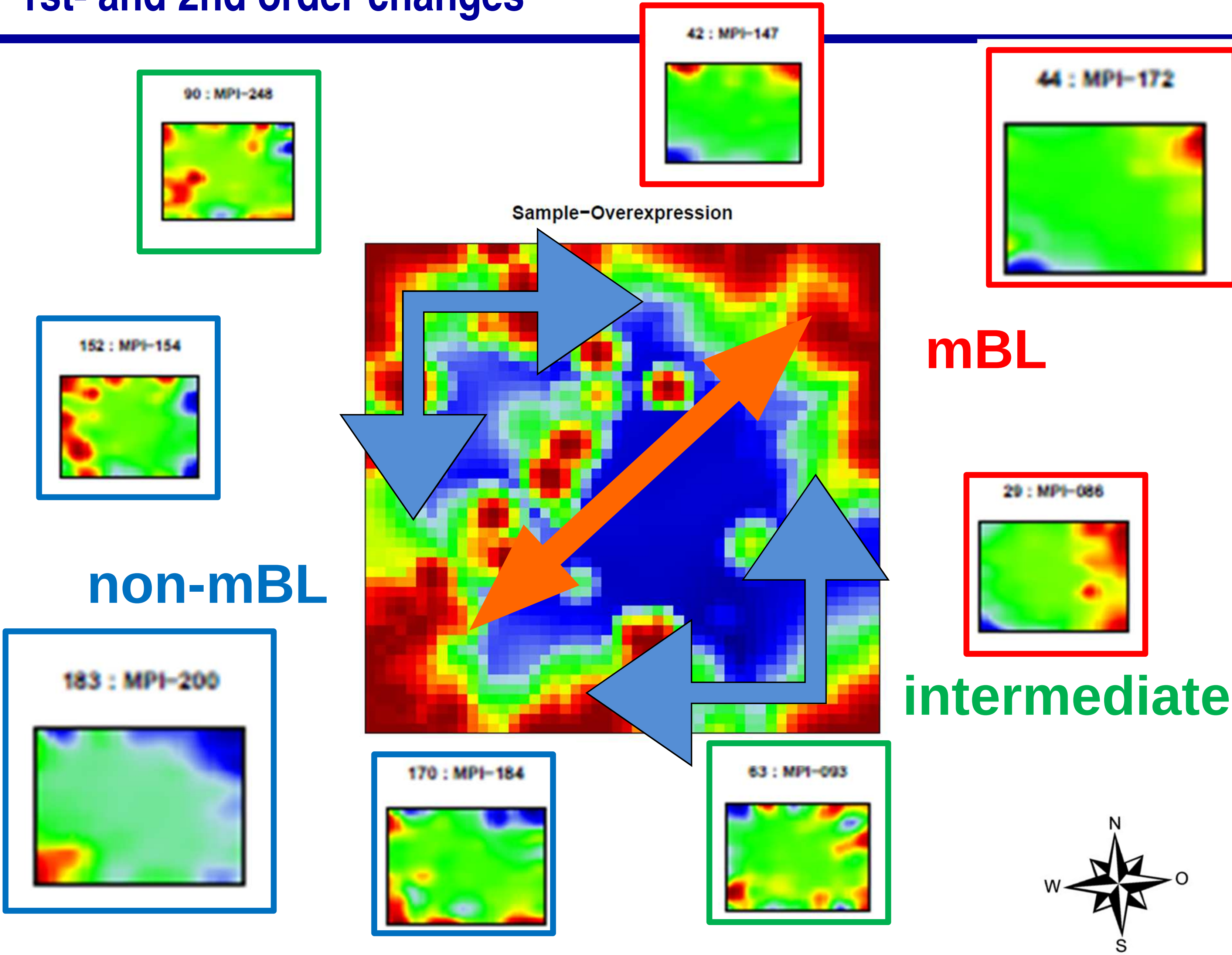
Michael Hummel, Ph.D., Stefan Bentink, M.S., Hilmar Berger, M.D., Wolfram Klapper, M.D., Swen Wessendorf, M.D., Thomas F.E. Barth, M.D., Heinz-Wolfram Bernd, M.D., Sergio B. Cogliatti, M.D., Judith Dierlamm, M.D., Ph.D., Alfred C. Feller, M.D., Martin-Leo Hansmann, M.D., Eugenia Haralambieva, M.D., Lana Harder, M.D.,

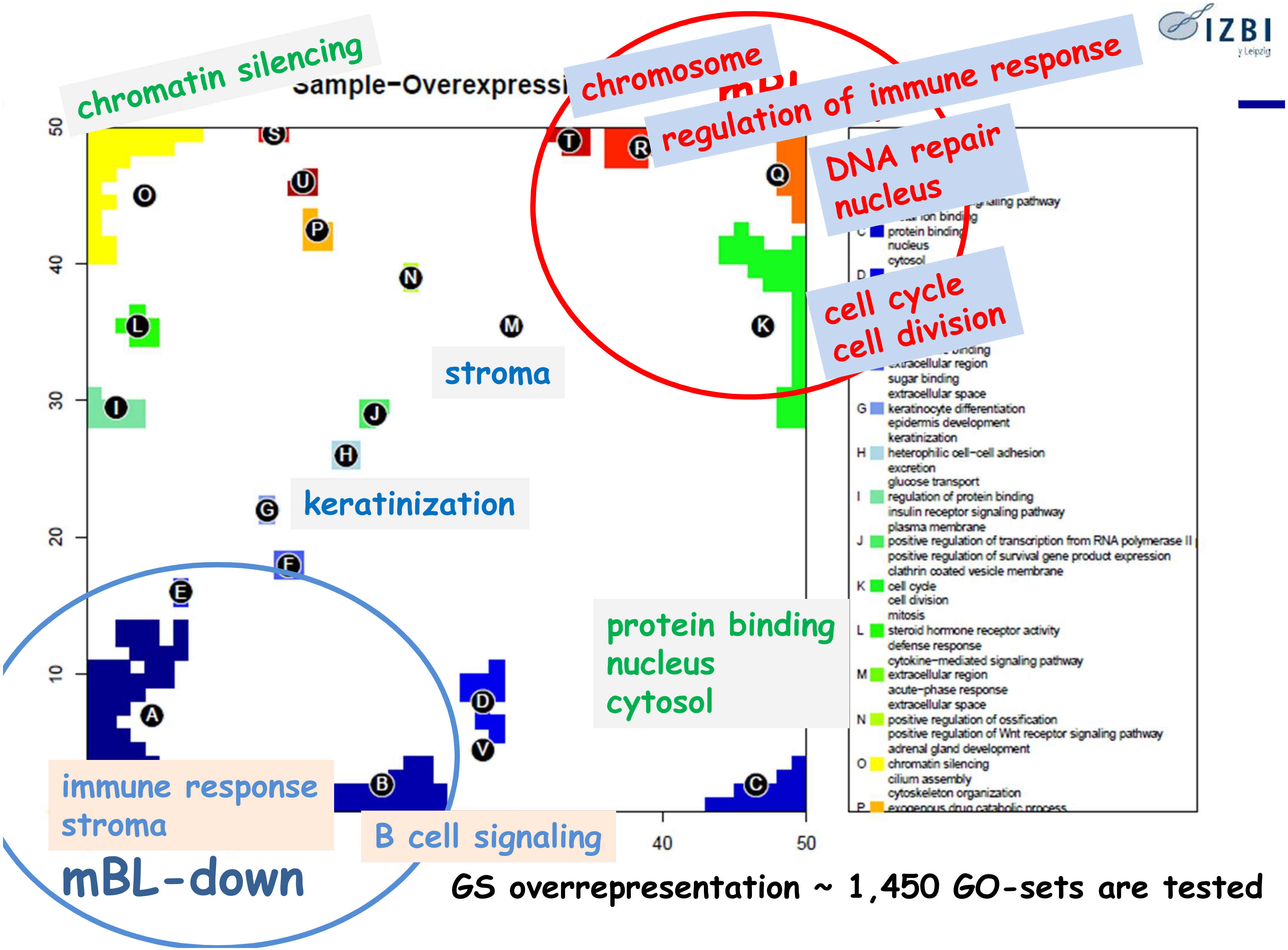
Supporting maps: Profiling map



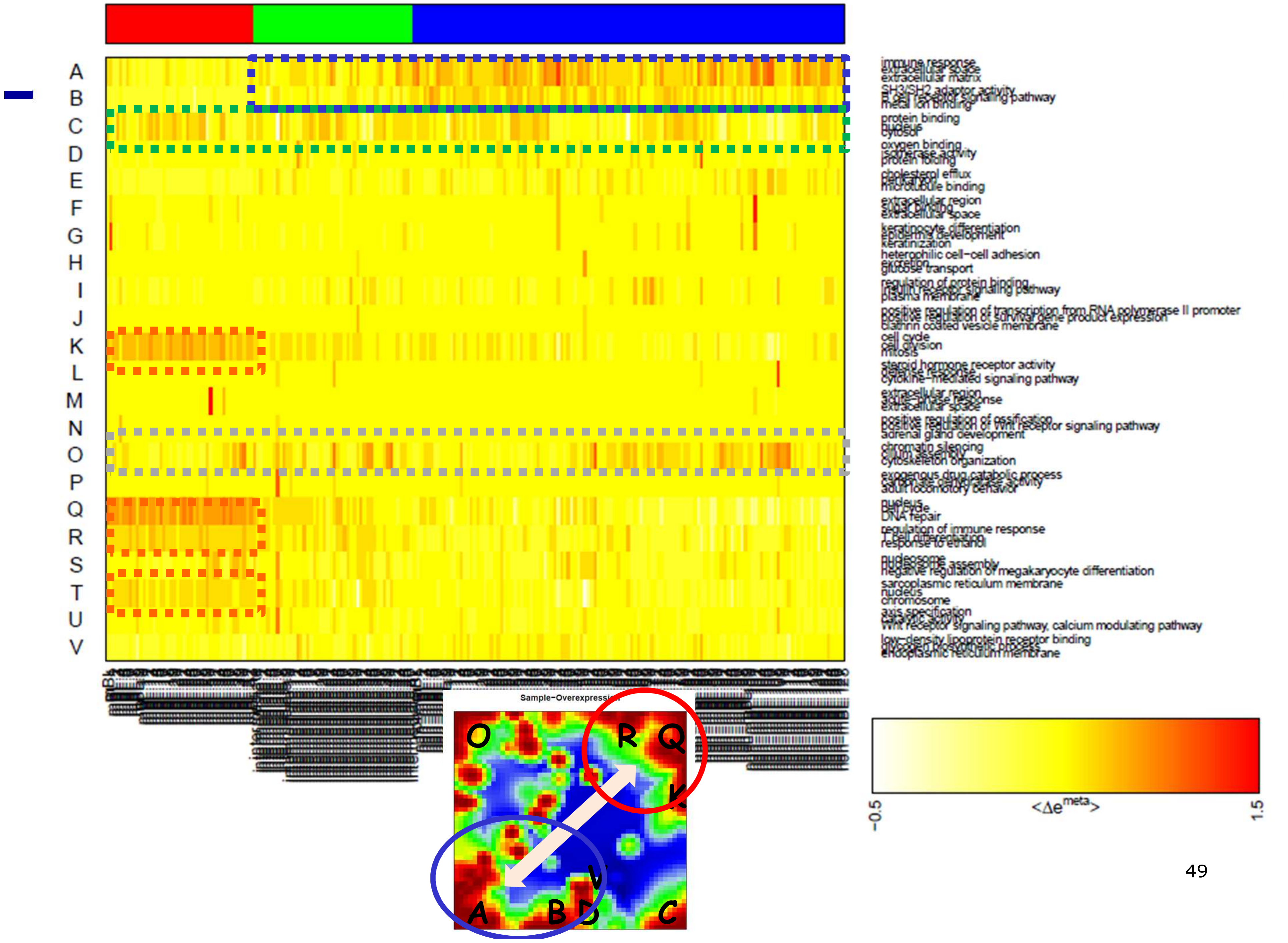


1st- and 2nd order changes

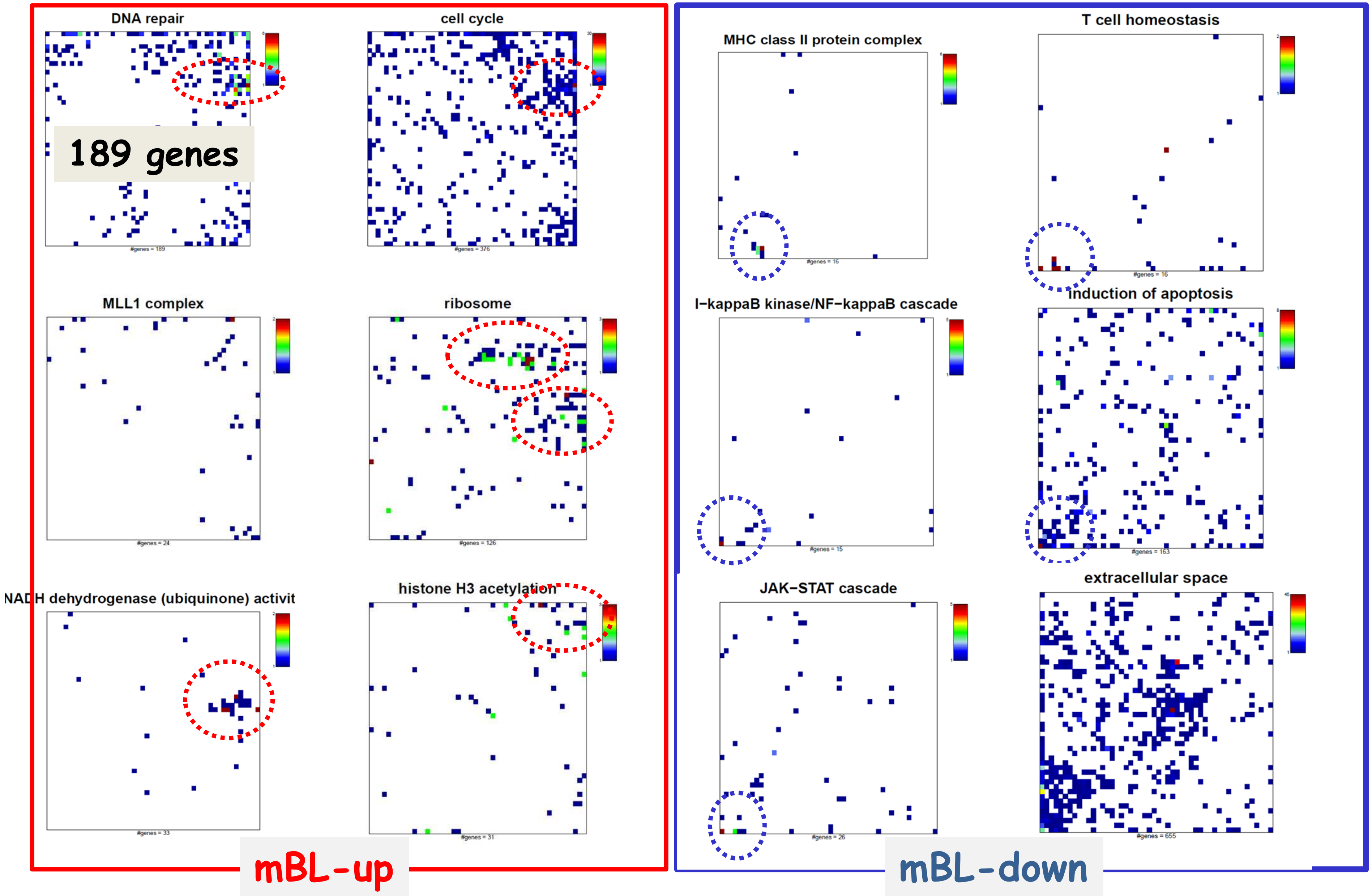




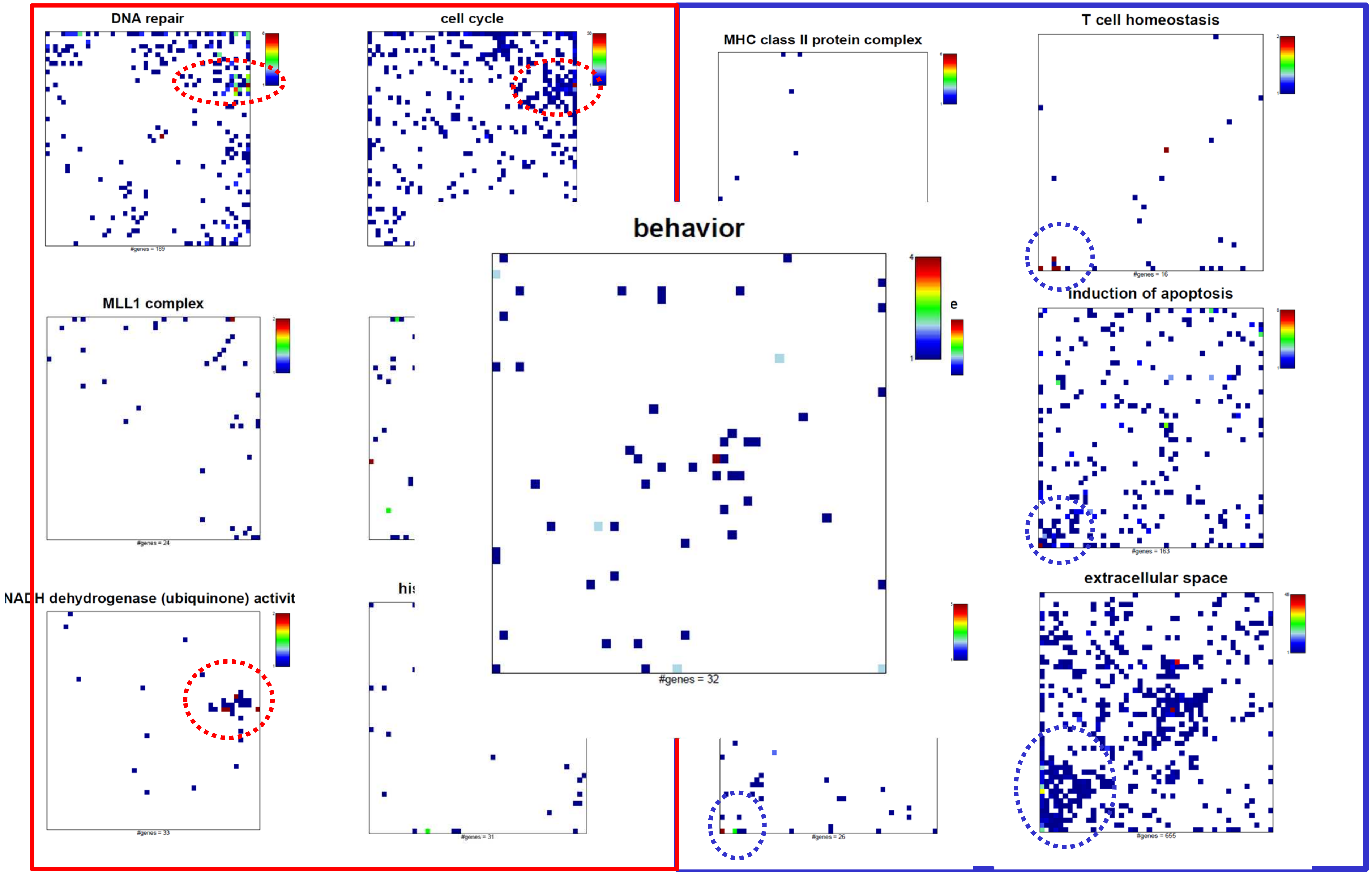
chromatin silencing
Sample-Overexpressi
chromosome
regulation of immune response
DNA repair
nucleus
cell cycle
cell division
IZBI
stroma
keratinization
protein binding
nucleus
cytosol
immune response
stroma
B cell signaling
mBL-down
GS overrepresentation ~ 1,450 GO-sets are tested
50
40
30
20
10
40
50
O
S
U
P
N
L
M
I
J
H
G
F
E
A
B
D
V
C
T
R
Q
K
signaling pathway
ion binding
C protein binding
nucleus
cytosol
D
extracellular region
sugar binding
extracellular space
G keratinocyte differentiation
epidermis development
keratinization
H heterophilic cell-cell adhesion
excretion
glucose transport
I regulation of protein binding
insulin receptor signaling pathway
plasma membrane
J positive regulation of transcription from RNA polymerase II
positive regulation of survival gene product expression
clathrin coated vesicle membrane
K cell cycle
cell division
mitosis
L steroid hormone receptor activity
defense response
cytokine-mediated signaling pathway
M extracellular region
acute-phase response
extracellular space
N positive regulation of ossification
positive regulation of Wnt receptor signaling pathway
adrenal gland development
O chromatin silencing
cilium assembly
cytoskeleton organization
P exogenous drug catabolic process


A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
heterophilic cell−cell adhesion
glucose transport
positive regulation of transcription from RNA polymerase II promoter
clathrin coated vesicle membrane
steroid hormone receptor activity
cytokine−mediated signaling pathway
cytoskeleton organization
exogenous drug catabolic process
regulation of immune response
sarcoplasmic reticulum membrane
chromosome
Wnt receptor signaling pathway, calcium modulating pathway
low−density lipoprotein receptor binding
Sample−Overexpression
−0.5
1.5

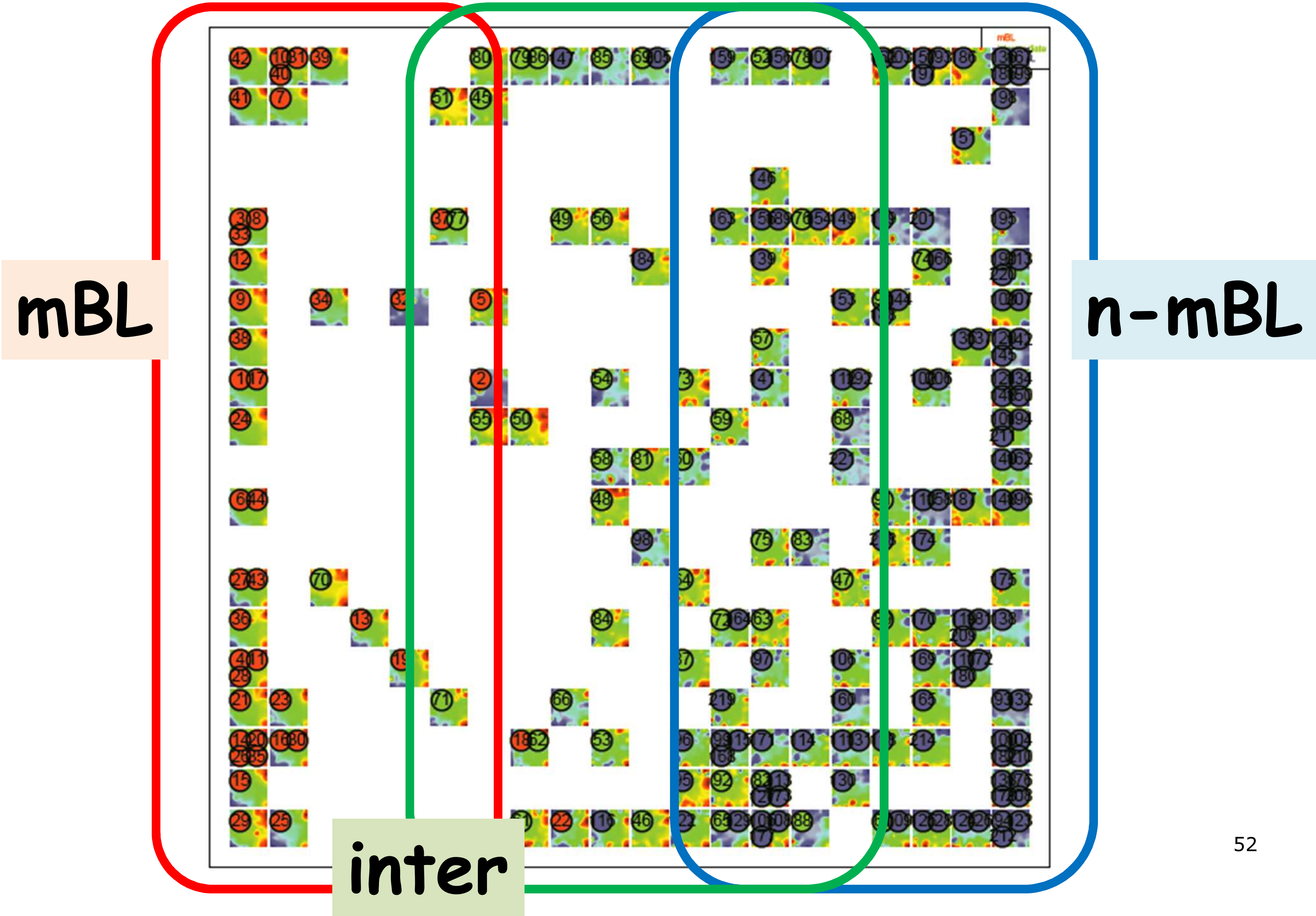
Go-geneset maps



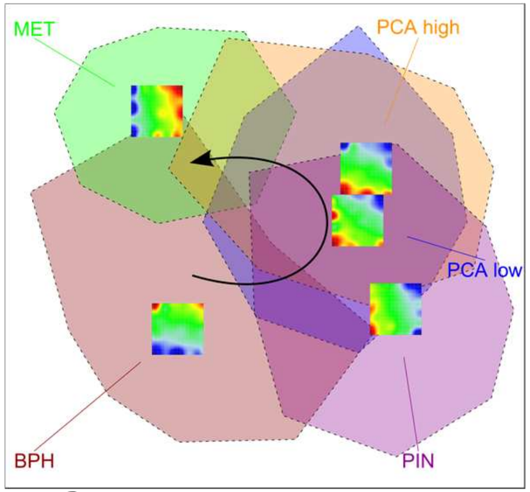
Go-geneset maps



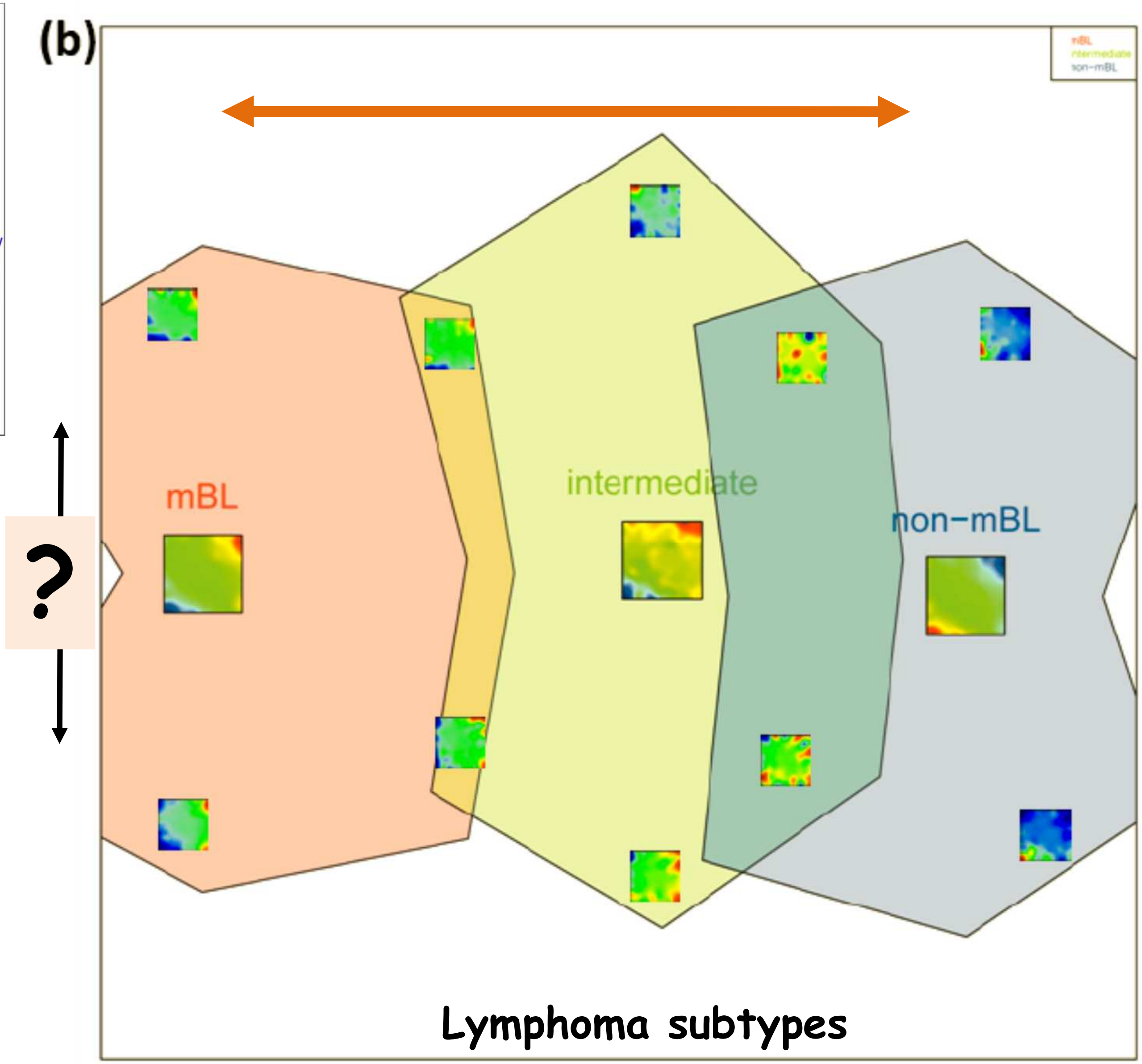
Similarity relations: 2nd level SOM



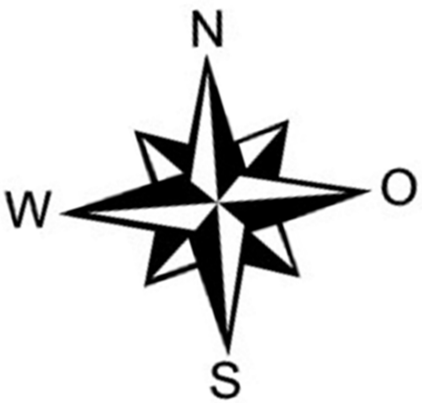
The problem is virtually one-dimensional



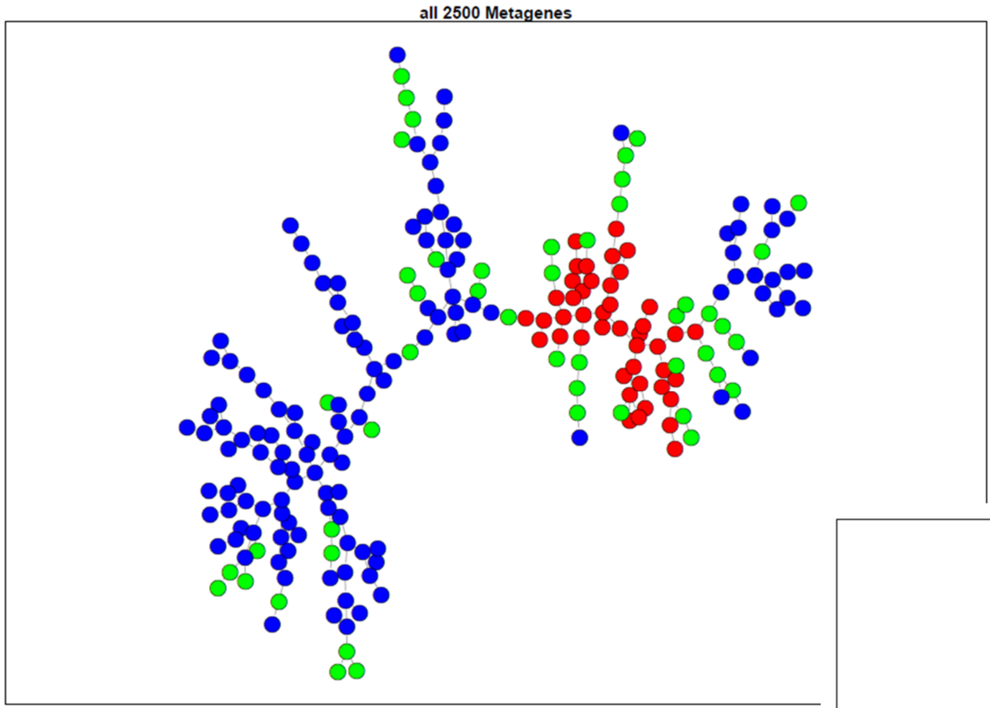
Prostate cancer progression



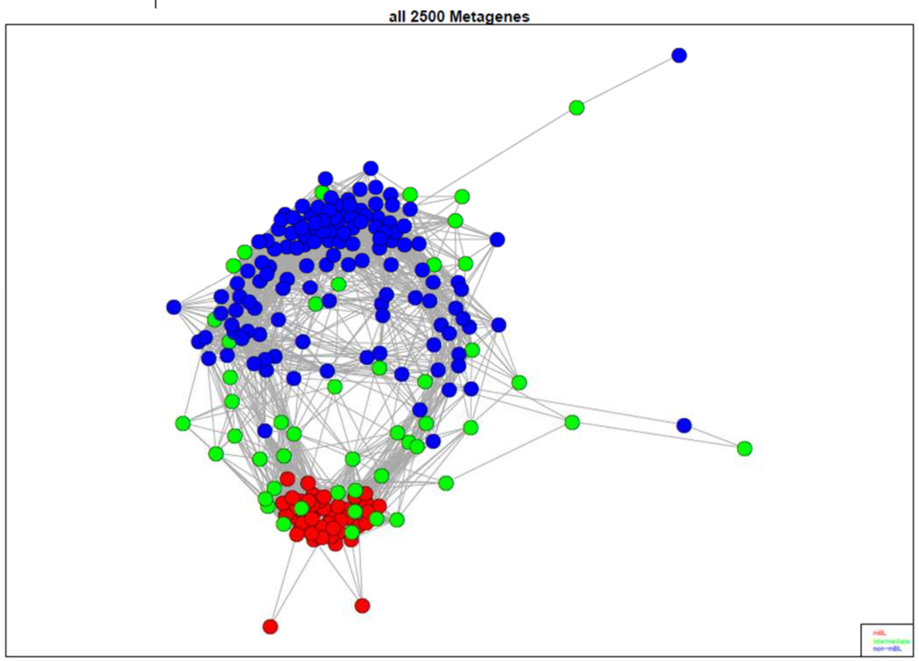
Lymphoma subtypes



Korrelation networks

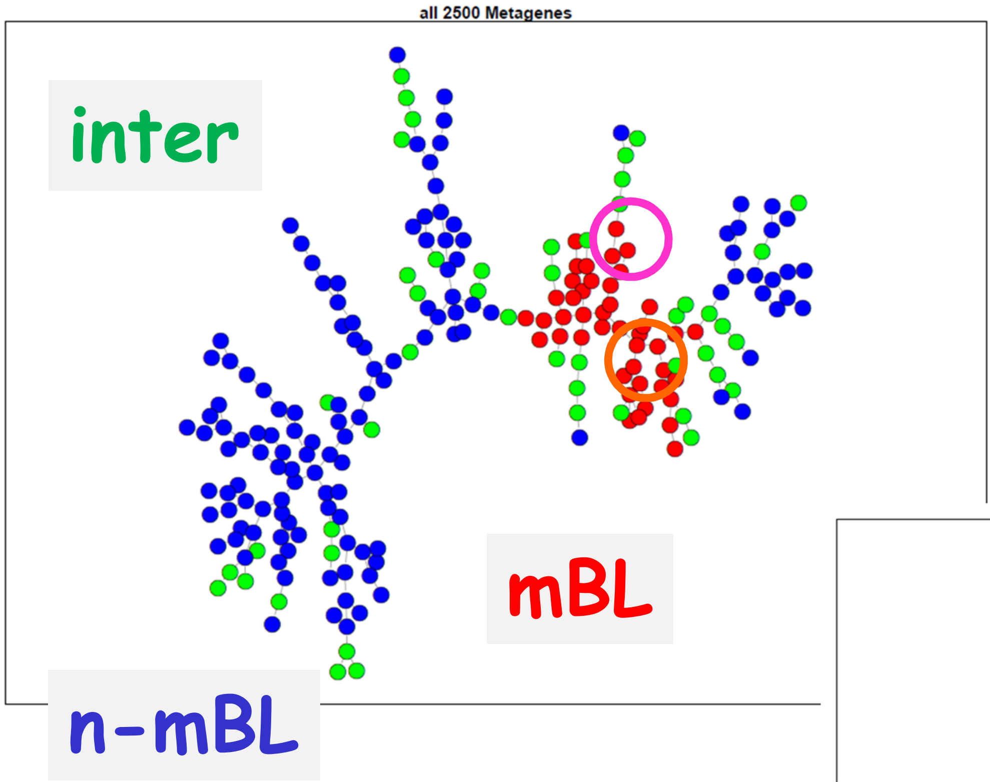


Maximum spanning tree:
Connects strongest pairwise correlations



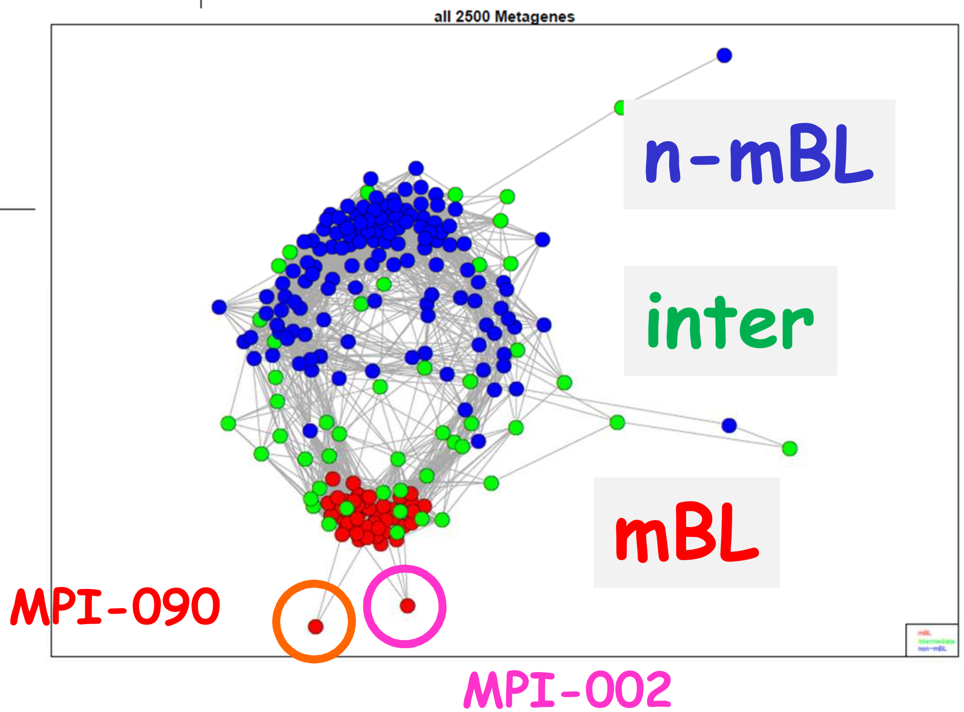
Correlation net:
Connects all samples mutually correlated with r>0.5

Correlation networks

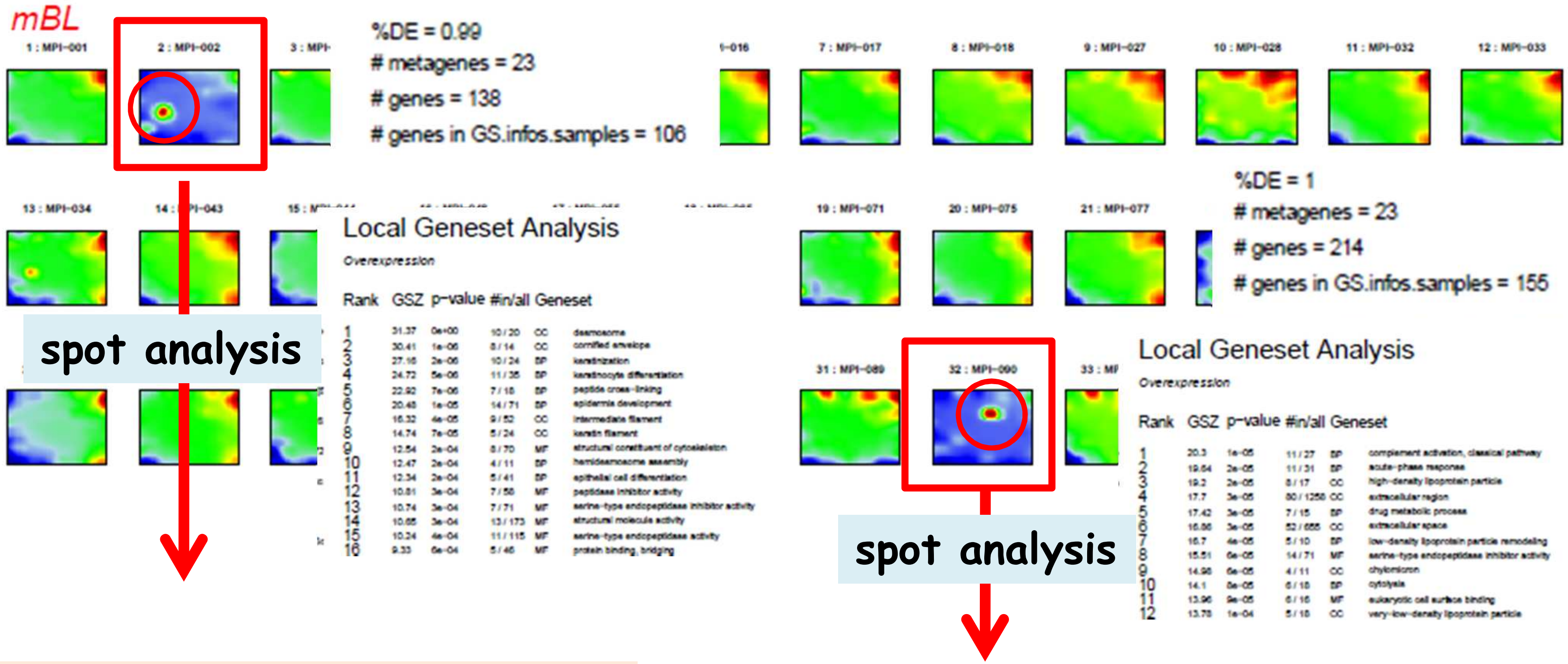


Maximum spanning tree:
Connects strongest pairwise correlations

Correlation net:
Connects all samples mutually correlated with r>0.5



Detecting and analyzing ‚contaminations'



Contamination:
Endothel
Keratin
→ Healthy lymph node tissue

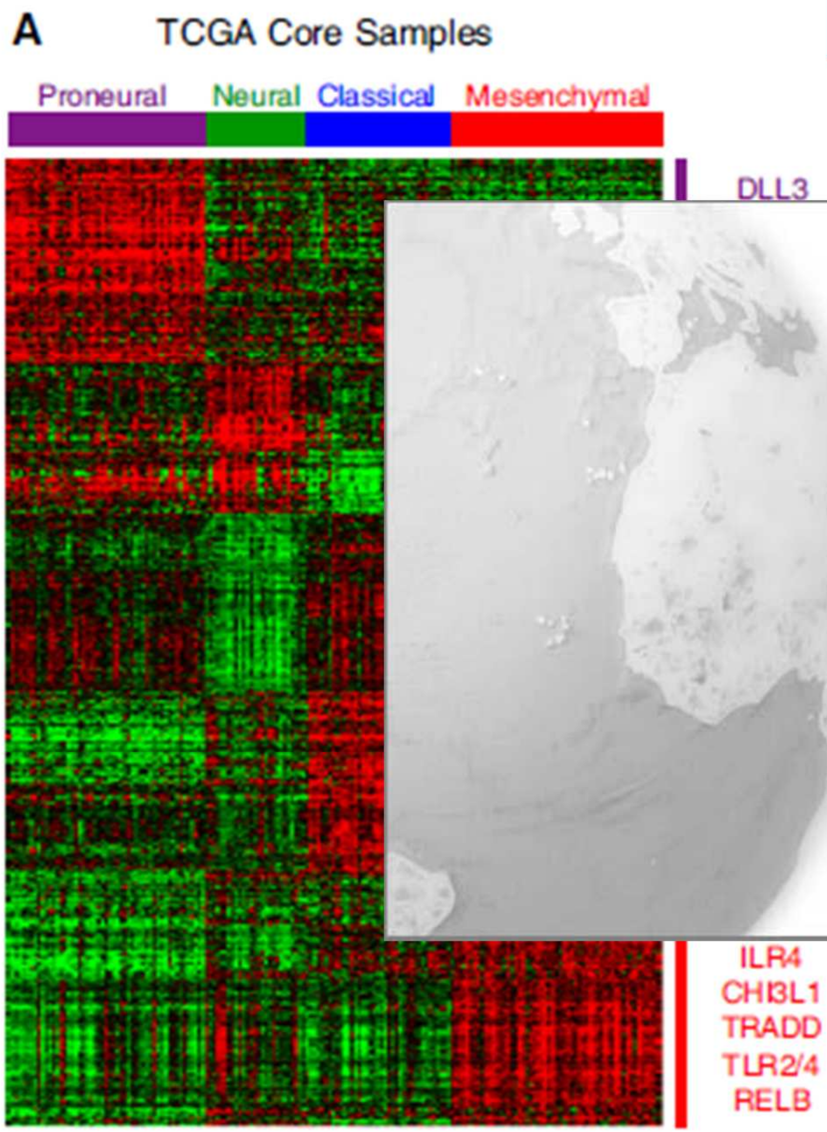
Contamination:
C-reactive protein
Albumin
Complement activation
Acute phase response

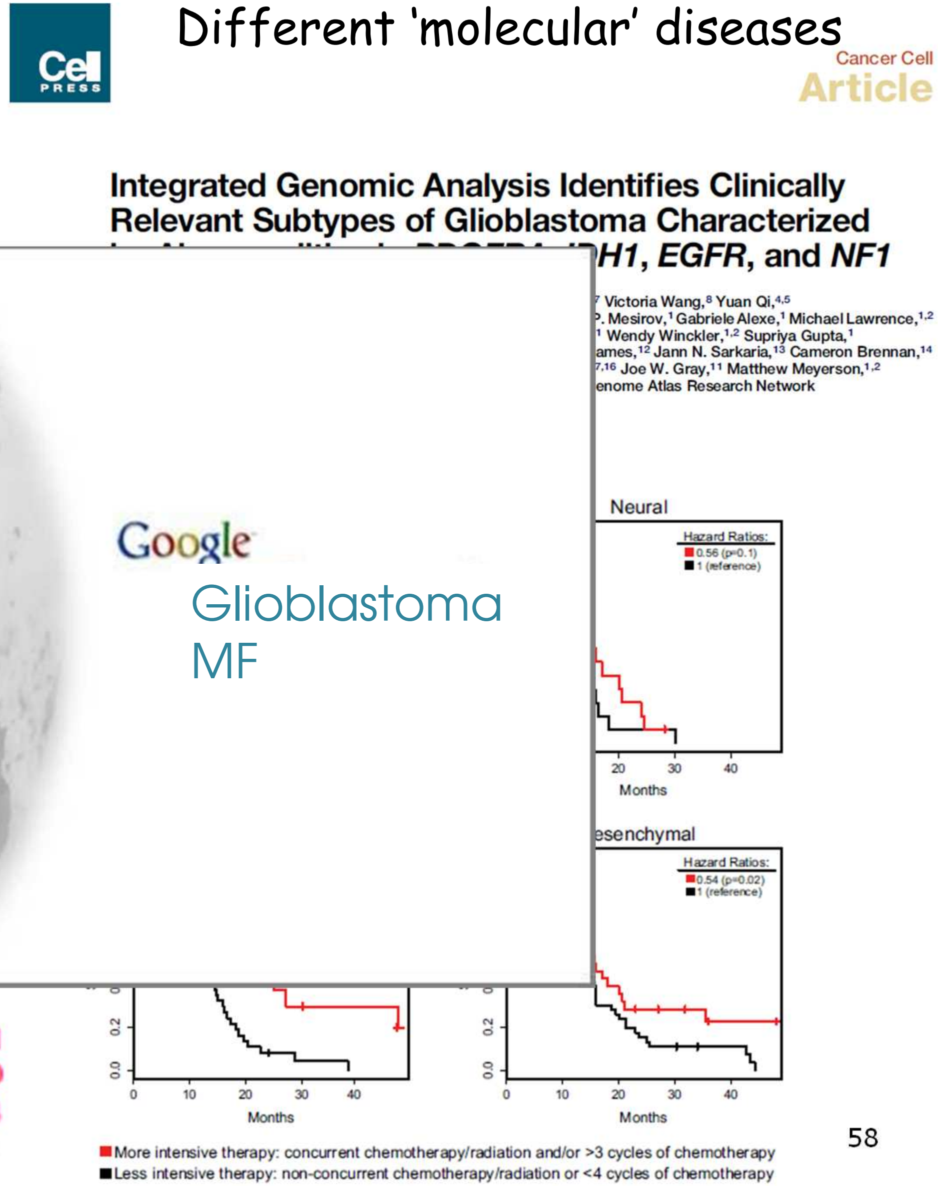
SOM and cancer

1. Spot characteristics of cancer subtypes
2. Similarity analysis: relations between subtypes
3. Individual portraits of the samples
4. Outlier-/ healty tissue- identification
5. GSEA: Assignment of biological processes/components associated with dysfunctions

Starting point: GMF subtypes

Different 'molecular' diseases

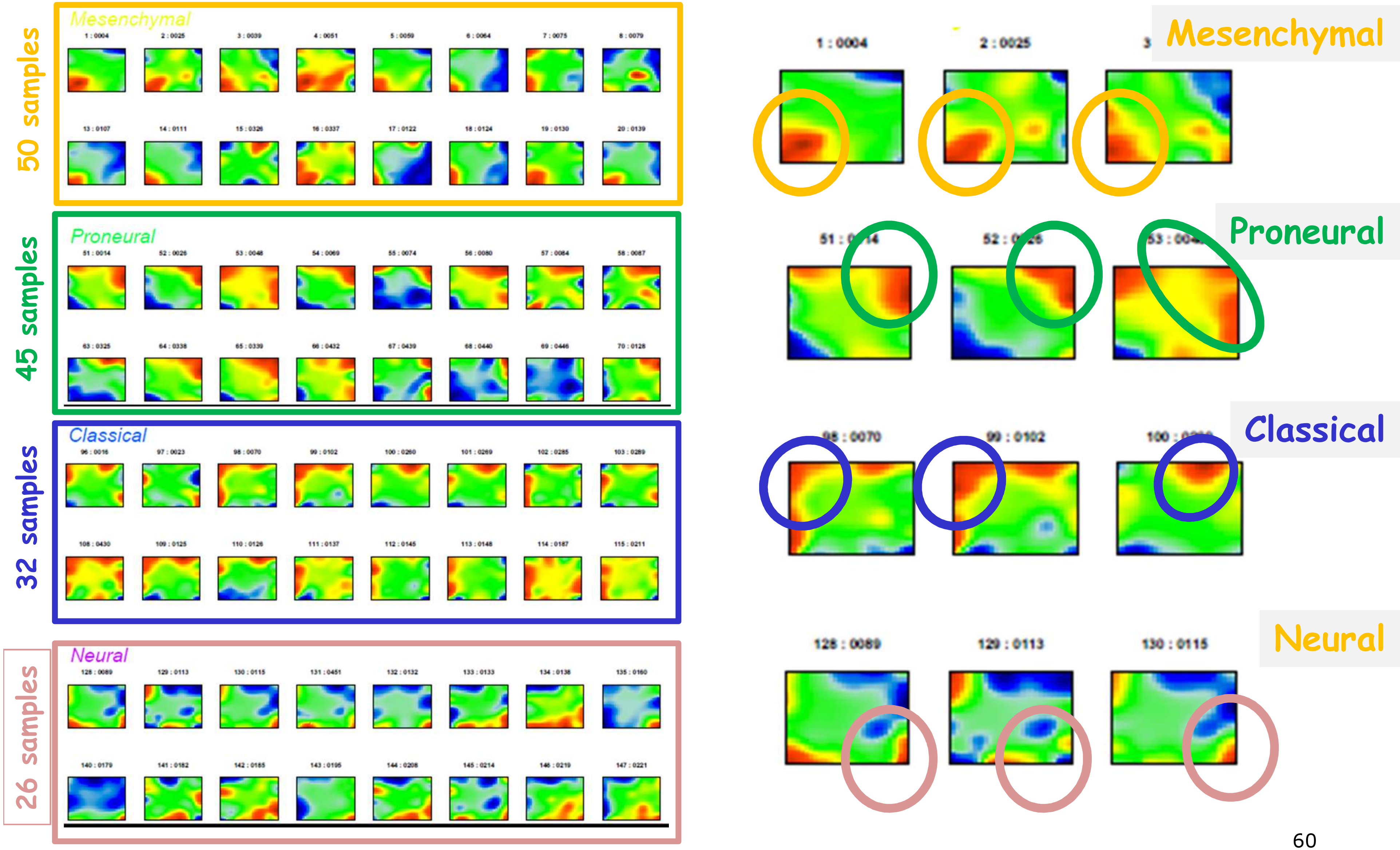




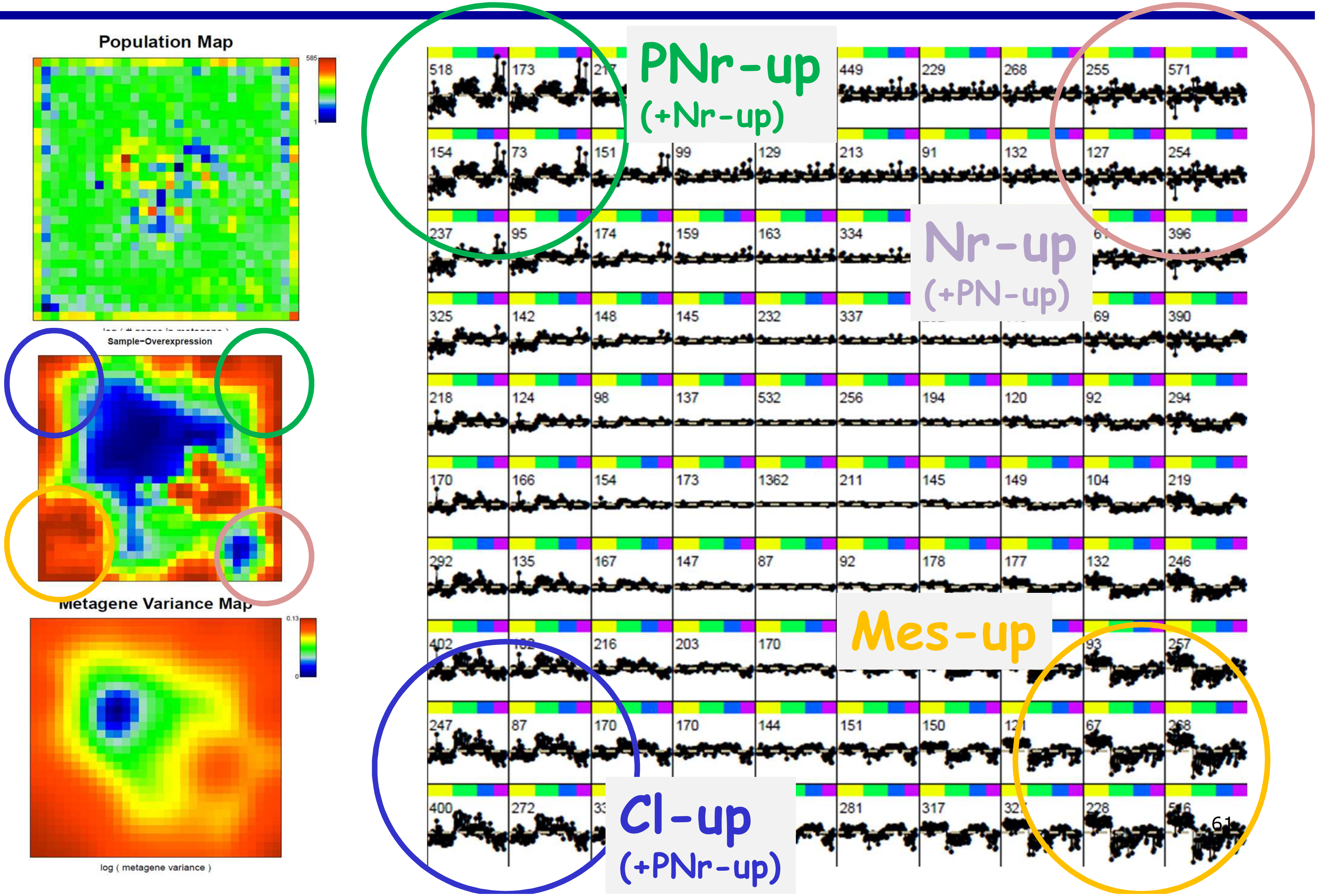
- Affy-Level 1 expression data (*.cel-files)
- Hook preprocessing + Quantile normalization
- Quality control → 153 samples
- → Separate story
- → Class labels of the paper

- → Original classification:
- → Larger data set
- → RMA preprocessing

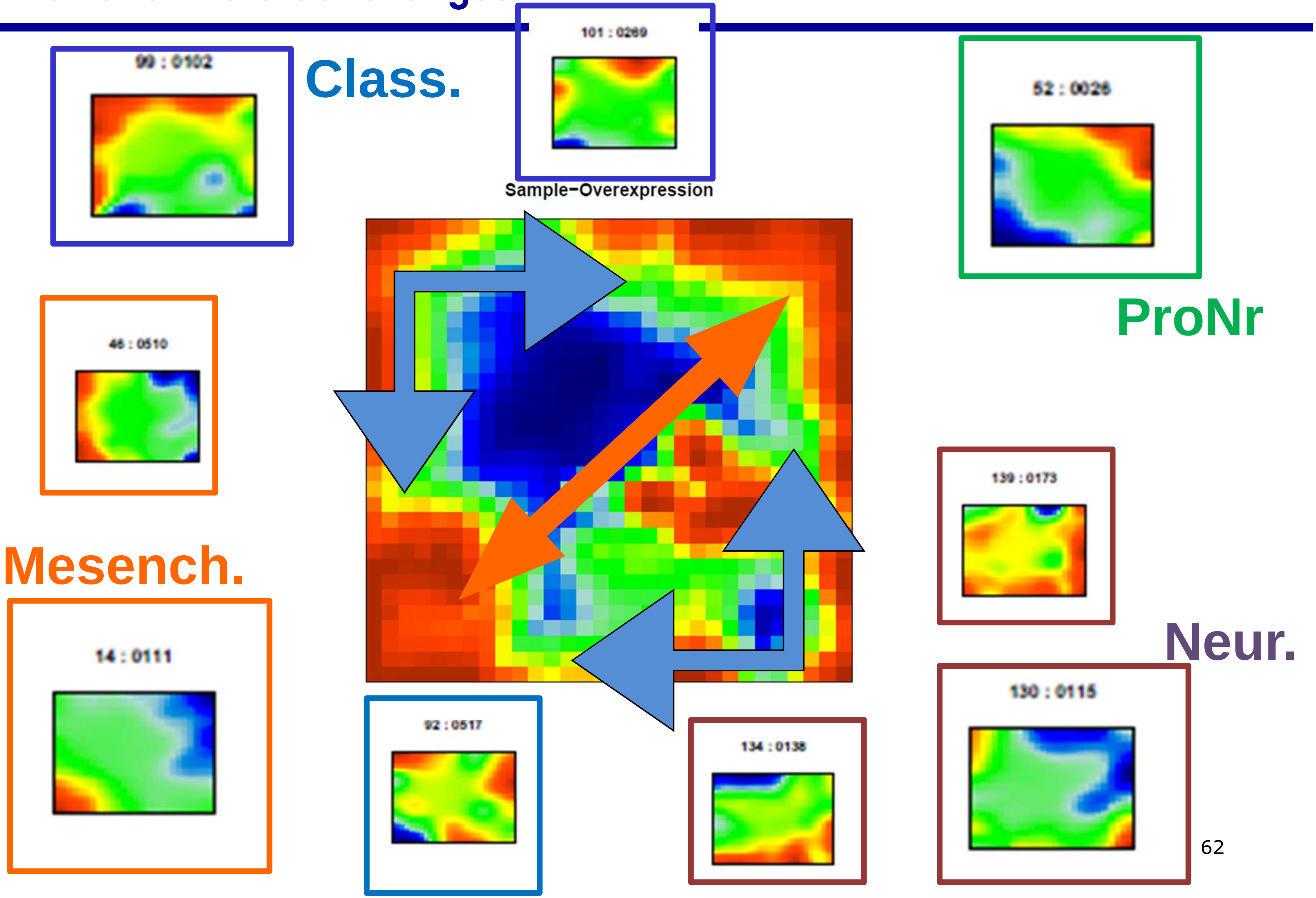
Disentangling GMF subtypes



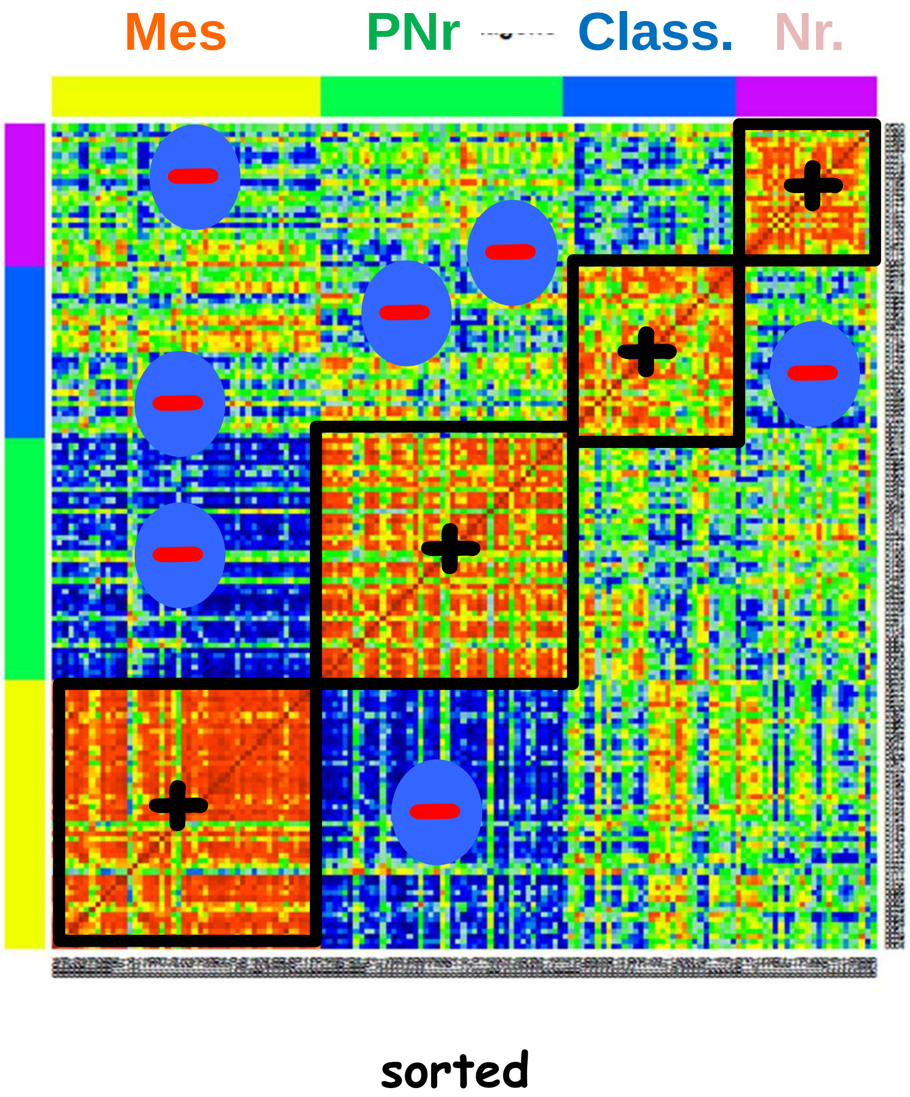
Supporting maps: Profiling map



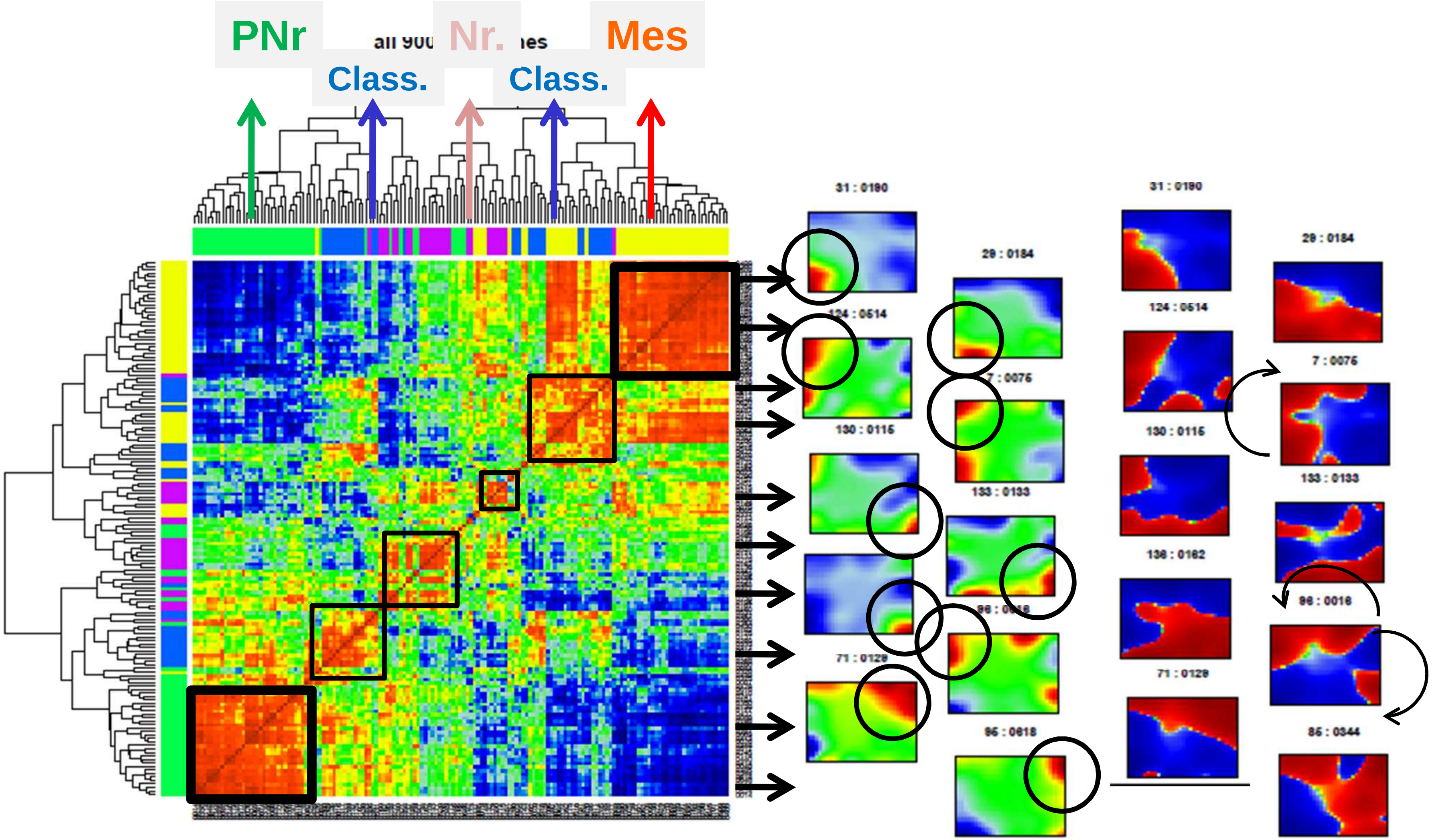
1st- and 2nd order changes



Pairwise correlation map of the portrays

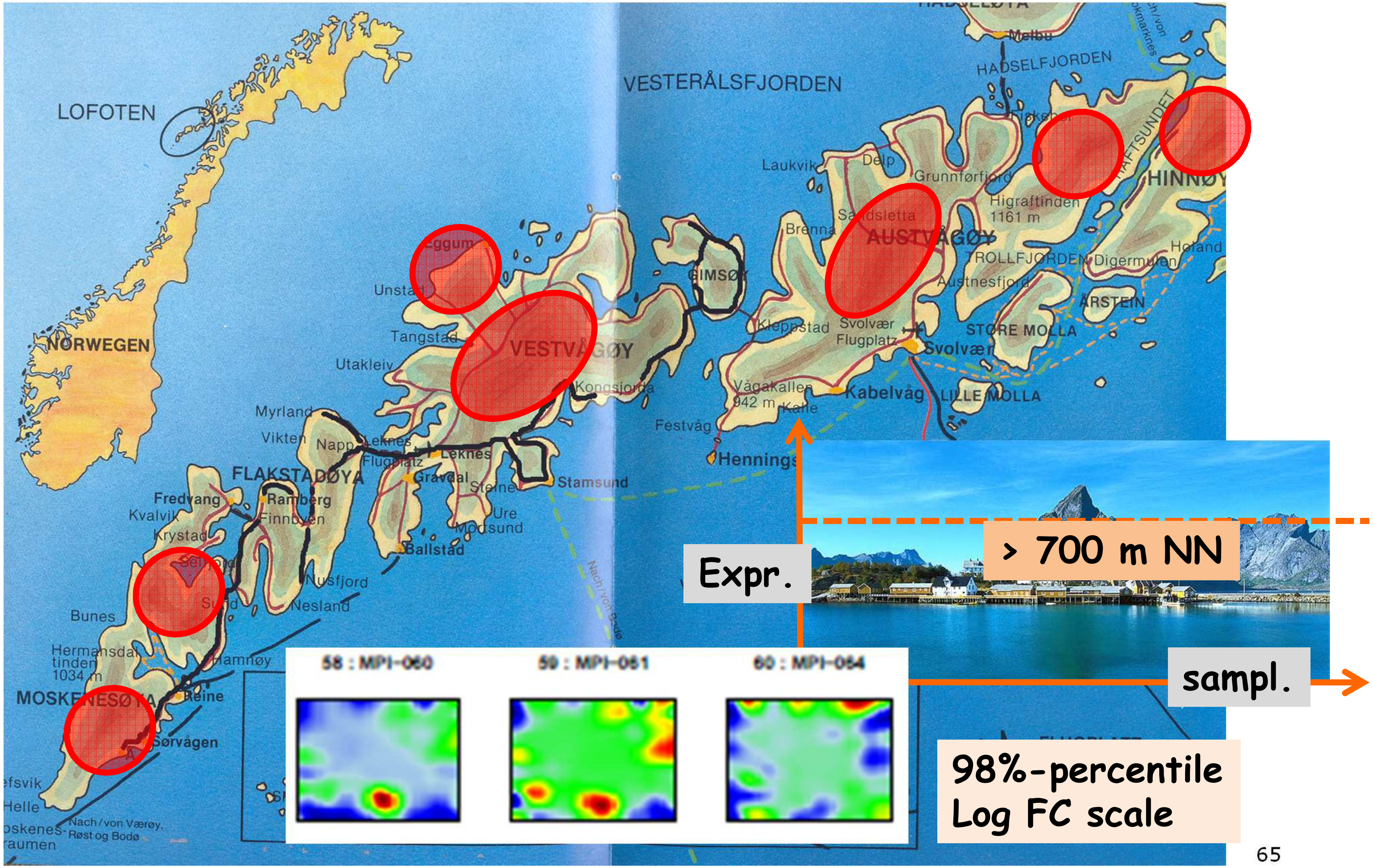


Pairwise correlation map

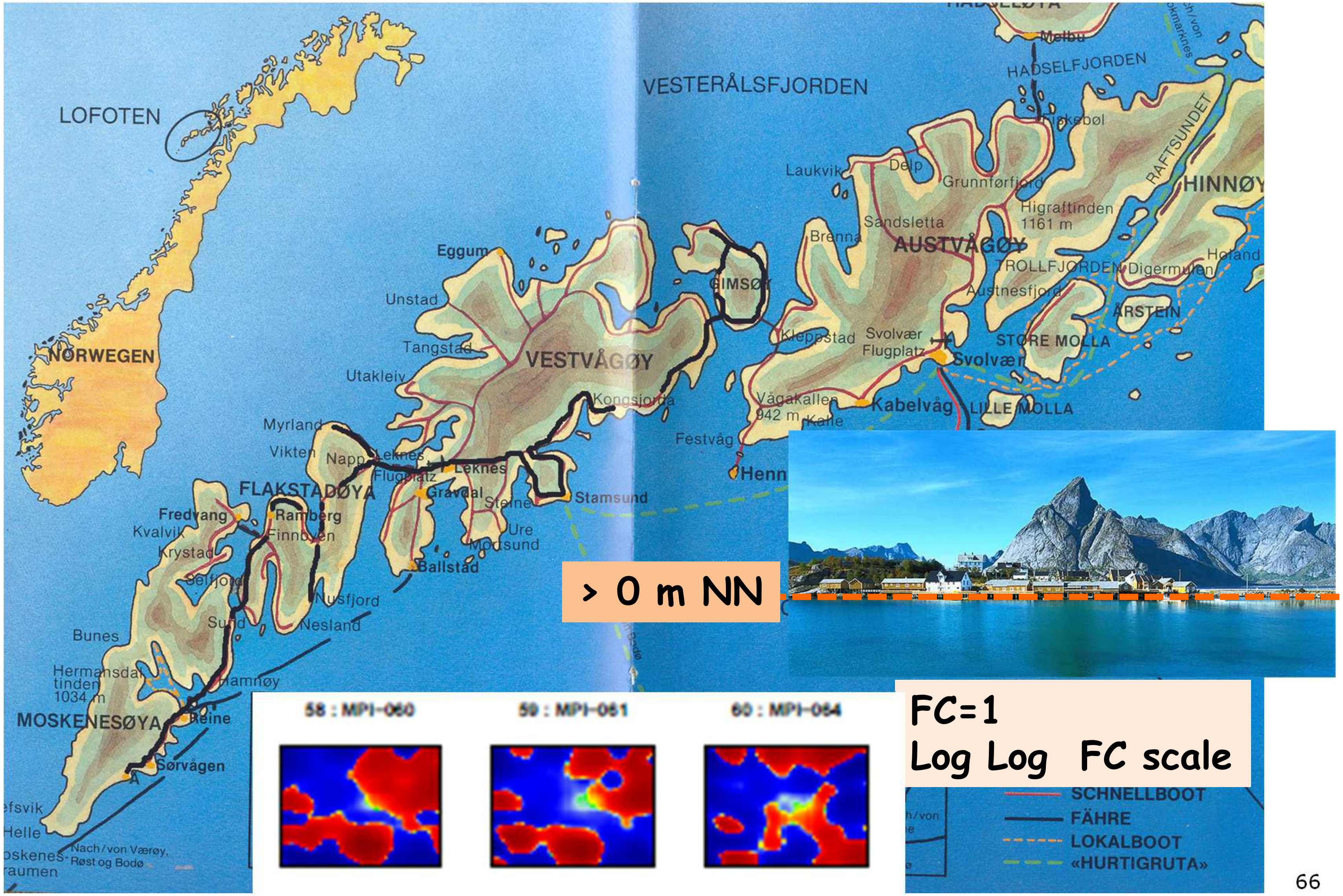


HClustered

2 scales: high expression



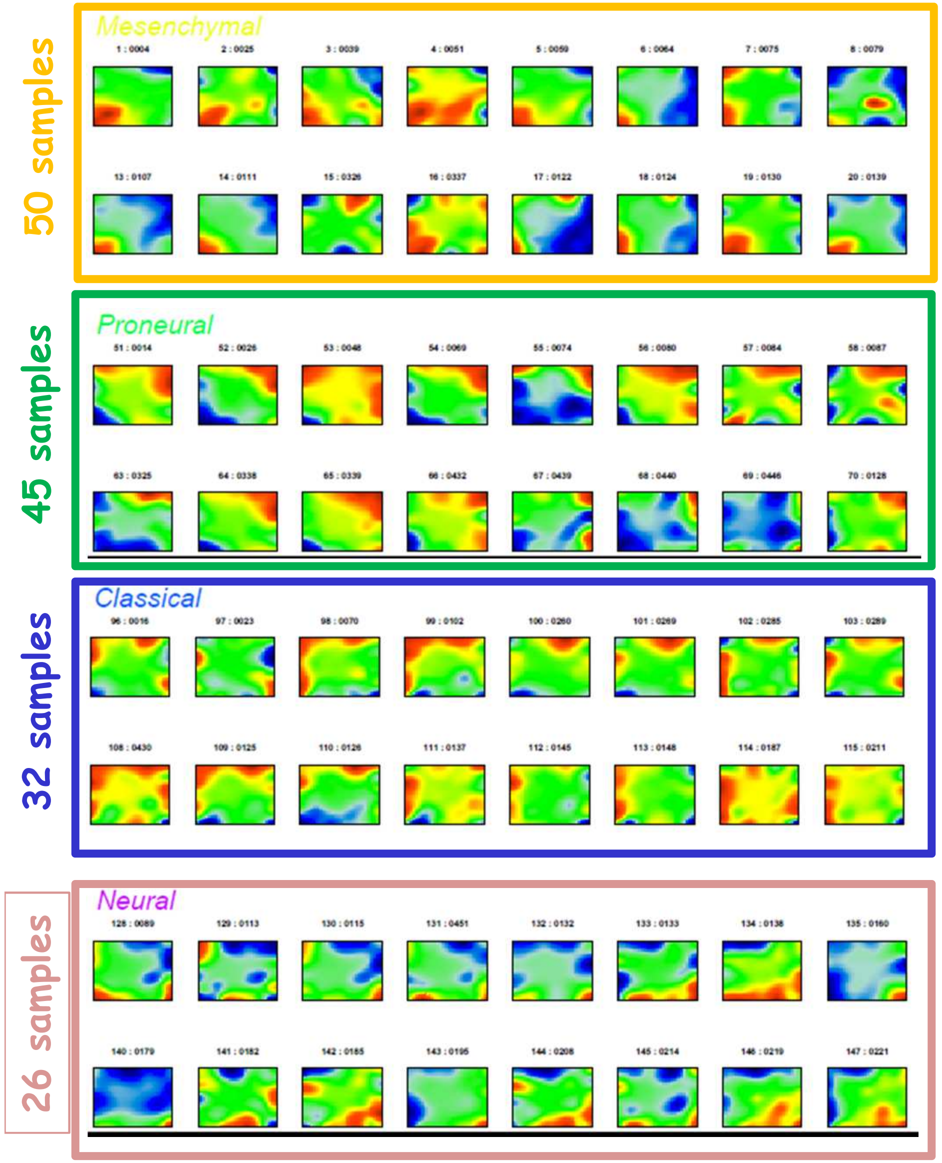
2 scales: average expression (log log FC)



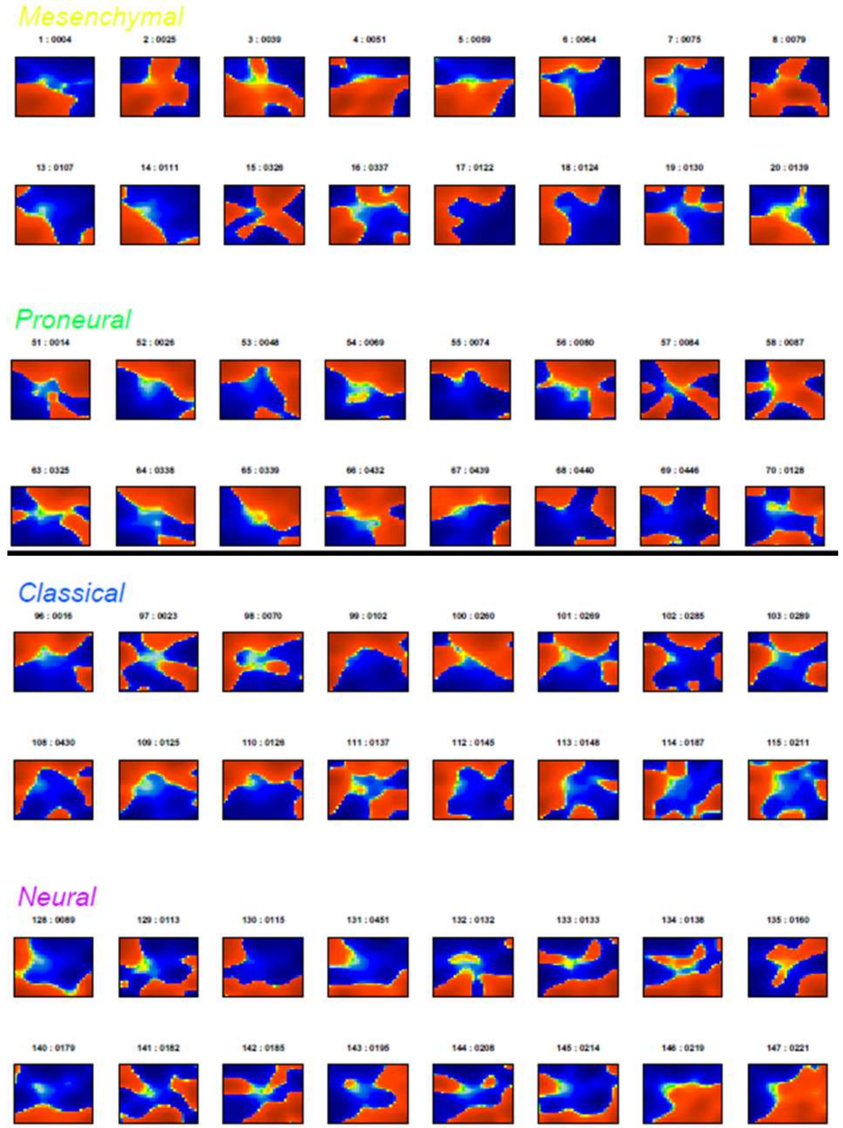
> 0 m NN

FC=1
Log Log FC scale

Disentangling GMF subtypes

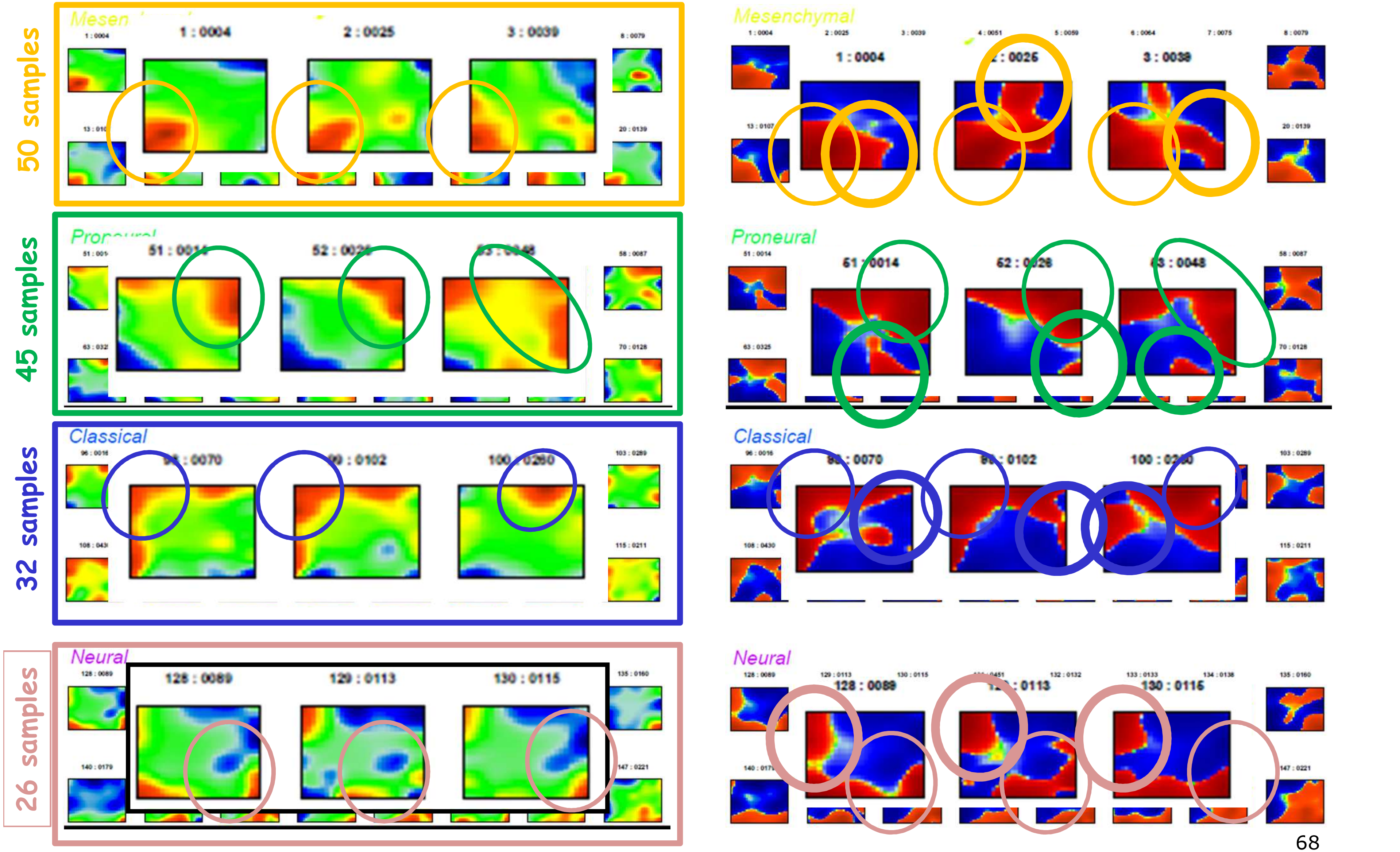


log FC-scale: top expression



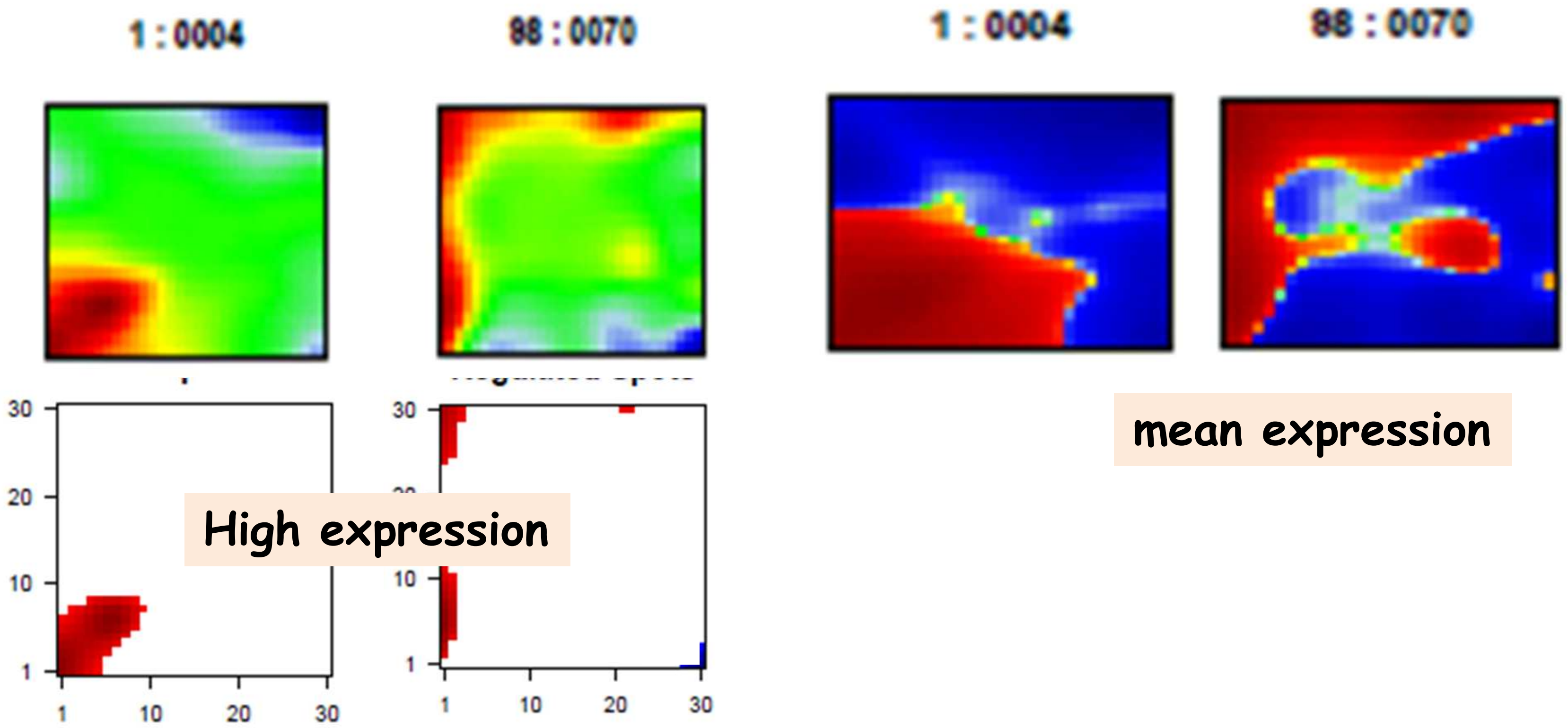
Log log FC-scale: up-down regulated

Disentangling GMF subtypes



Log log FC-scale: up-down regulated

Topological measures: GMF



Characterizing the fuzziness of the expression landscape:

# of red pixels: # of highly expressed metagenes

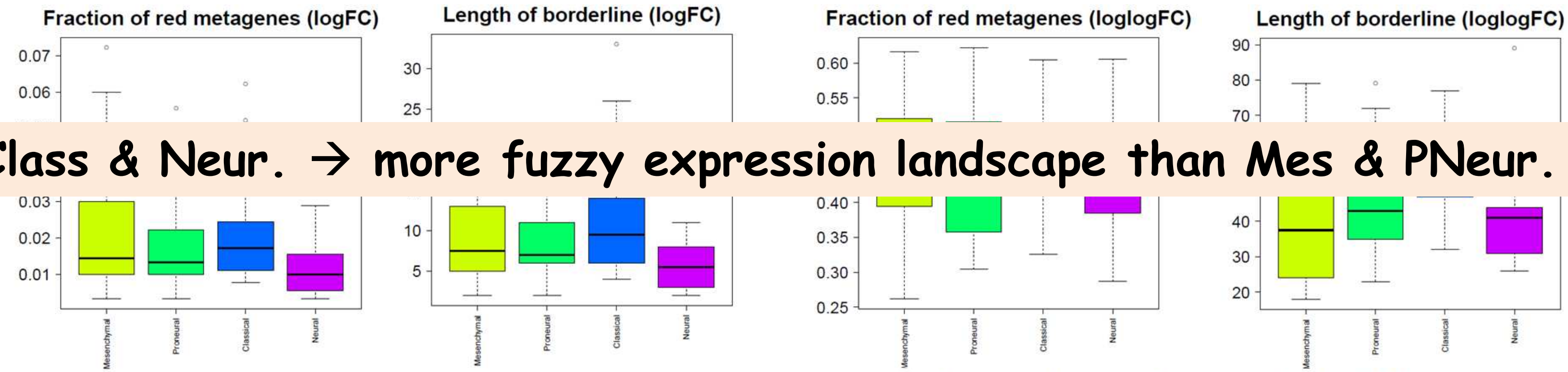
# of red pixels forming the borderline:

Compactness: area/ border line

Topological measures: GMF

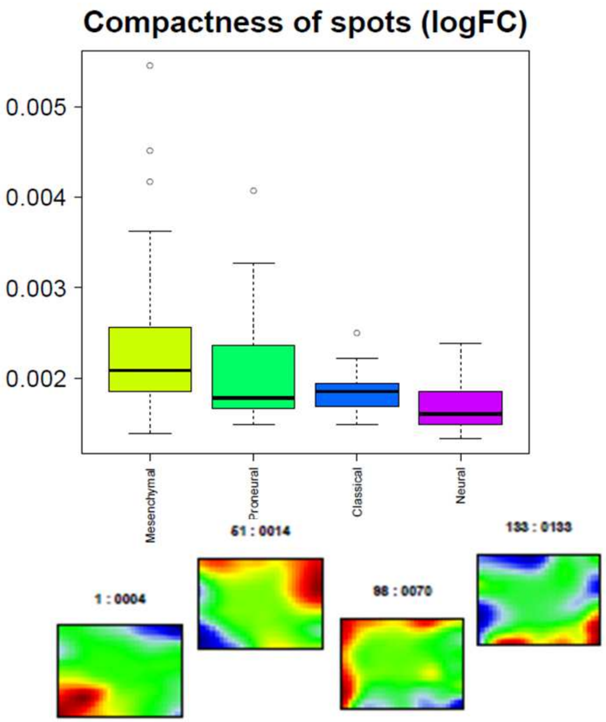
High expression

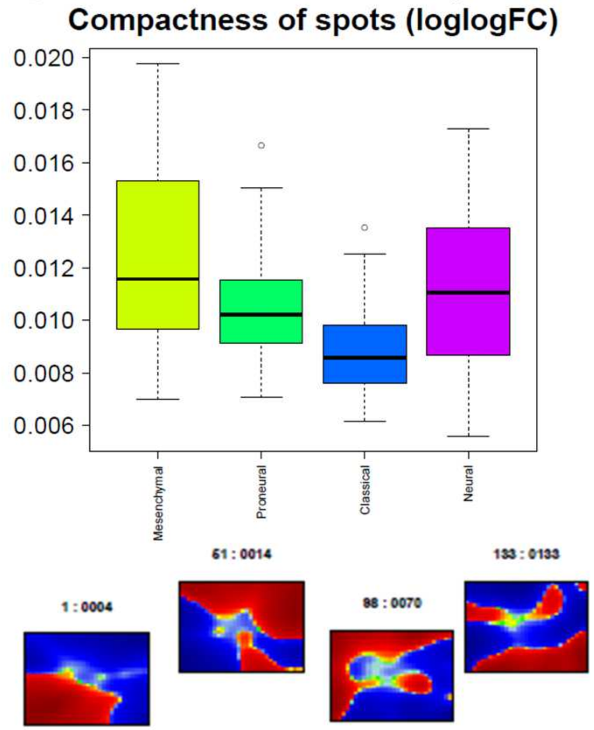
mean expression



Class & Neur. → more fuzzy expression landscape than Mes & PNeur.

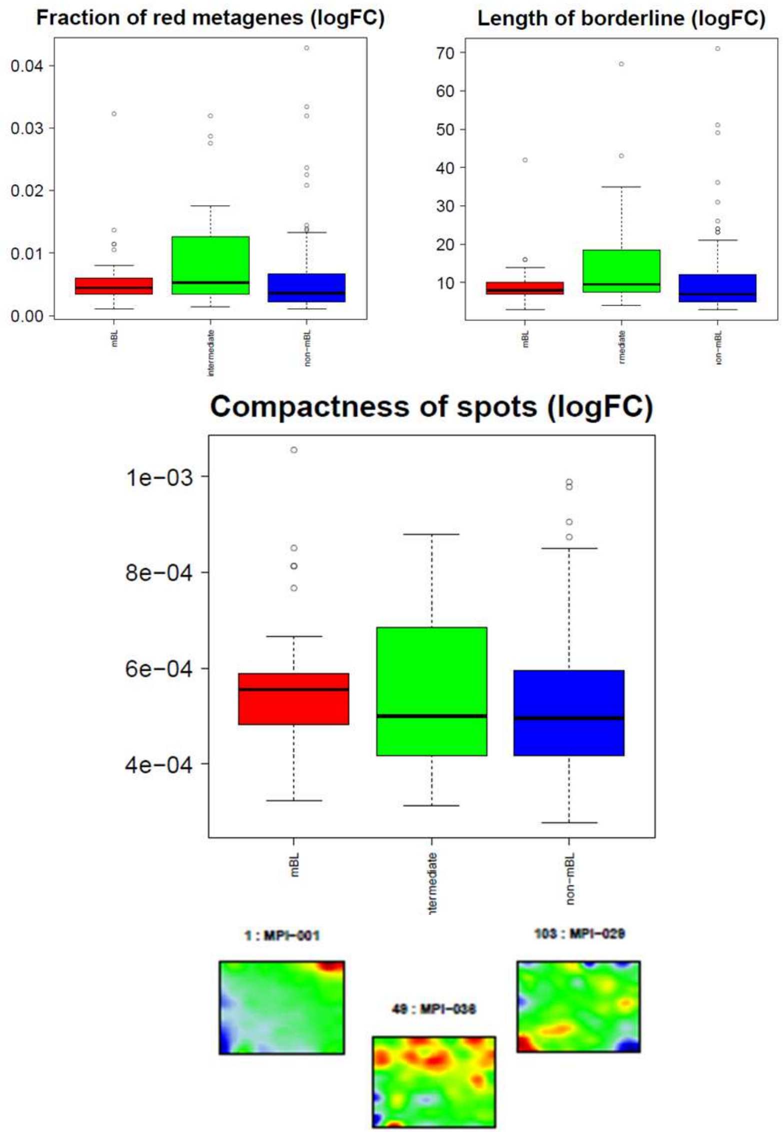
fuzziness



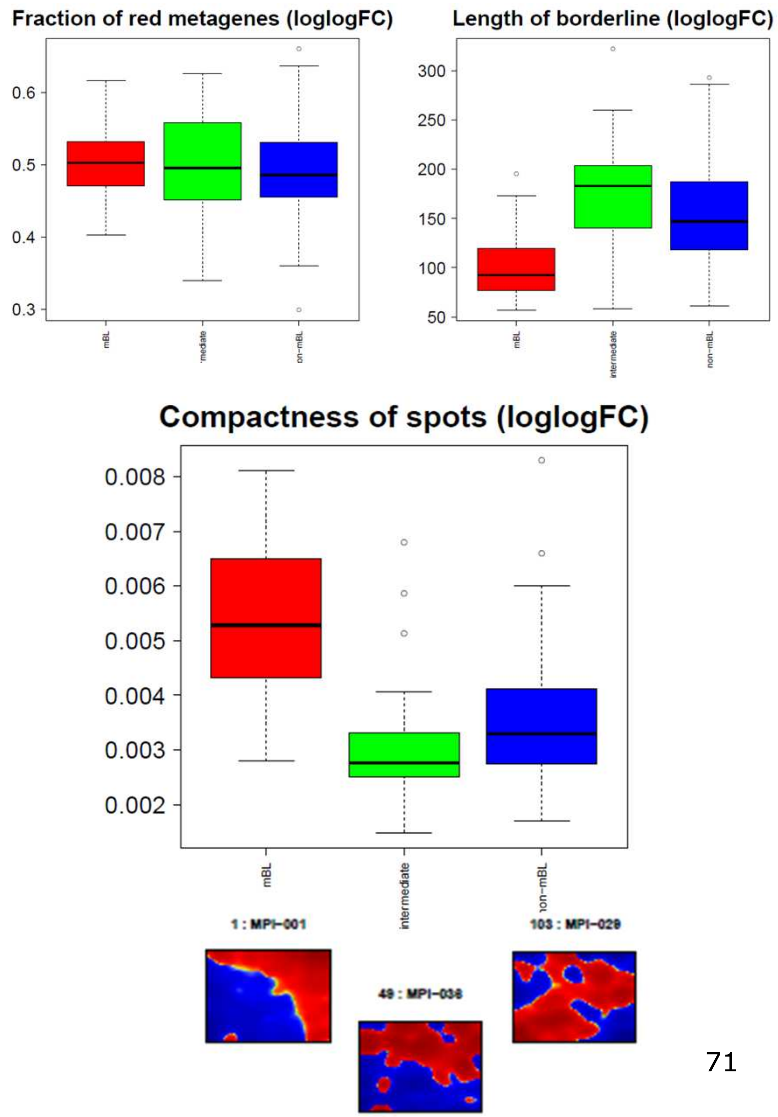


Topological measures: Lymphoma

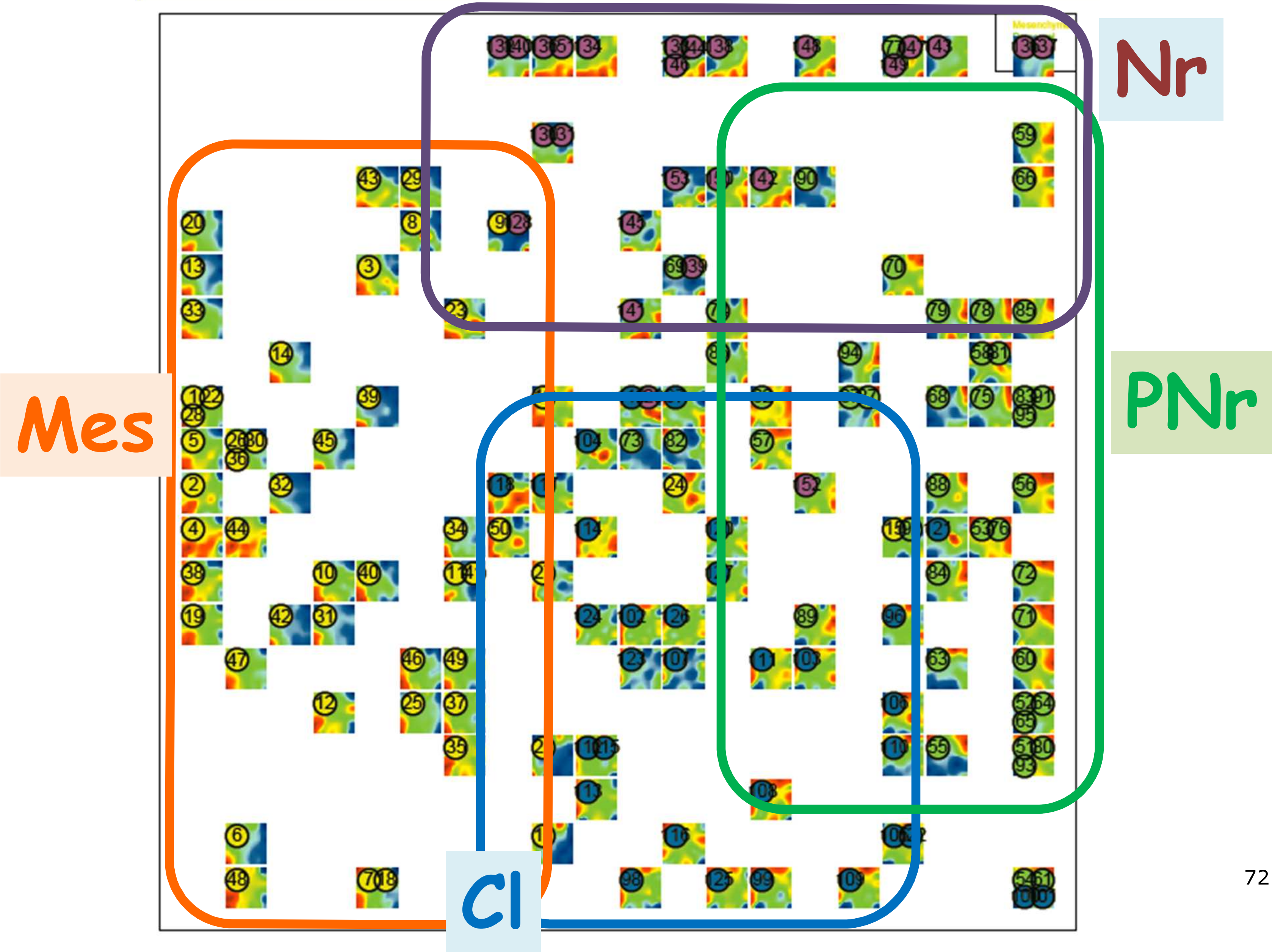
High expression



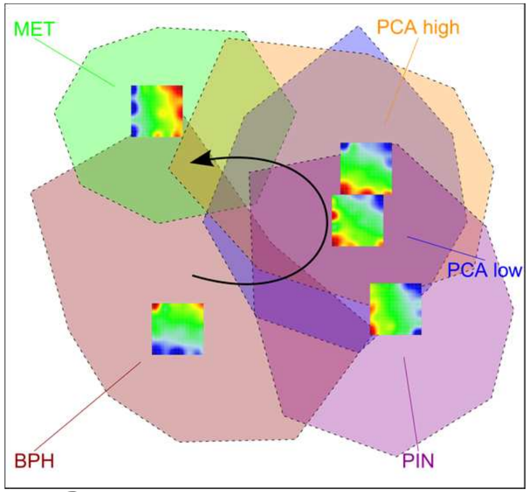
mean expression



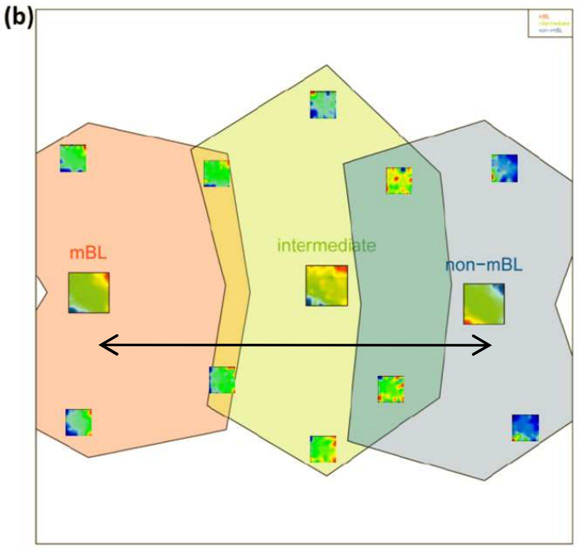
Similarity relations: 2nd level SOM



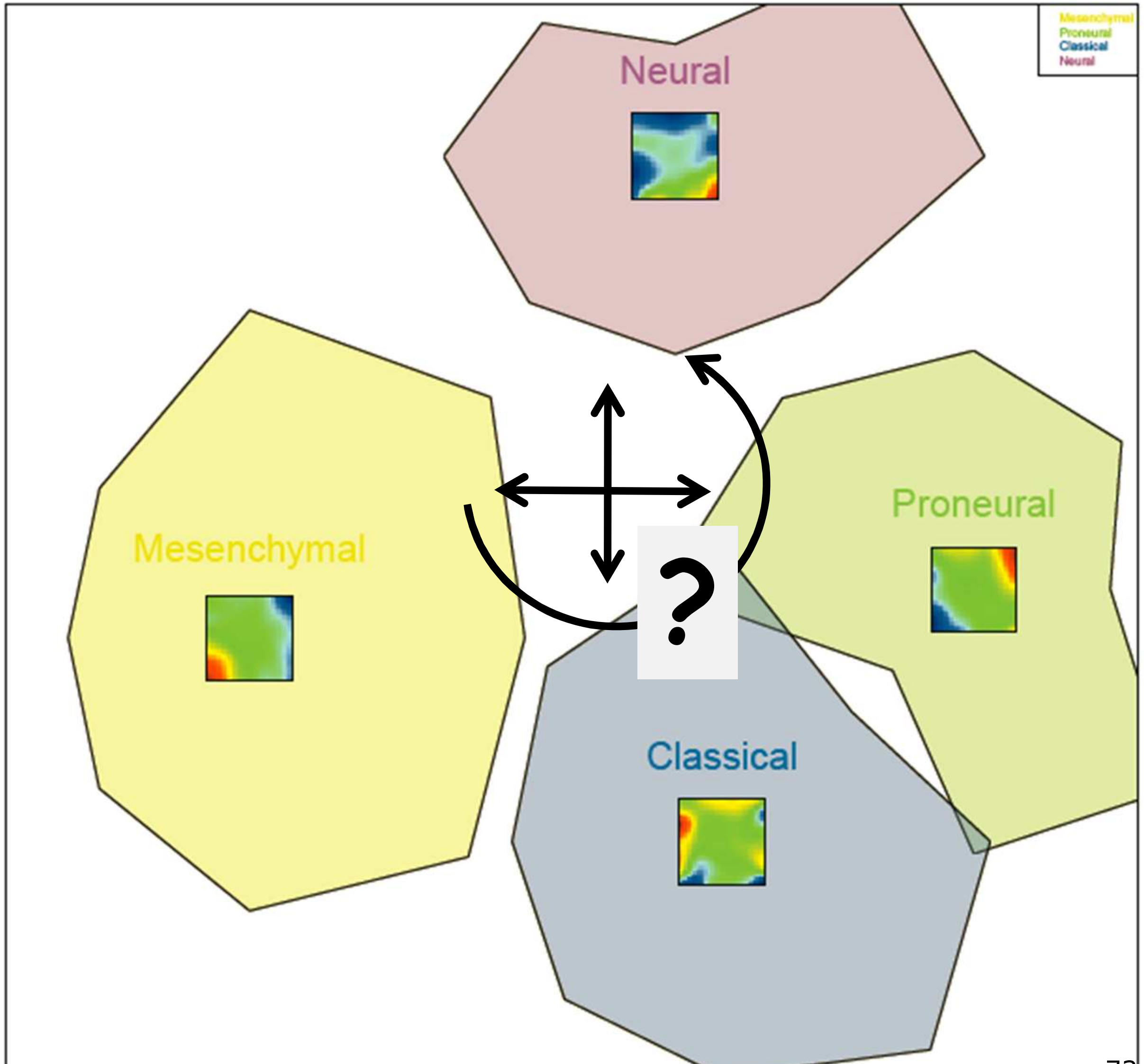
2nd level SOM: GMF



Prostate cancer progression

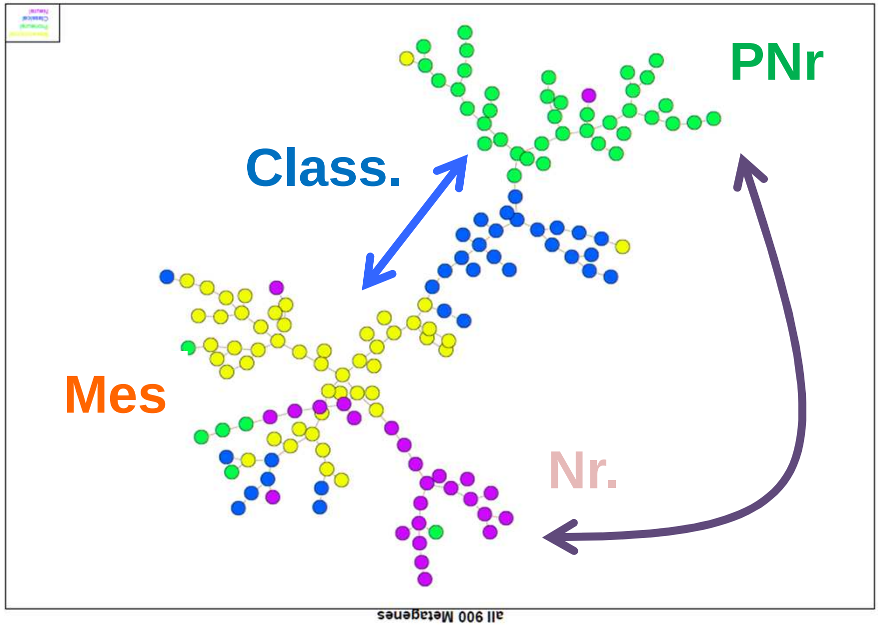


Lymphoma subtypes

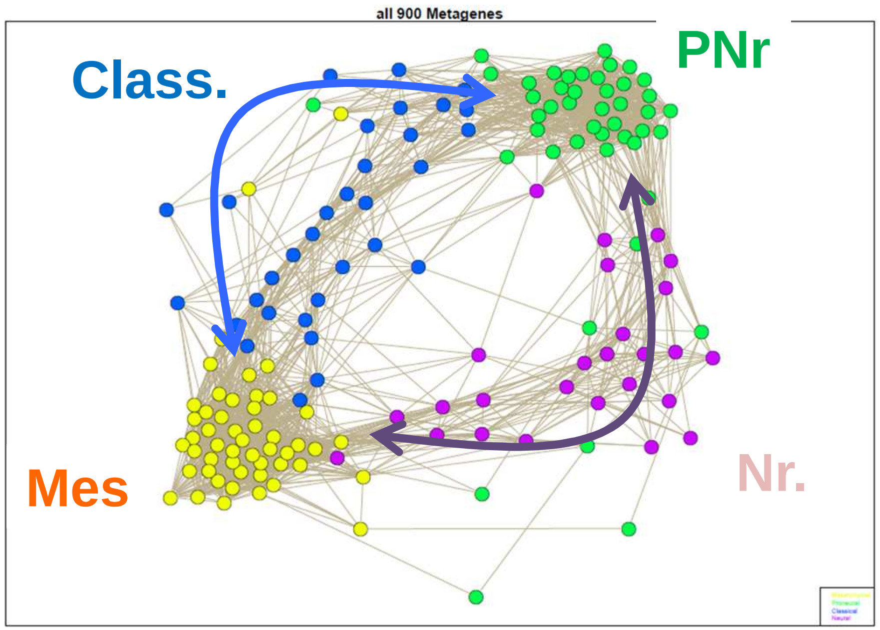


GMF subtypes

Correlation networks: GMF



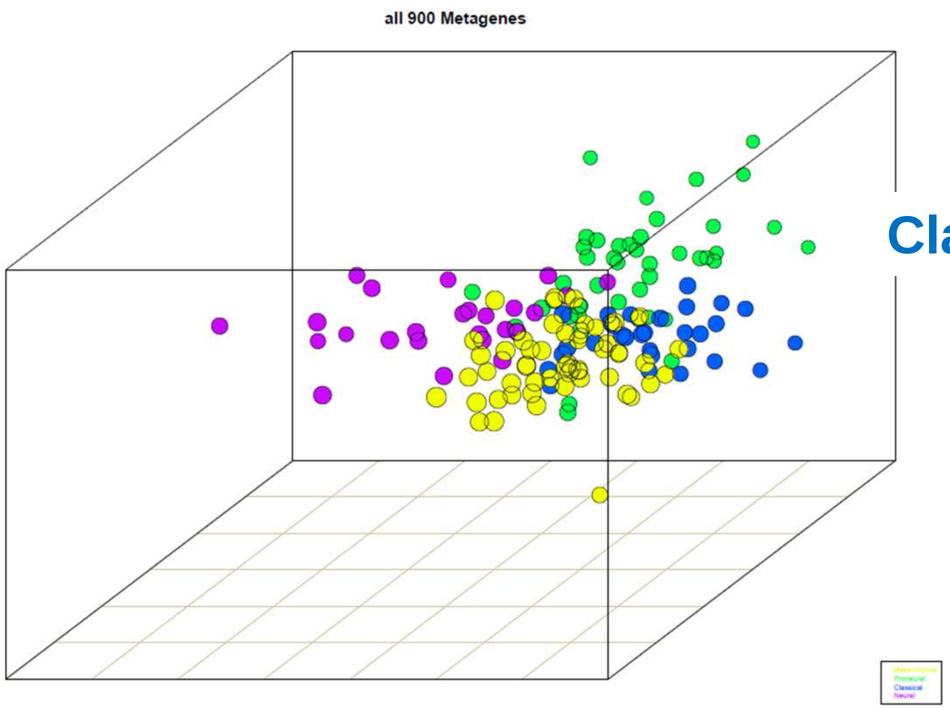
Maximum spanning tree:
Connects strongest pairwise correlations

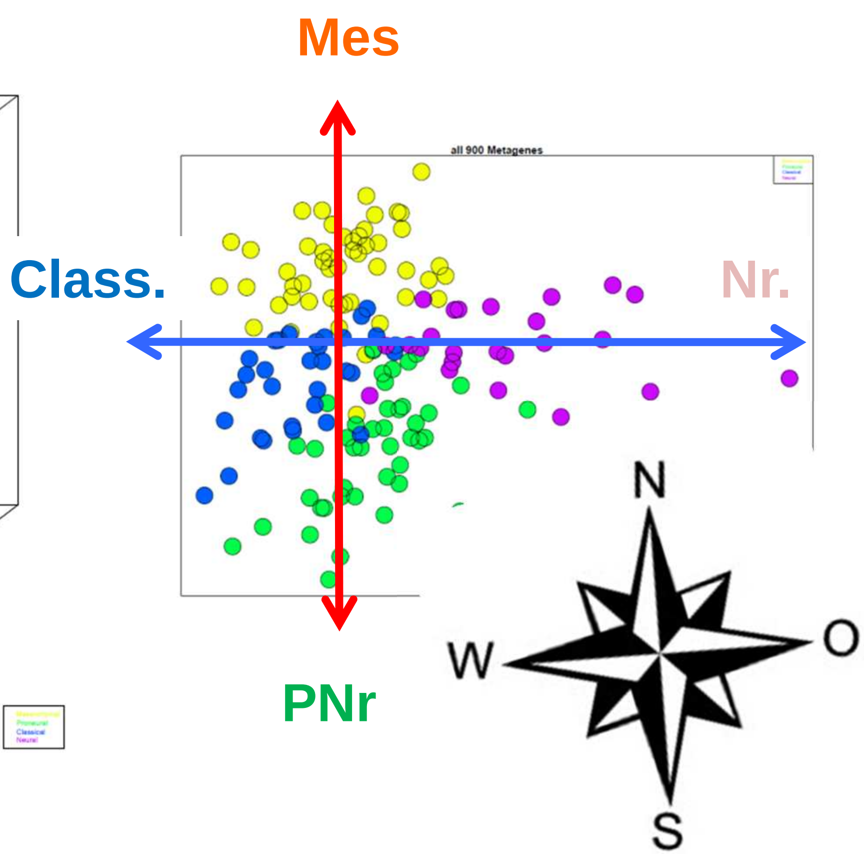


Correlation net:
Connects all samples mutually correlated with r>0.5

Class & Neur. = different intermediate pattern between Mes & PNeur.

Independent component analysis

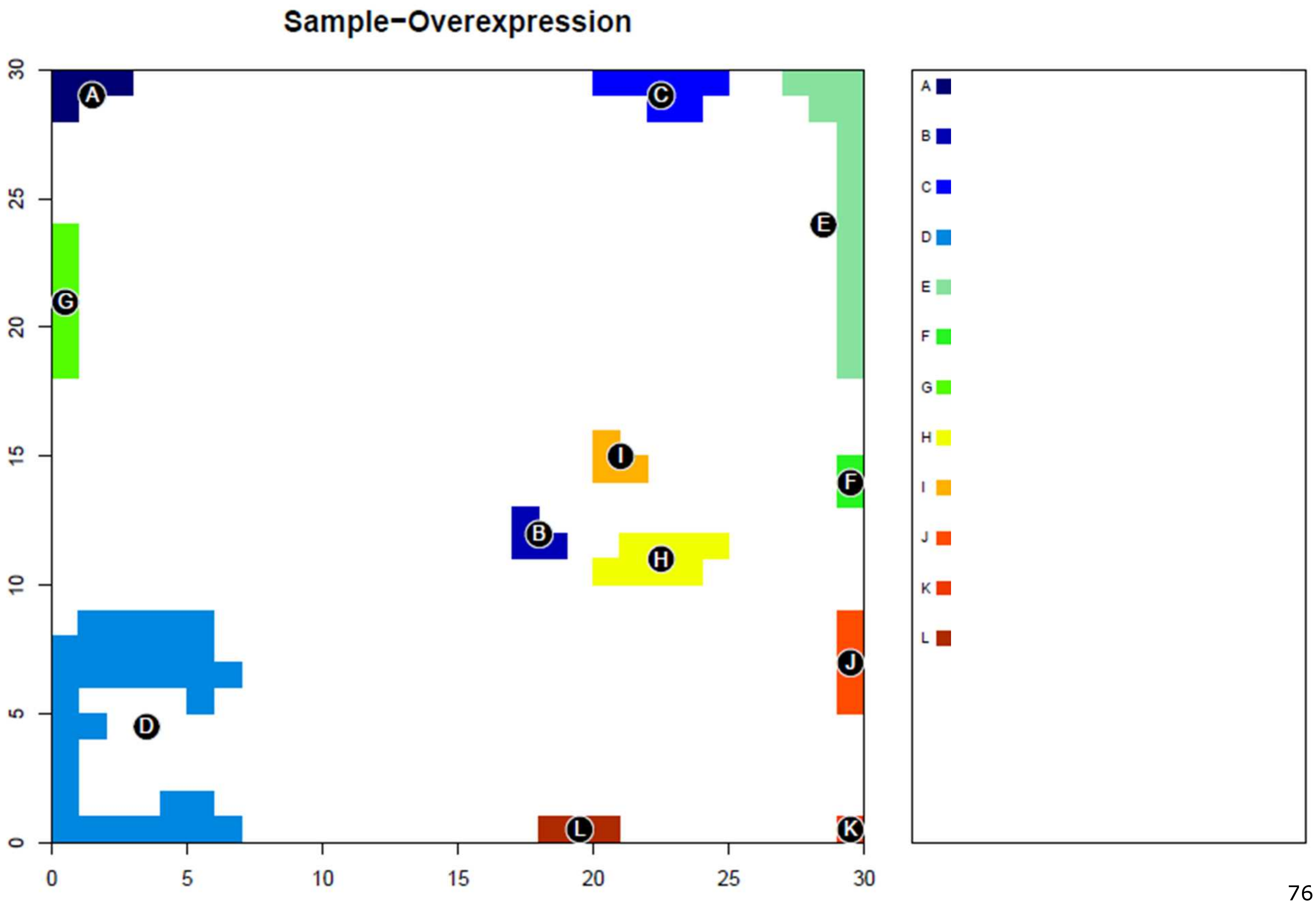




ICA: perpendicular axes → independent sets of features

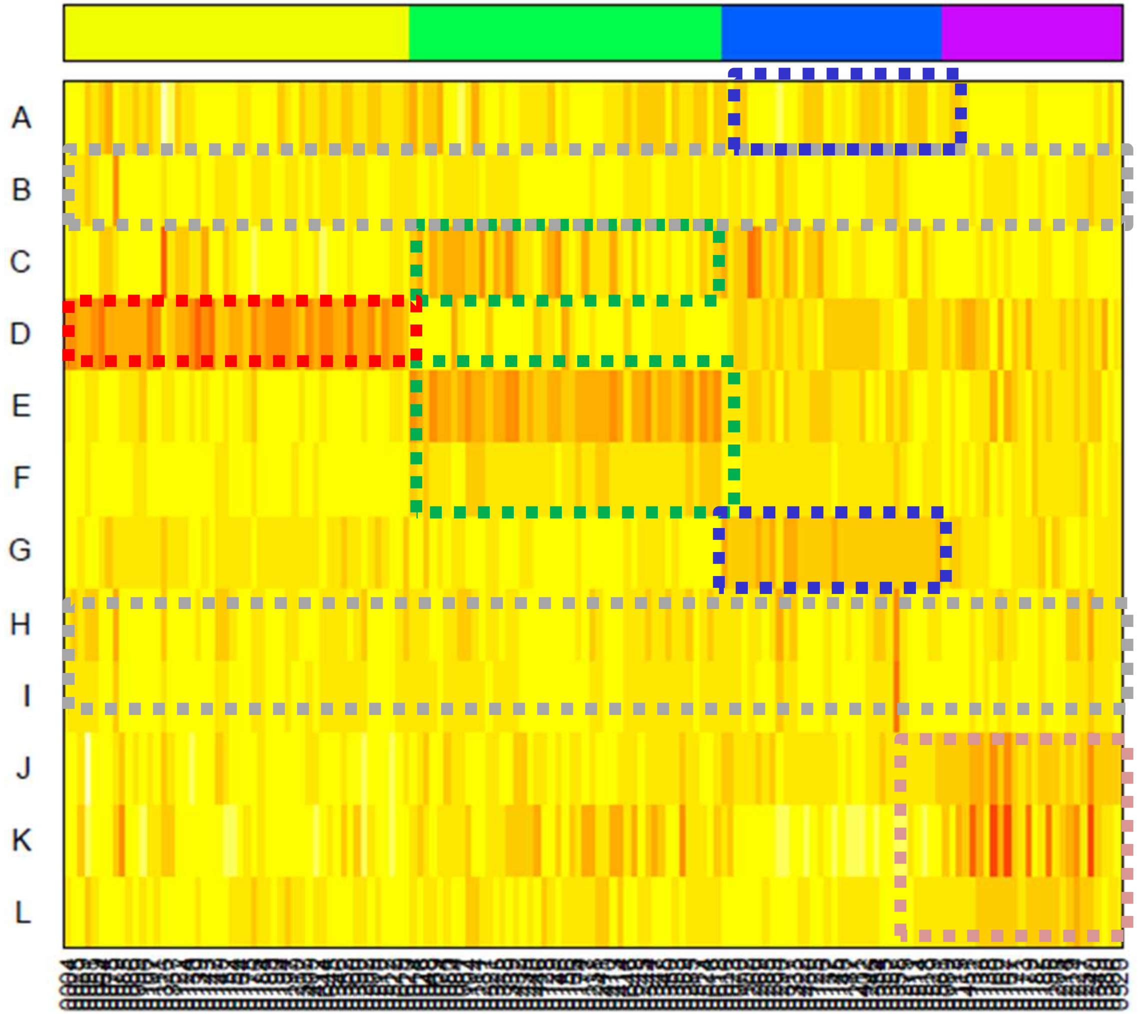
Orthogonal sets of genes (Class & Neur.) vs (Mes & Pneur)

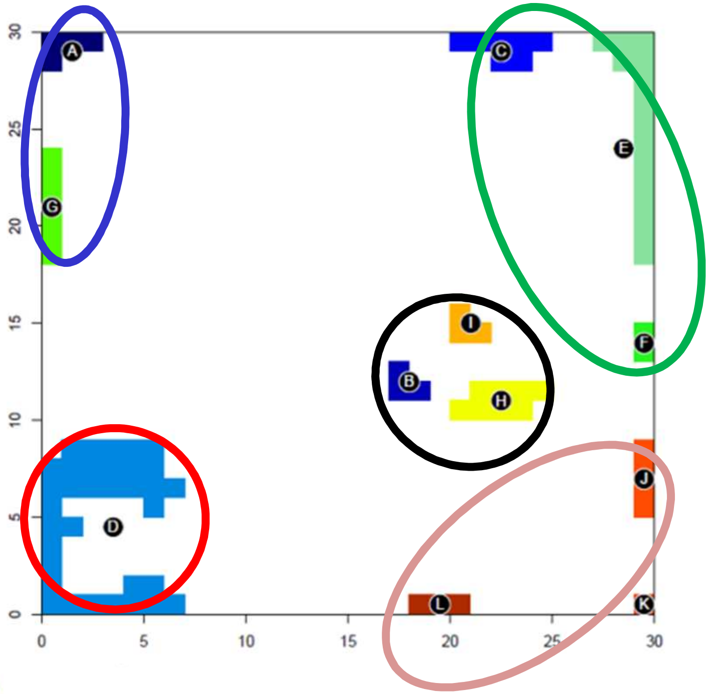
98-percentile: 12 overexpression spots: A…L



Spot heatmap: co-occurance in different subtypes

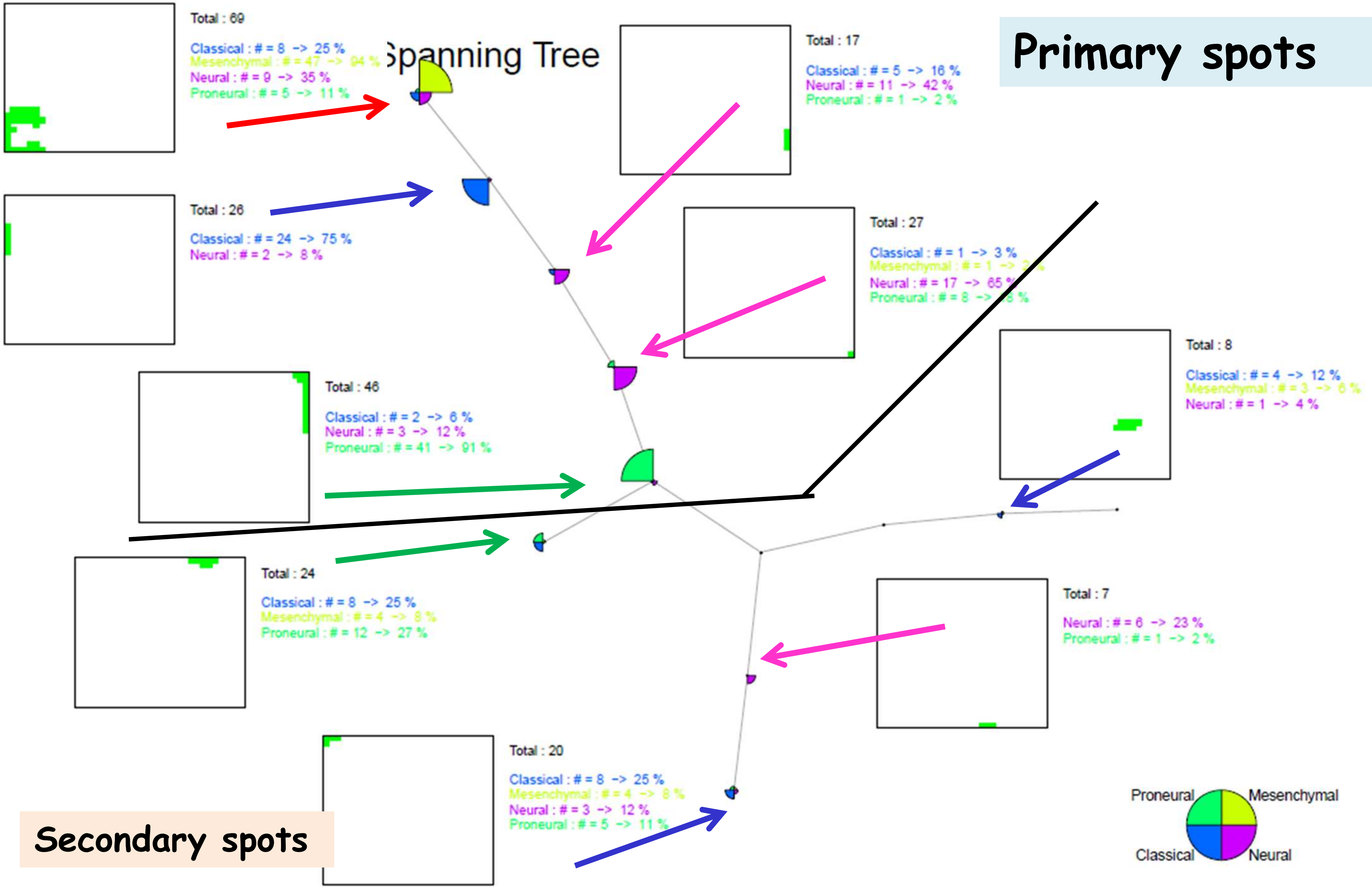
Mes PNr Class. Nr.



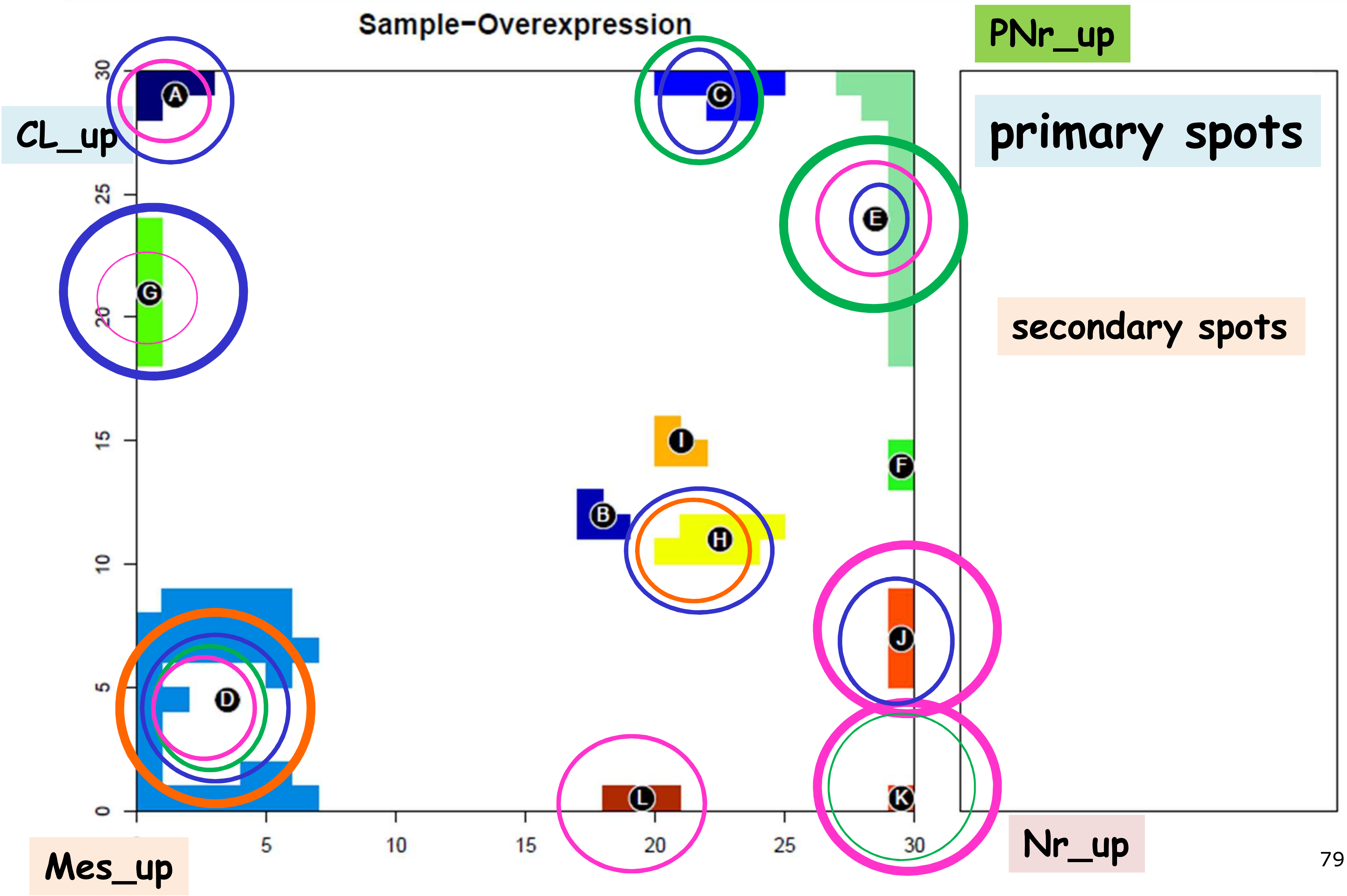


IZBI
University Leipzig

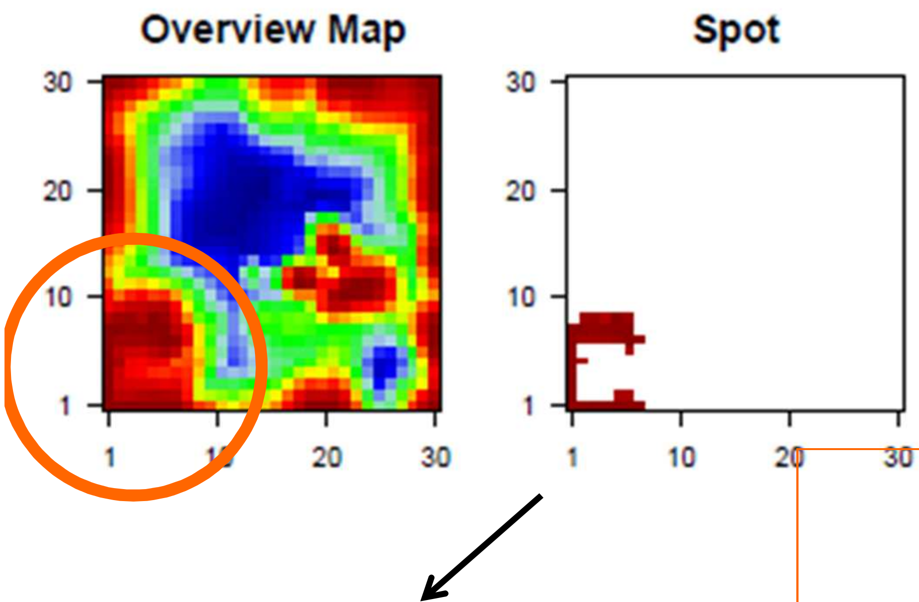
Spot-occurance tree



Spots in concert



Spot analysis



Overrepresentation:
→set-members in the spot (HG)

Overexpression:
→Expression of the set

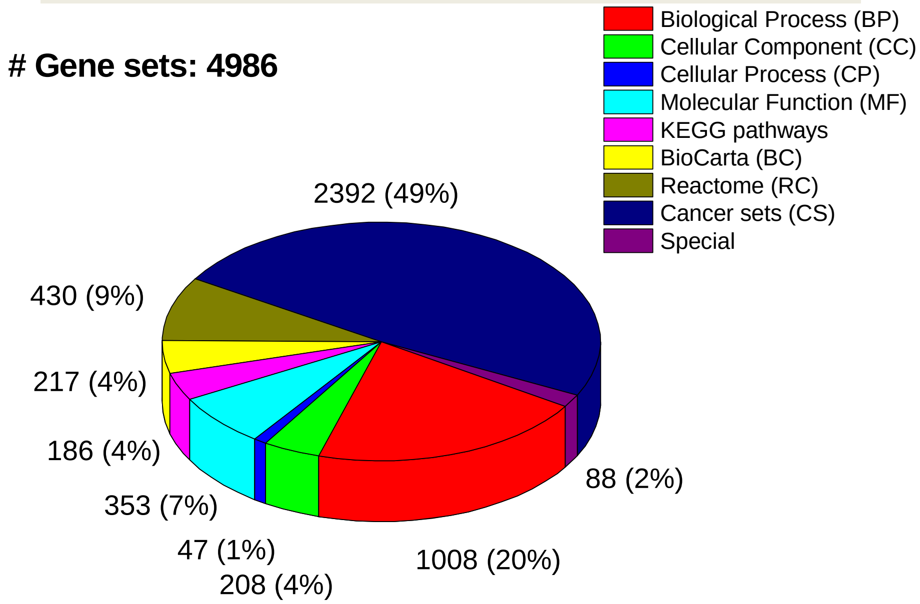
GS-Enrichment Z-score:
→ Overrepr + -expression

metagenes = 34
genes = 1043

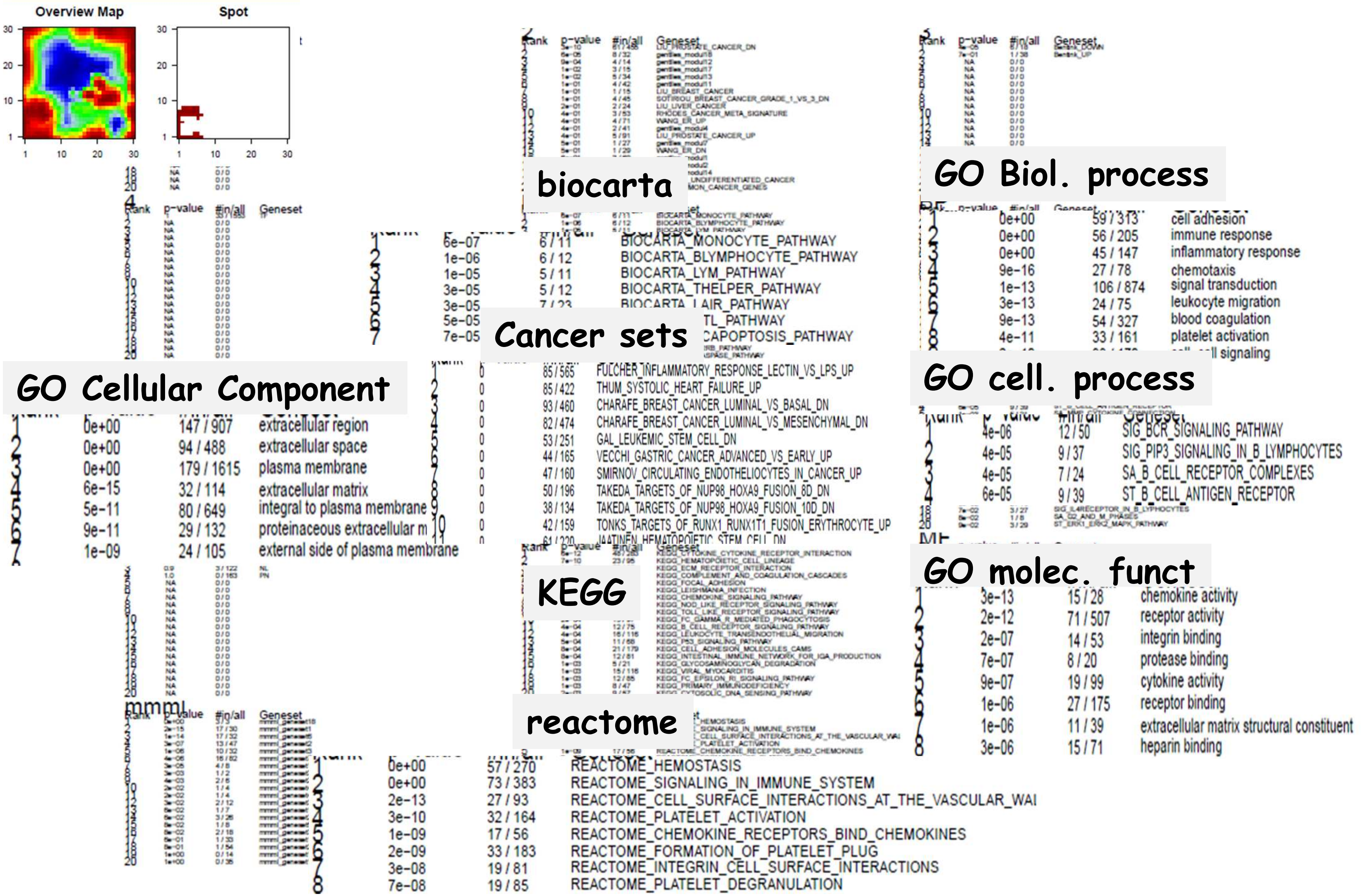
<r> metagenes = 0.85
<r> genes = 0.33

samples with spot = 69 (45.1 %)

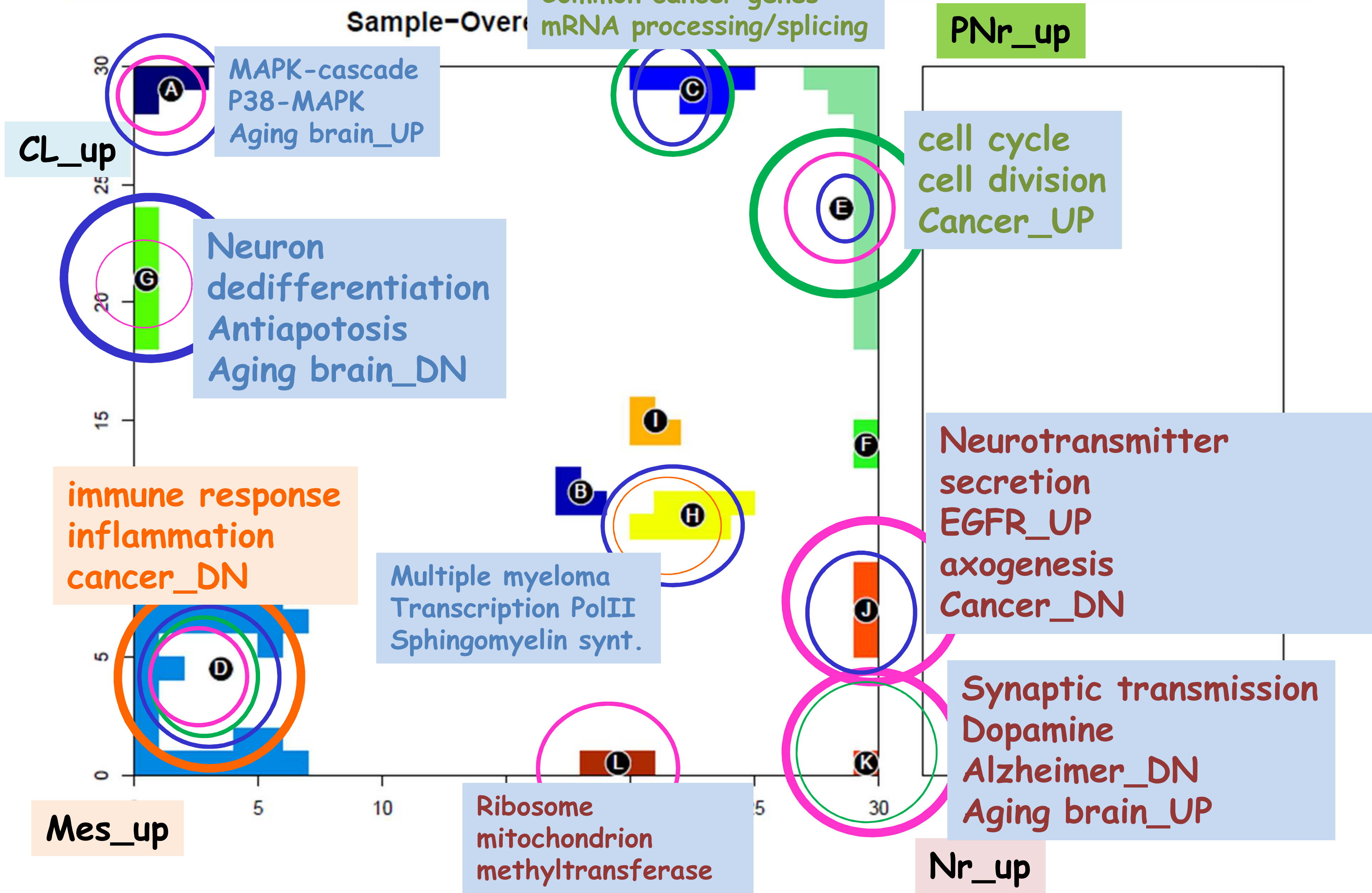
Classical : 8 (25 %)
Mesenchymal : 47 (94 %)
Neural : 9 (34.6 %)
Proneural : 5 (11.1 %)



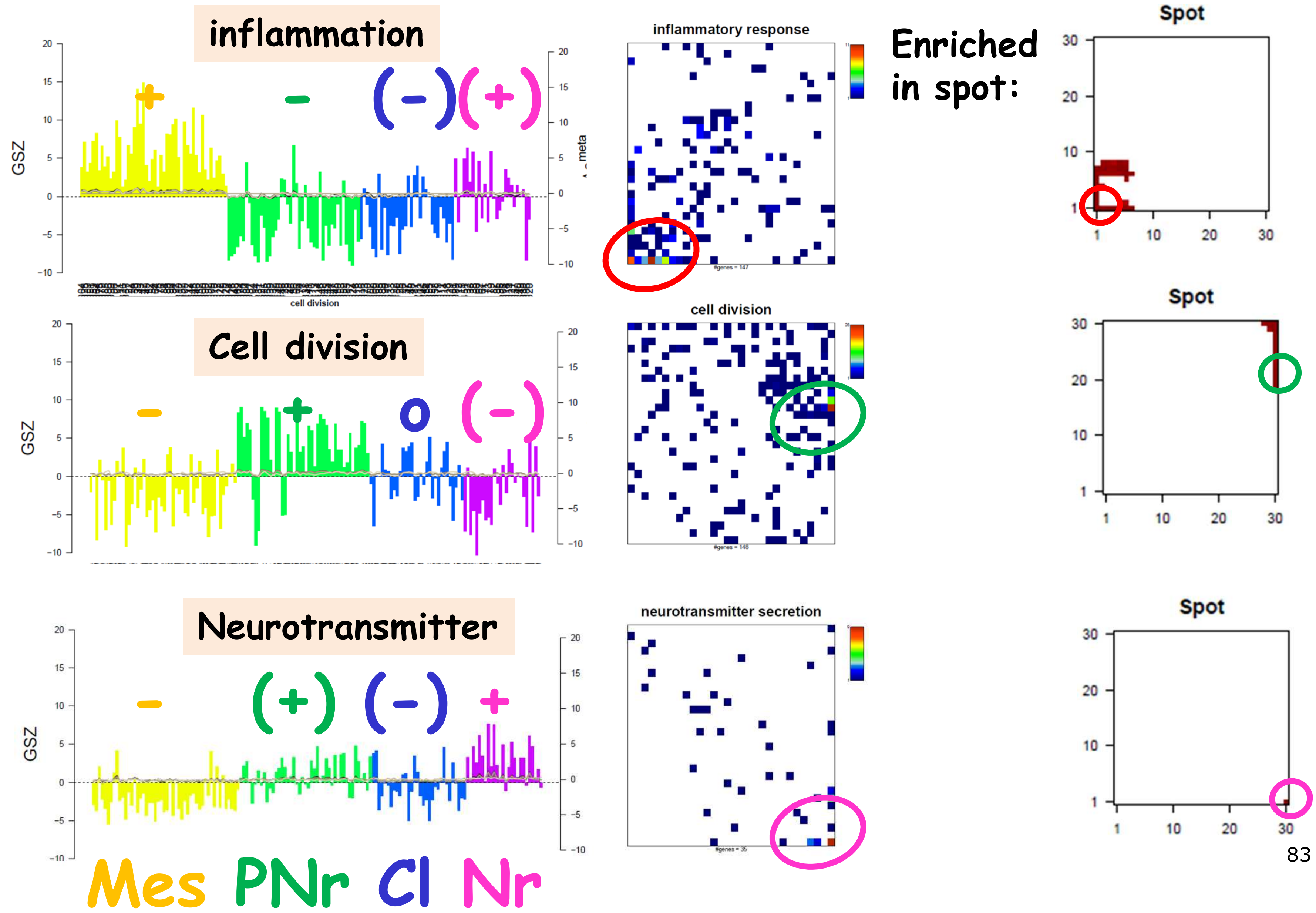
GeneSet enrichment reports

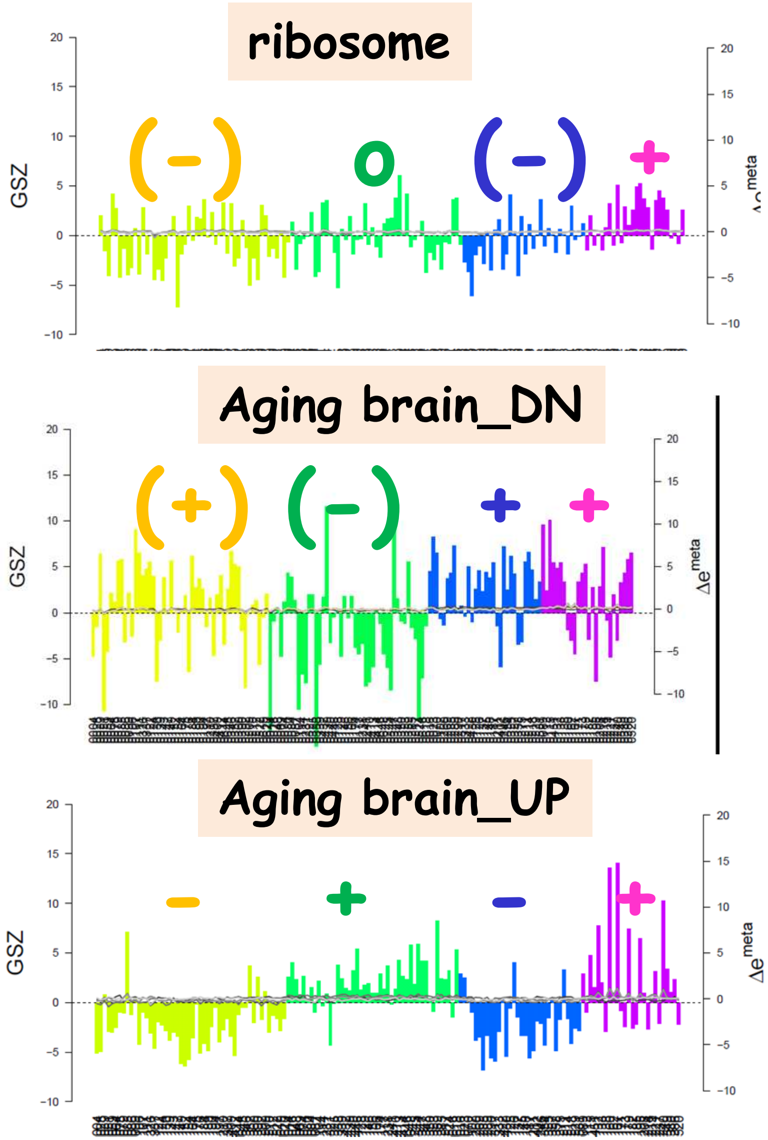


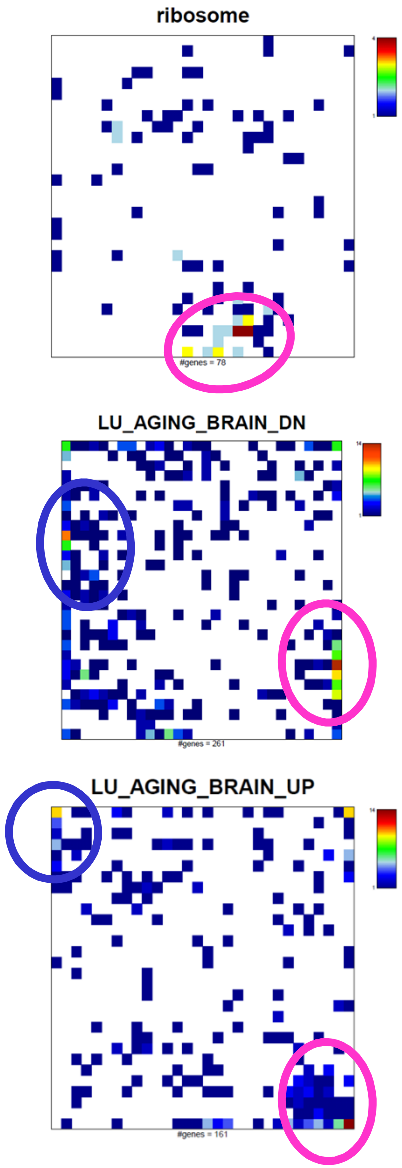
Functional assignment of s

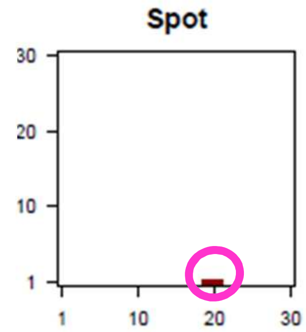


Gene set profiles and Gene set maps





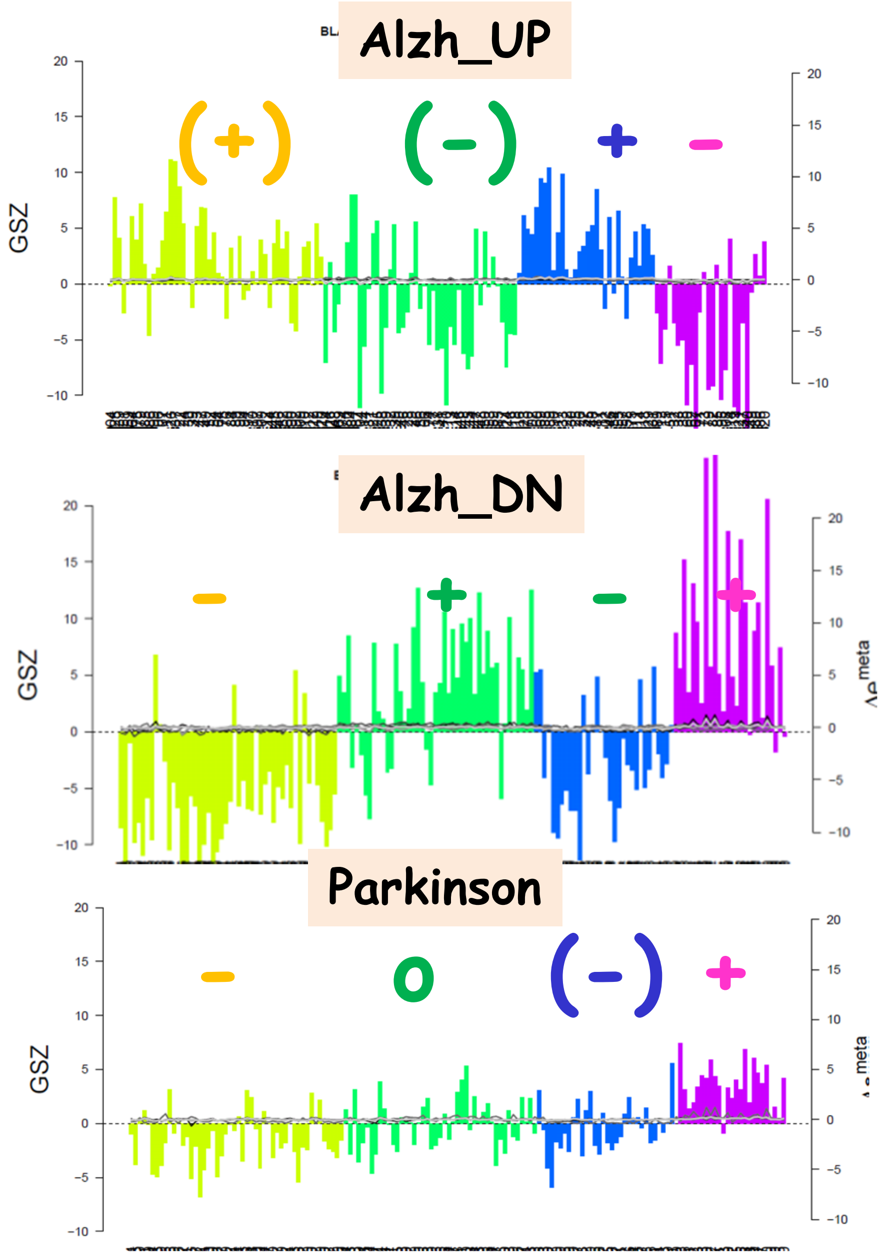


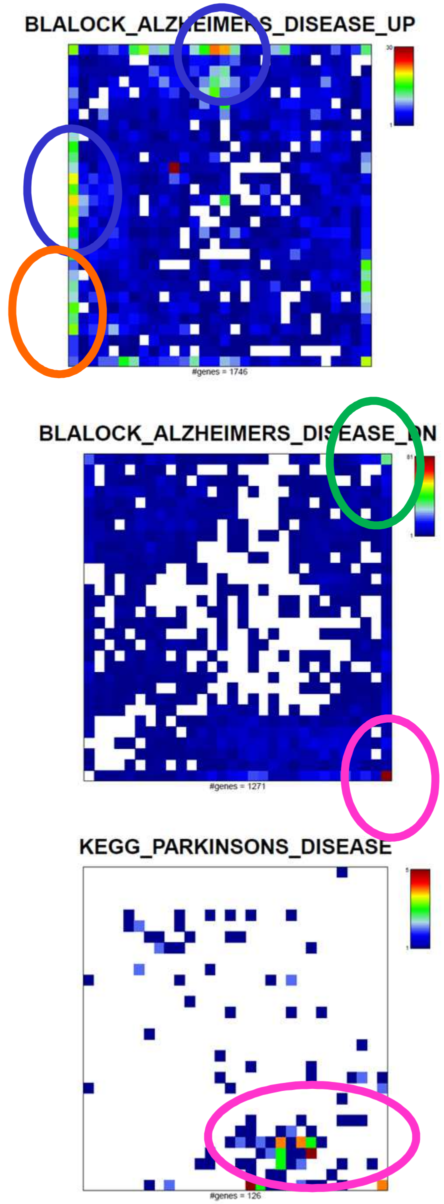


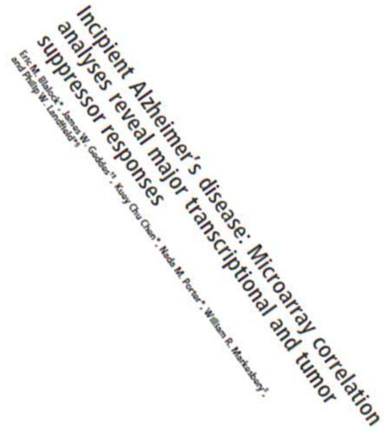
Gene regulation and DNA damage in the ageing human brain

Tao Lu[1], Ying Pan[1], Shyan-Yuan Kao[1], Cheng Li[2], Isaac Kohane[3], Jennifer Chan[4] & Bruce A. Yankner[1]

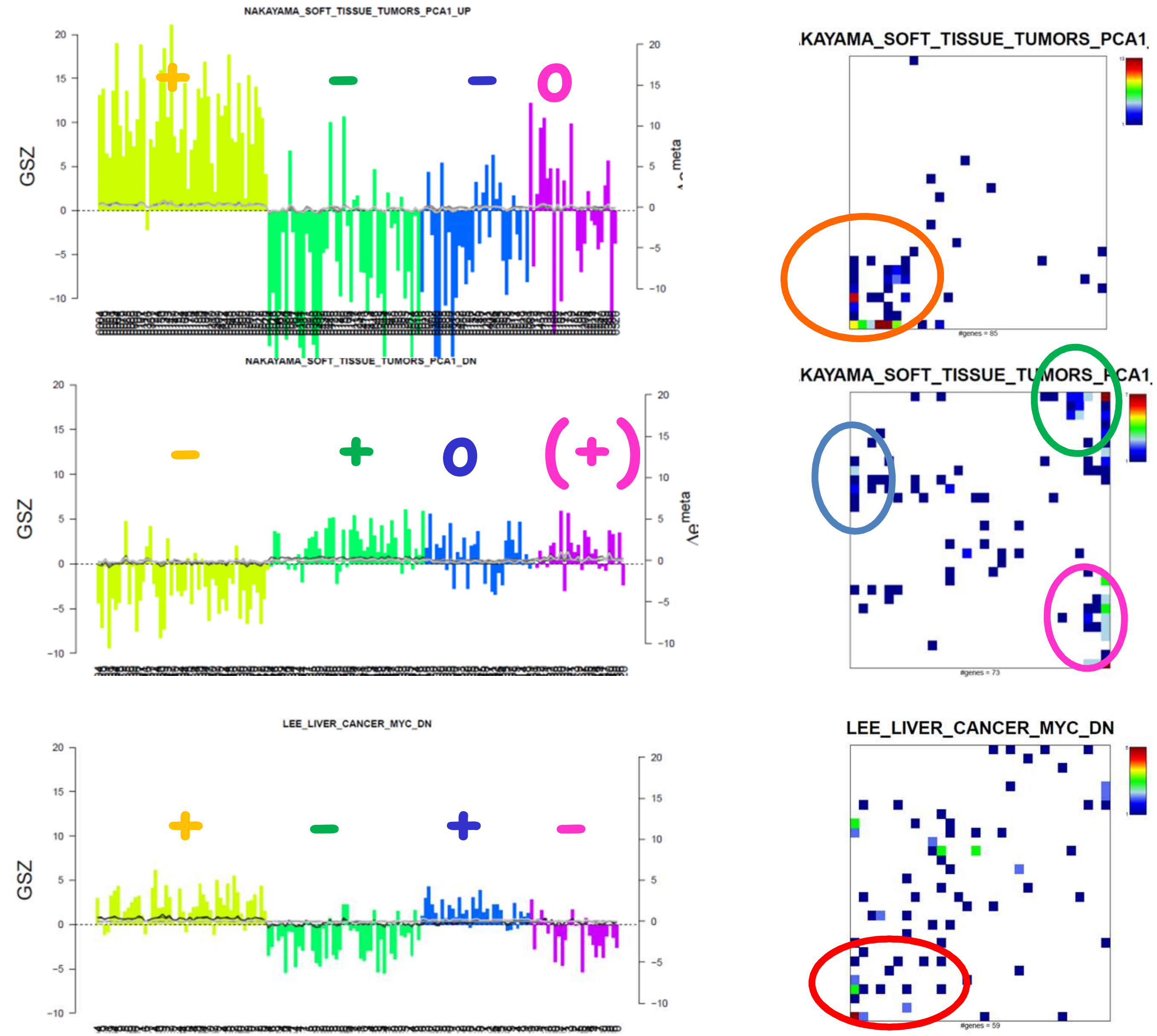
Brain diseases







Cancer

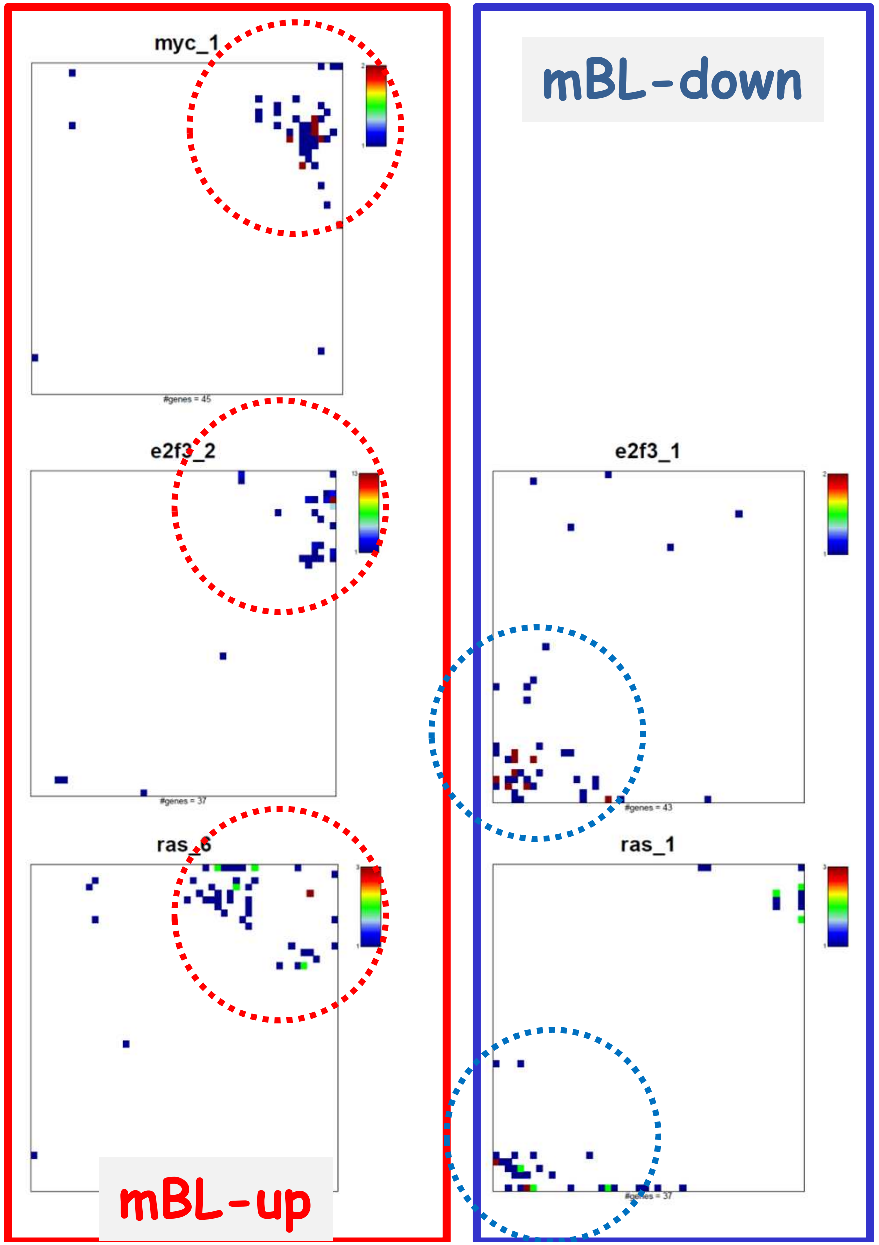


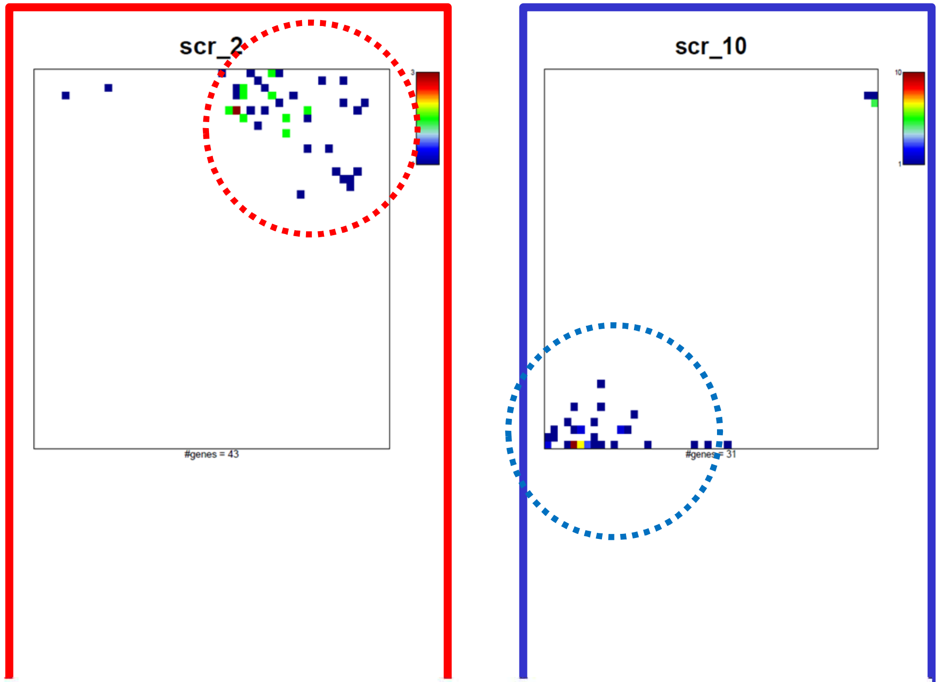
Soft tissue tumors PC1_UP

Soft tissue tumors PC1_DN

Liver Carc.

Pathway activation sets → B-cell Lymphoma





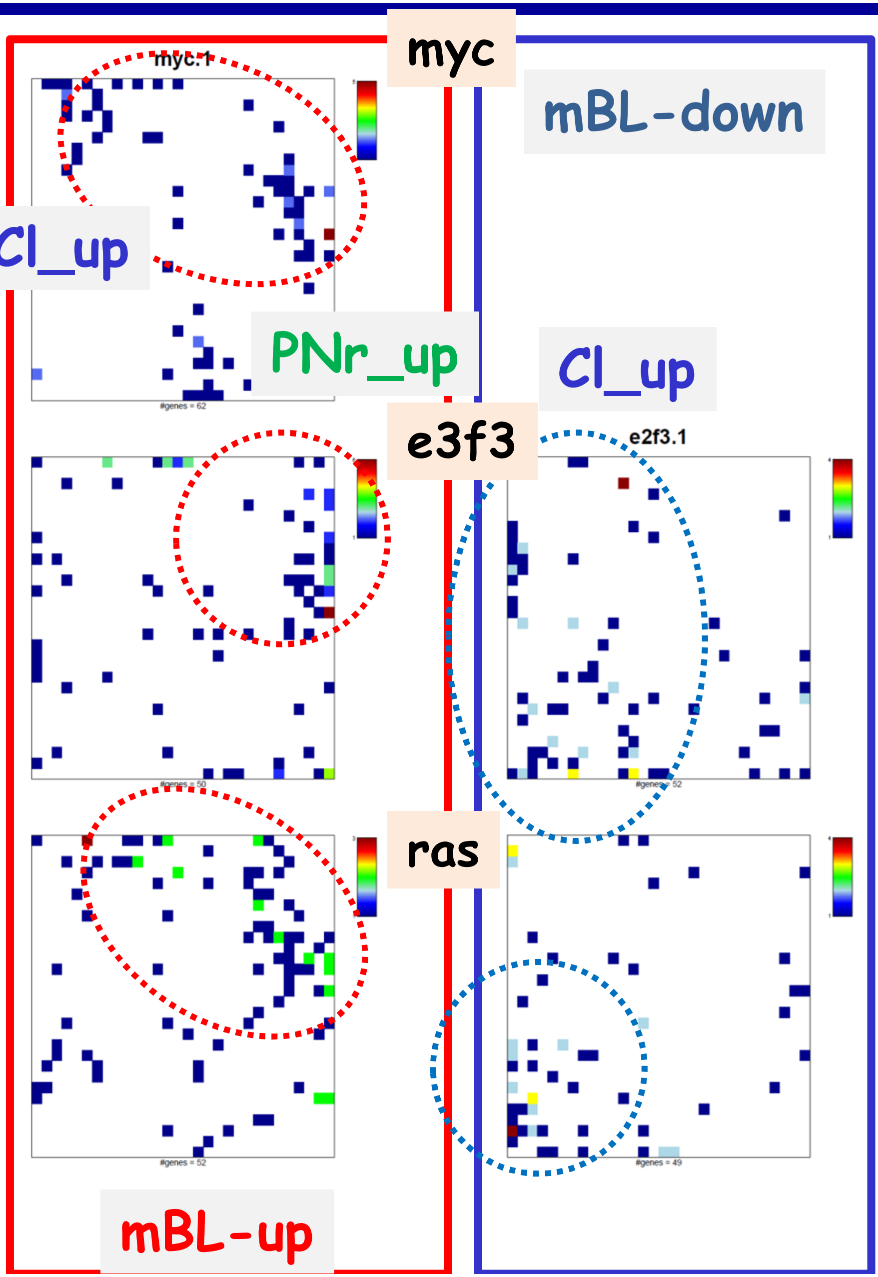
Oncogenic pathway signatures in human cancers as a guide to targeted therapies

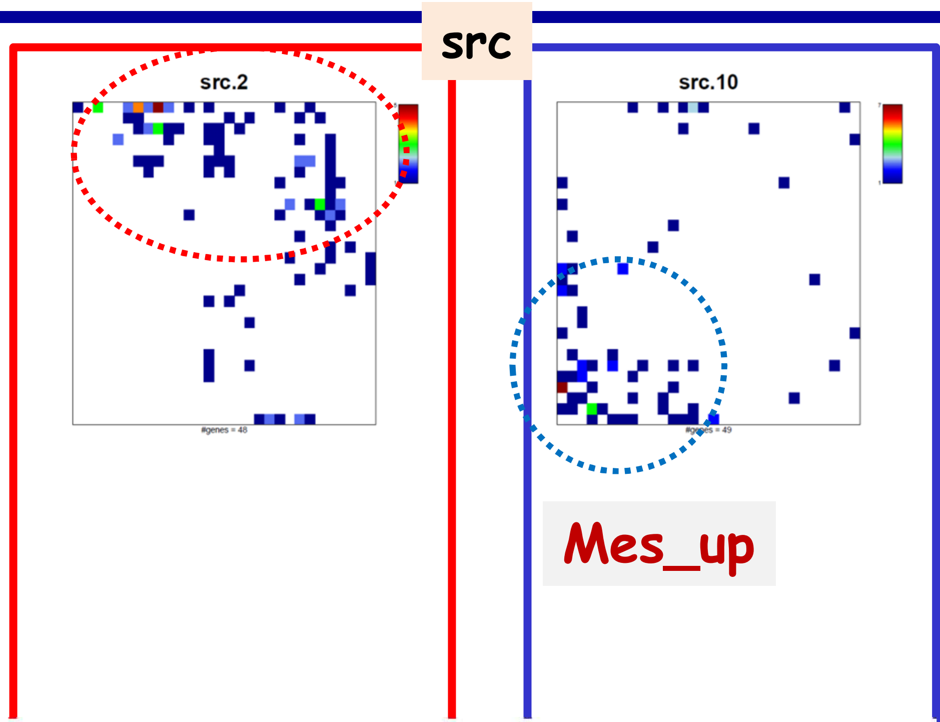
Andrea H. Bild[1,2], Guang Yao[1,2], Jeffrey T. Chang[1,2], Quanli Wang[1], Anil Potti[1,4], Dawn Chasse[1,2], Mary-Beth Joshi[3], David Harpole[3], Johnathan M. Lancaster[7], Andrew Berchuck[5], John A. Olson Jr[1,3], Jeffrey R. Marks[3], Holly K. Dressman[1,2], Mike West[6] & Joseph R. Nevins[1,2]

Pathway activation patterns in diffuse large B-cell lymphomas

S Bentink[1], S Wessendorf[2], C Schwaenen[2], M Rosolowski[3], W Klapper[4], A Rosenwald[5], G Ot
AC Feller[7], M-L Hansmann[8], D Hasenclever[3], M Hummel[9], D Lenze[9], P Möller[10], B Stuerzen
L Truemper[11], H Stein[9], R Siebert[12] and R Spang[1] for the Molecular Mechanisms in Malignant Ly
Deutsche Krebshilfe

Pathway activation sets → GMF





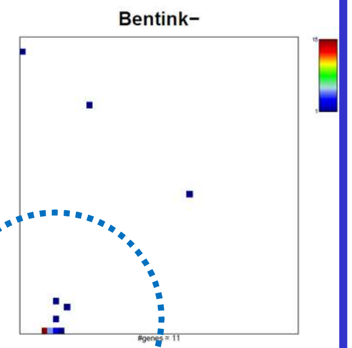
Oncogenic pathway signatures in human cancers as a guide to targeted therapies

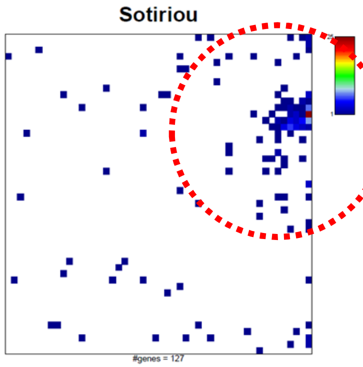
Andrea H. Bild[1,2], Guang Yao[1,2], Jeffrey T. Chang[1,2], Quanli Wang[1], Anil Potti[1,4], Dawn Chasse[1,2], Mary-Beth Joshi[3], David Harpole[3], Johnathan M. Lancaster[7], Andrew Berchuck[5], John A. Olson Jr[1,3], Jeffrey R. Marks[3], Holly K. Dressman[1,2], Mike West[6] & Joseph R. Nevins[1,2]

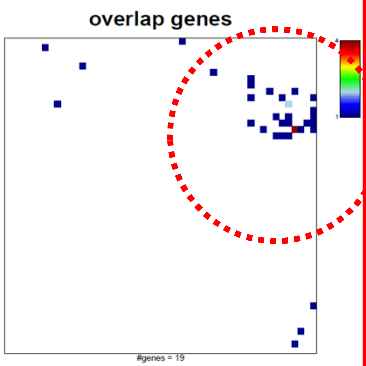
Pathway activation patterns in diffuse large B-cell lymphomas

S Bentink[1], S Wessendorf[2], C Schwaenen[2], M Rosolowski[3], W Klapper[4], A Rosenwald[5], G Ot
AC Feller[7], M-L Hansmann[8], D Hasenclever[3], M Hummel[9], D Lenze[9], P Möller[10], B Stuerzen
L Truemper[11], H Stein[9], R Siebert[12] and R Spang[1] for the Molecular Mechanisms in Malignant Ly
Deutsche Krebshilfe

Cancer sets → B-cell Lymphoma







The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JUNE 8, 2006 VOL. 354 NO. 23

A Biologic Definition of Burkitt's Lymphoma from Transcriptional and Genomic Profiling

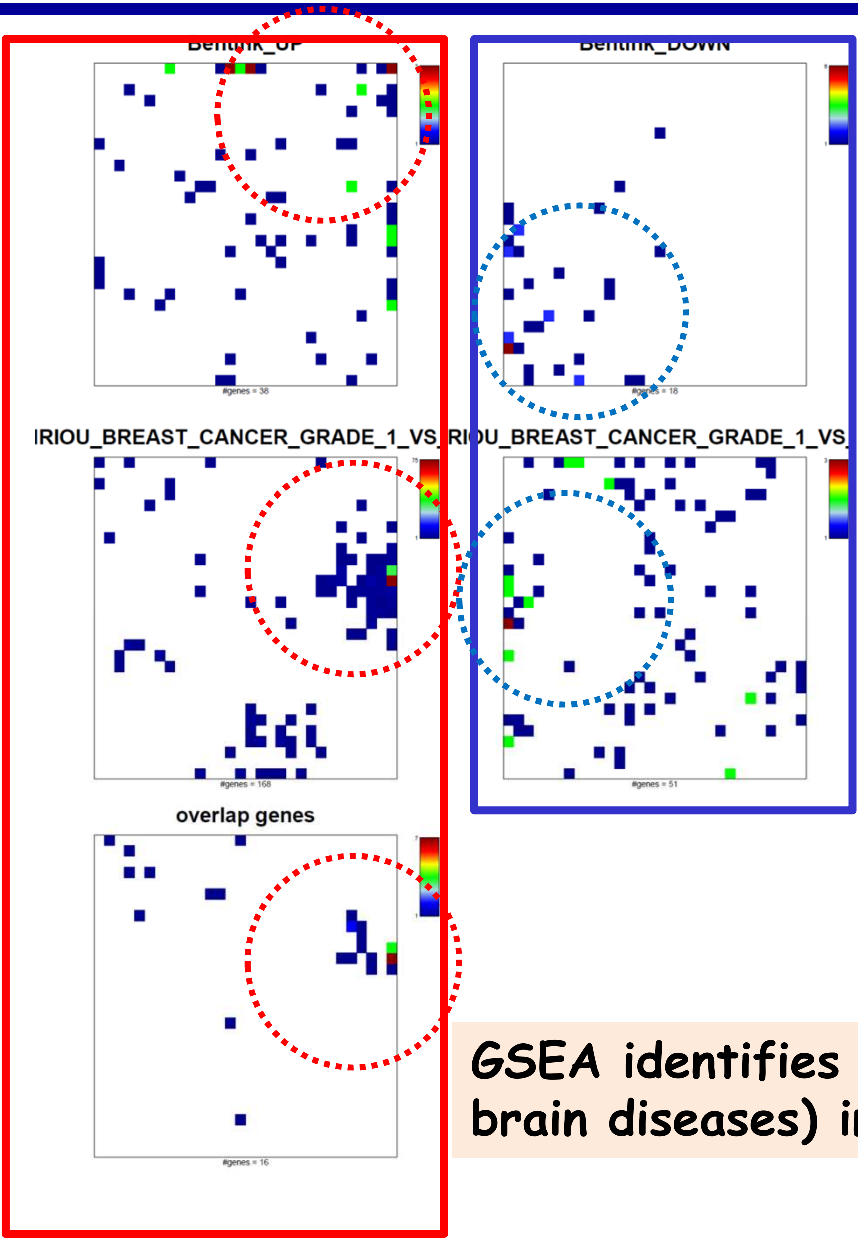
Michael Hummel, Ph.D., Stefan Bentink, M.S., Hilmar Berger, M.D., Wolfram Klapper, M.D., Swen Wessendorf, M.D., Thomas F.E. Barth, M.D., Heinz-Wolfram Bernd, M.D., Sergio B. Cogliatti, M.D., Judith Dierlamm, M.D., Ph.D., Alfred C. Feller, M.D., Martin-Leo Hansmann, M.D., Eugenia Haralambieva, M.D., Lana Harder, M.D.,

Sotiriou C, et al. (2006) Gene expression profiling in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 98: 262–272.

MYC regulation of a "poor-prognosis" metastatic cancer cell state

Anita Wolfer[a,b,1], Ben S. Wittner[a,b,1], Daniel Irimia[b,c,d], Richard J. Flavin[b,e], Mathieu Lupien[b,f,2], Ruwanthi N. Gunawardane[b,3], Clifford A. Meyer[g], Eric S. Lightcap[h], Pablo Tamayo[i], Jill P. Mesirov[i], X. Shirley Liu[g],

Cancer sets → GMF



Cancer Cell
Article

Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*

Roel G.W. Verhaak,[1,2,17] Katherine A. Hoadley,[3,4,17] Elizabeth Purdom,[7] Victoria Wang,[8] Yuan Qi,[4,5] Matthew D. Wilkerson,[4,5] C. Ryan Miller,[4,6] Li Ding,[9] Todd Golub,[1,10] Jill P. Mesirov,[1] Gabriele Alexe,[1] Michael Lawrence,[1,2] Michael O'Kelly,[1,2] Pablo Tamayo,[1] Barbara A. Weir,[1,2] Stacey Gabriel,[1] Wendy Winckler,[1,2] Supriya Gupta,[1] Lakshmi Jakkula,[11] Heidi S. Feiler,[11] J. Graeme Hodgson,[12] C. David James,[12] Jann N. Sarkaria,[13] Cameron Brennan,[14] Ari Kahn,[15] Paul T. Spellman,[11] Richard K. Wilson,[9] Terence P. Speed,[7,16] Joe W. Gray,[11] Matthew Meyerson,[1,2] Gad Getz,[1] Charles M. Perou,[3,4,8] D. Neil Hayes,[4,5,*] and The Cancer Genome Atlas Research Network

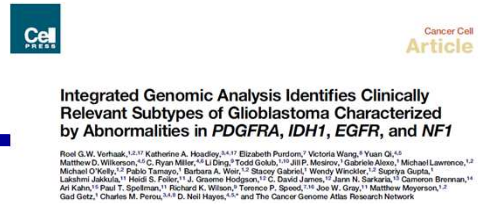
Sotiriou C, et al. (2006) Gene expression profiling in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 98: 262–272.

MYC regulation of a "poor-prognosis" metastatic cancer cell state

Anita Wolfer[a,b,1], Ben S. Wittner[a,b,1], Daniel Irimia[b,c,d], Richard J. Flavin[b,e], Mathieu Lupien[b,f,2], Ruwanthi N. Gunawardane[b,3], Clifford A. Meyer[g], Eric S. Lightcap[h], Pablo Tamayo[i], Jill P. Mesirov[i], X. Shirley Liu[g],

GSEA identifies a large numbers of signatures (cancer, brain diseases) in GMF with subtype-specific occurance

Subtype-specific gene sets

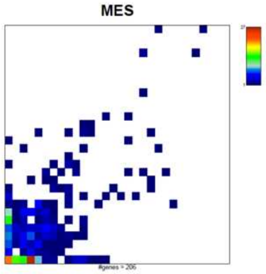


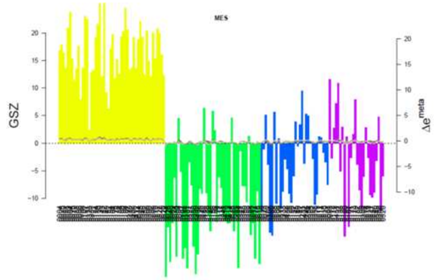
Cancer Cell
Article

Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*

Roel G.W. Verhaak,[1,2,17] Katherine A. Hoadley,[3,4,17] Elizabeth Purdom,[7] Victoria Wang,[8] Yuan Qi,[4,5] Matthew D. Wilkerson,[4,5] C. Ryan Miller,[4,6] Li Ding,[9] Todd Golub,[1,10] Jill P. Mesirov,[1] Gabriele Alexe,[1] Michael Lawrence,[1,2] Michael O'Kelly,[1,2] Pablo Tamayo,[1] Barbara A. Weir,[1,2] Stacey Gabriel,[1] Wendy Winckler,[1,2] Supriya Gupta,[1] Lakshmi Jakkula,[11] Heidi S. Feiler,[11] J. Graeme Hodgson,[12] C. David James,[12] Jann N. Sarkaria,[13] Cameron Brennan,[14] Ari Kahn,[15] Paul T. Spellman,[11] Richard K. Wilson,[9] Terence P. Speed,[7,16] Joe W. Gray,[11] Matthew Meyerson,[1,2] Gad Getz,[1] Charles M. Perou,[3,4,8] D. Neil Hayes,[4,5,*] and The Cancer Genome Atlas Research Network

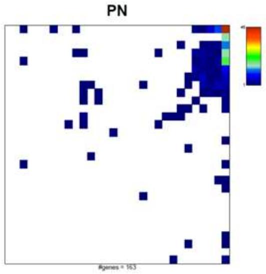
Mes_up

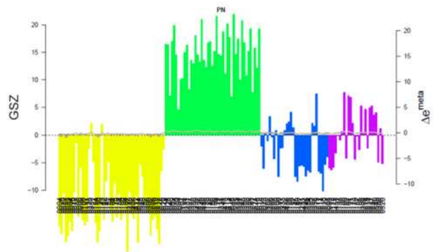




+ - (-) (-)

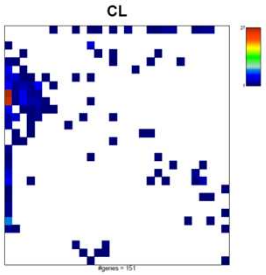
PNr_up

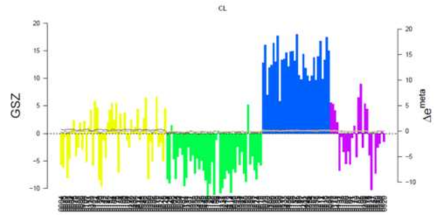




- + (-) (+)

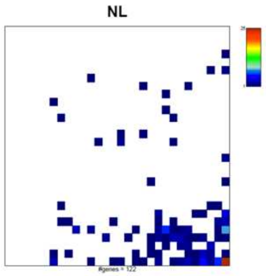
CL_up

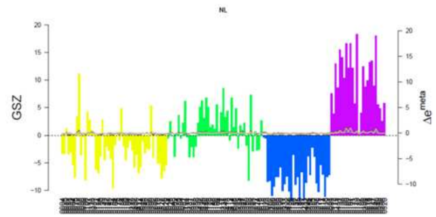




(-) - + (-)

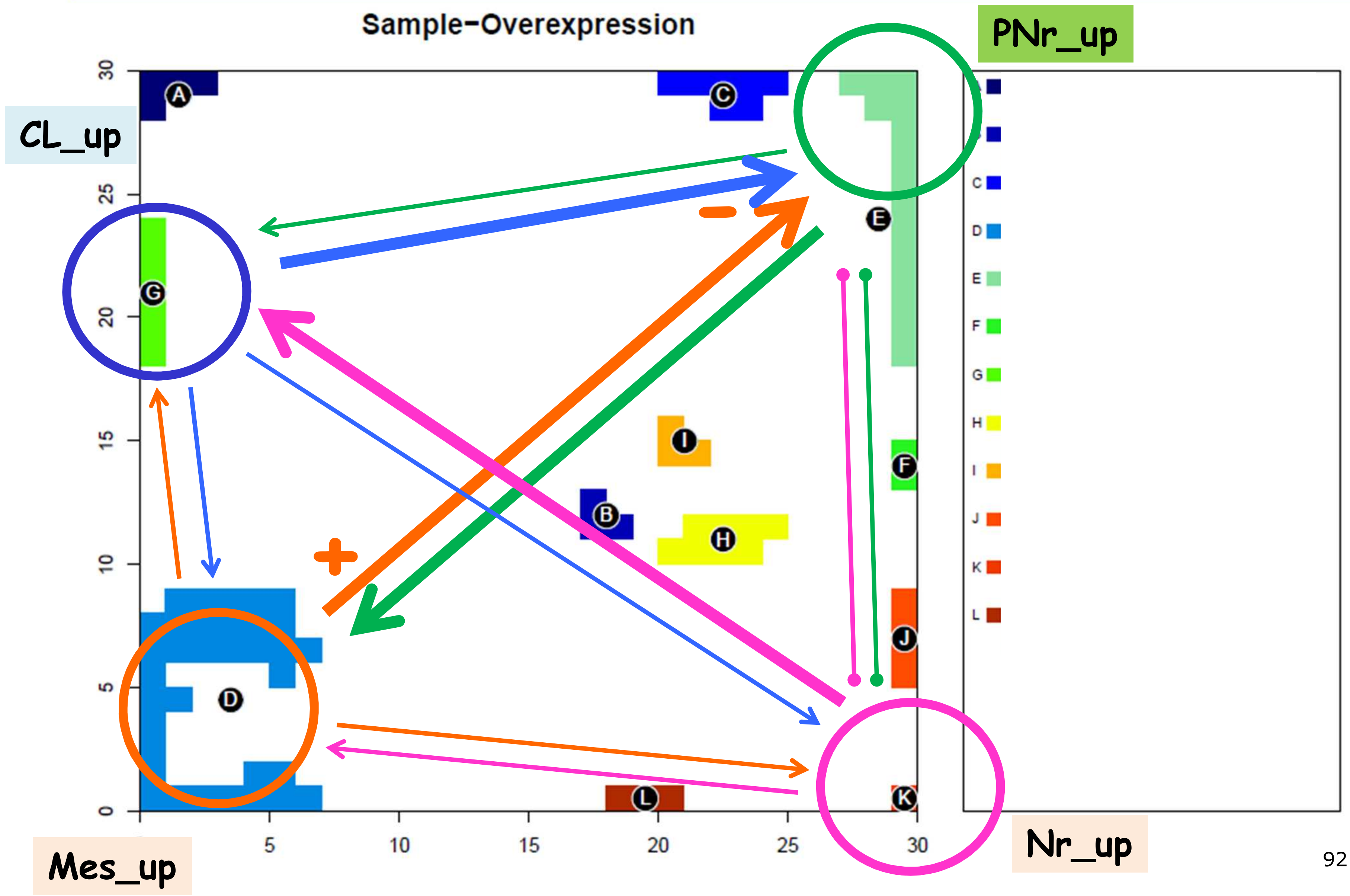
Nr_up



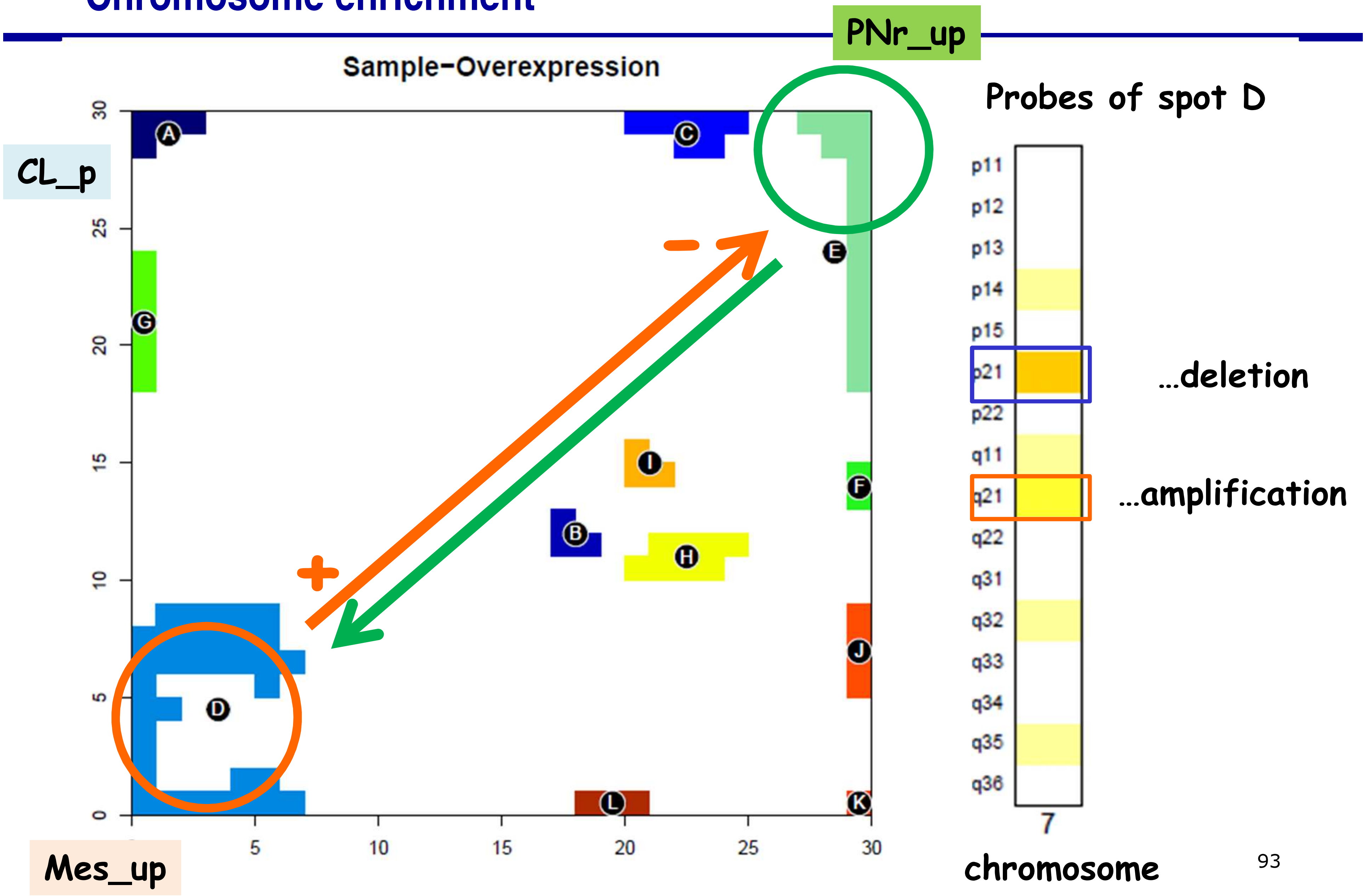


(-) (+) - +

Spots in concert



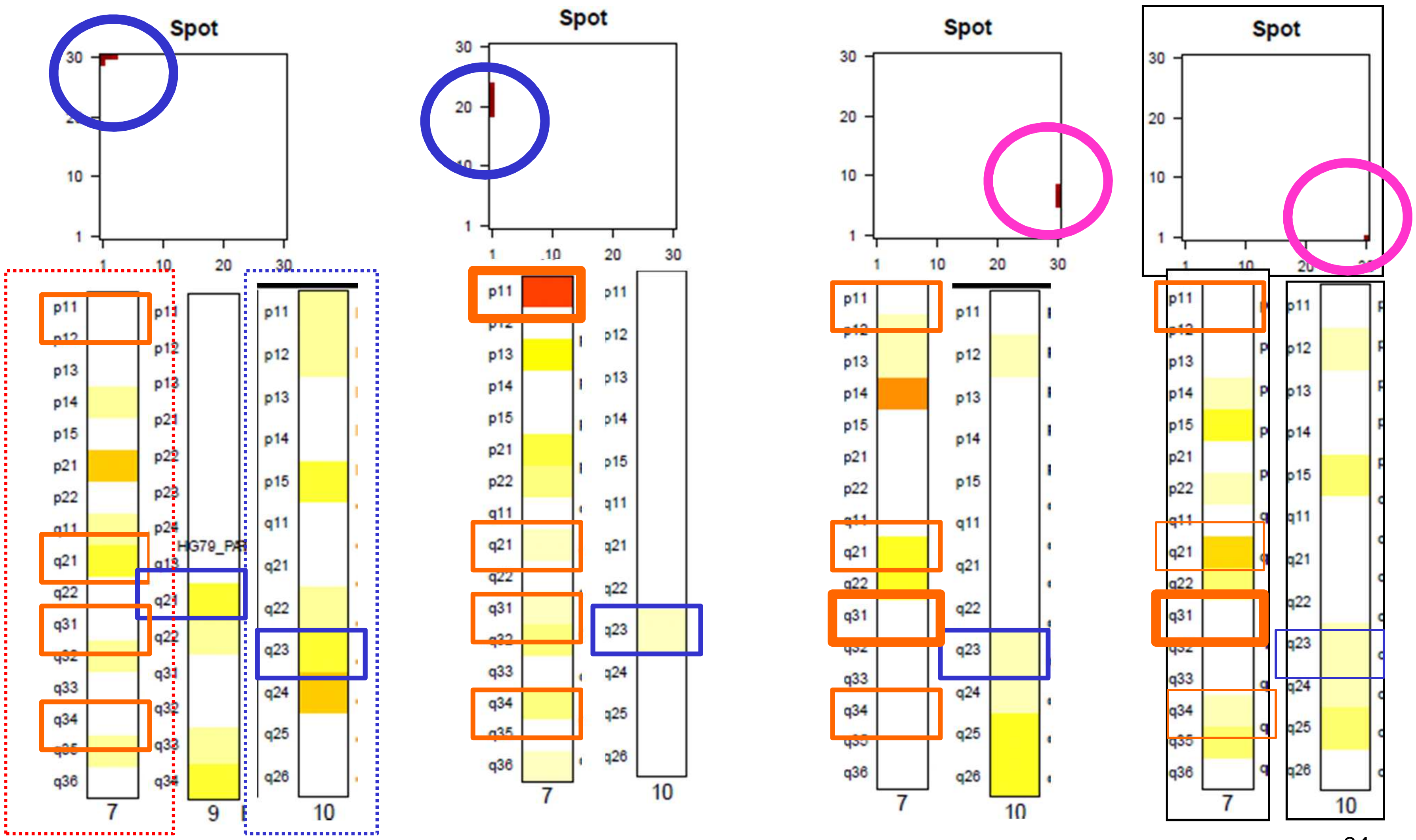
Chromosome enrichment



Chromosome enrichment

Classical_UP/Nr_DN

Classical_DN/Nr_UP

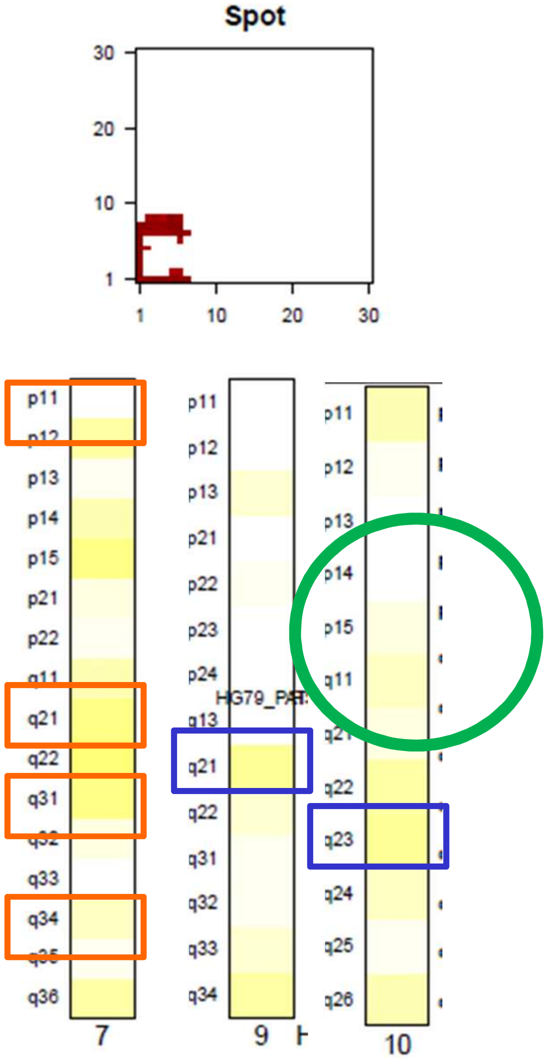


…amplification

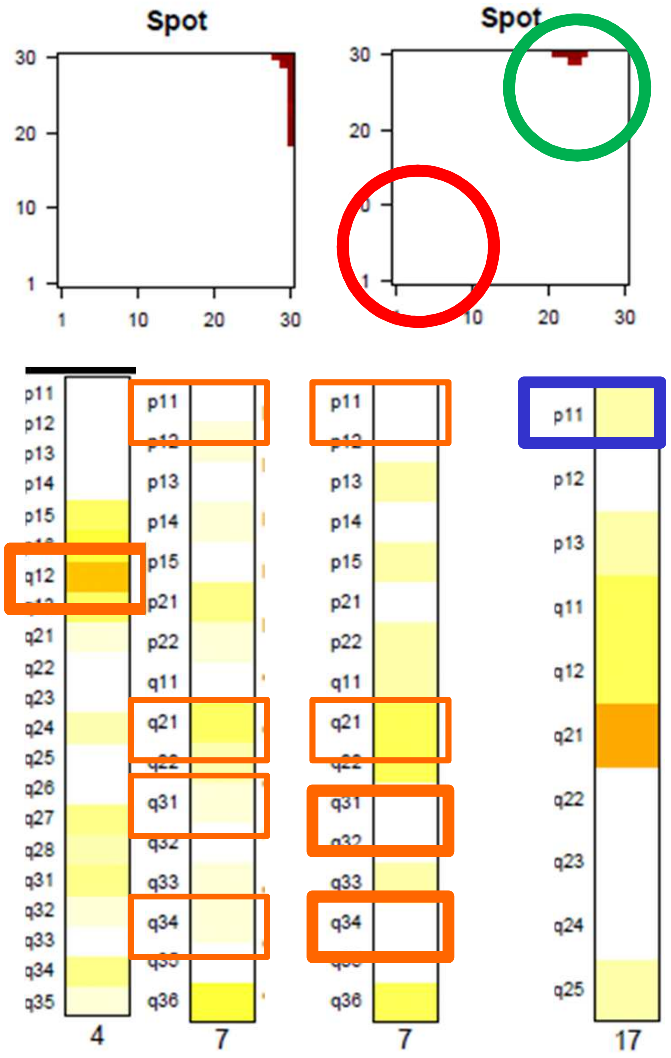
…deletion

Chromosome enrichment

Mes_UP/PNr_DN



Mes_DN/PNr_UP



Preliminary I: Chromosomal abberations

Classical
Amp:→ chr. 7
Del: → chr. 9,10

Neural
xxx

Mesenchymal
Del: → chr. 17

Proneural
Amp:→ chr. 4, 7

Cancer Cell
Article

Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*

Roel G.W. Verhaak,[1,2,17] Katherine A. Hoadley,[3,4,17] Elizabeth Purdom,[7] Victoria Wang,[8] Yuan Qi,[4,5] Matthew D. Wilkerson,[4,5] C. Ryan Miller,[4,6] Li Ding,[9] Todd Golub,[1,10] Jill P. Mesirov,[1] Gabriele Alexe,[1] Michael Lawrence,[1,2] Michael O'Kelly,[1,2] Pablo Tamayo,[1] Barbara A. Weir,[1,2] Stacey Gabriel,[1] Wendy Winckler,[1,2] Supriya Gupta,[1] Lakshmi Jakkula,[11] Heidi S. Feiler,[11] J. Graeme Hodgson,[12] C. David James,[12] Jann N. Sarkaria,[13] Cameron Brennan,[14] Ari Kahn,[15] Paul T. Spellman,[11] Richard K. Wilson,[9] Terence P. Speed,[7,16] Joe W. Gray,[11] Matthew Meyerson,[1,2] Gad Getz,[1] Charles M. Perou,[3,4,8] D. Neil Hayes,[4,5,*] and The Cancer Genome Atlas Research Network

Preliminary I: Chromosomal abberations

Classical
Ampl:→ chr. 7
Del: → chr. 9,10

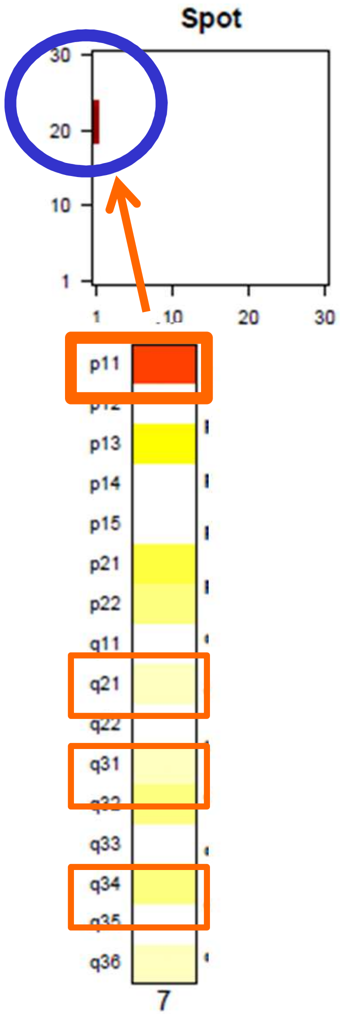
Neural
xxx

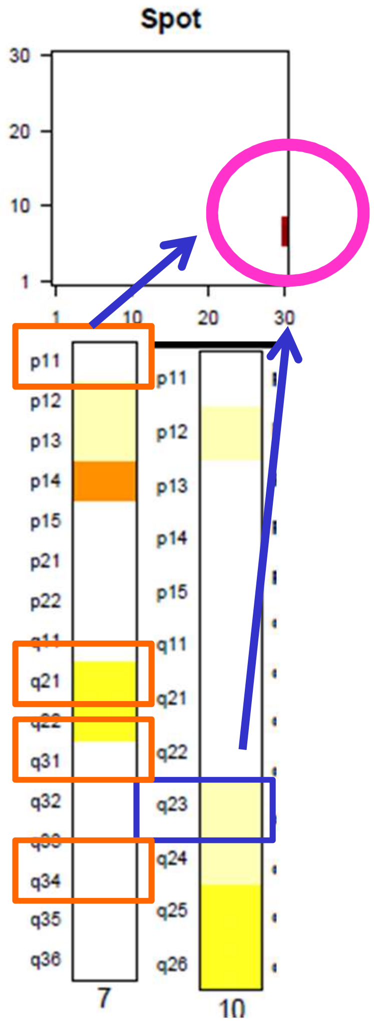
Cell PRESS

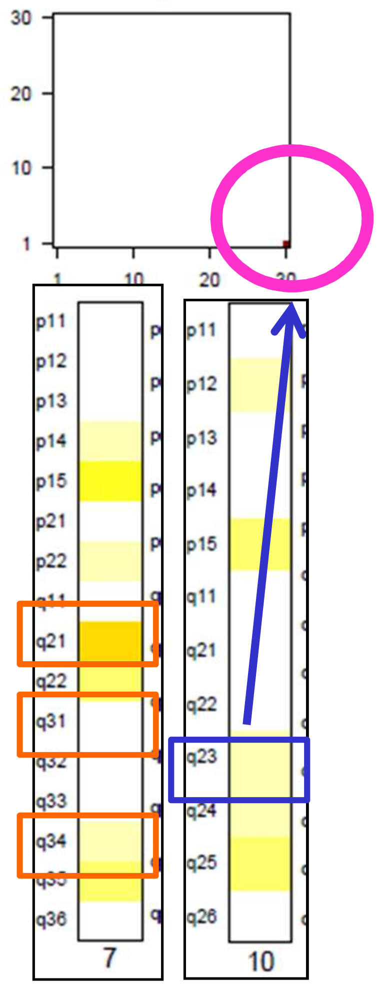
Cancer Cell
Article

Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*

Roel G.W. Verhaak,[1,2,17] Katherine A. Hoadley,[3,4,17] Elizabeth Purdom,[7] Victoria Wang,[8] Yuan Qi,[4,5] Matthew D. Wilkerson,[4,5] C. Ryan Miller,[4,6] Li Ding,[9] Todd Golub,[1,10] Jill P. Mesirov,[1] Gabriele Alexe,[1] Michael Lawrence,[1,2] Michael O'Kelly,[1,2] Pablo Tamayo,[1] Barbara A. Weir,[1,2] Stacey Gabriel,[1] Wendy Winckler,[1,2] Supriya Gupta,[1] Lakshmi Jakkula,[11] Heidi S. Feiler,[11] J. Graeme Hodgson,[12] C. David James,[12] Jann N. Sarkaria,[13] Cameron Brennan,[14] Ari Kahn,[15] Paul T. Spellman,[11] Richard K. Wilson,[9] Terence P. Speed,[7,16] Joe W. Gray,[11] Matthew Meyerson,[1,2] Gad Getz,[1] Charles M. Perou,[3,4,8] D. Neil Hayes,[4,5,*] and The Cancer Genome Atlas Research Network



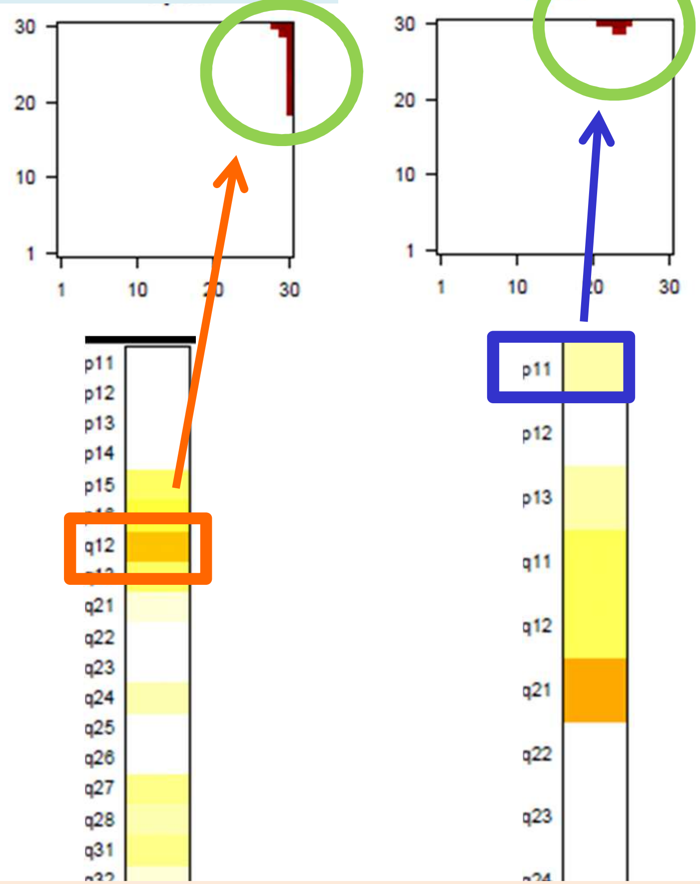




Preliminary I: Chromosomal abberations

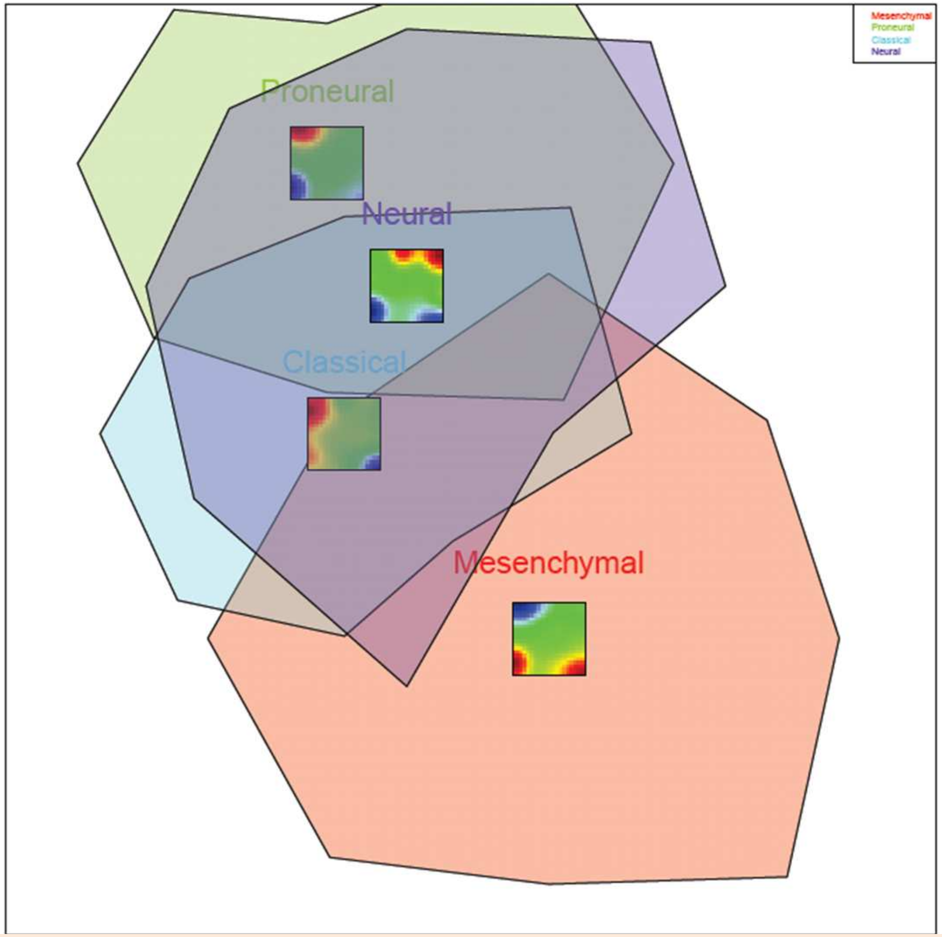
Mesenchymal
Del: → chr. 17

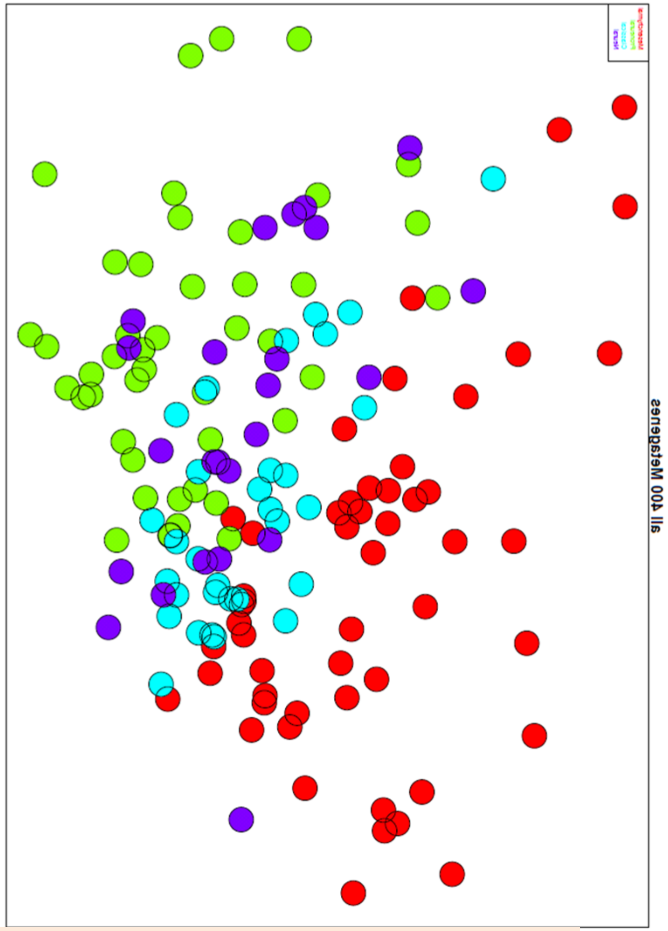
Proneural
Amp:→ chr. 4



Preliminary I: Spot-related chromosome-enrichment of gene activity correlates with CNV

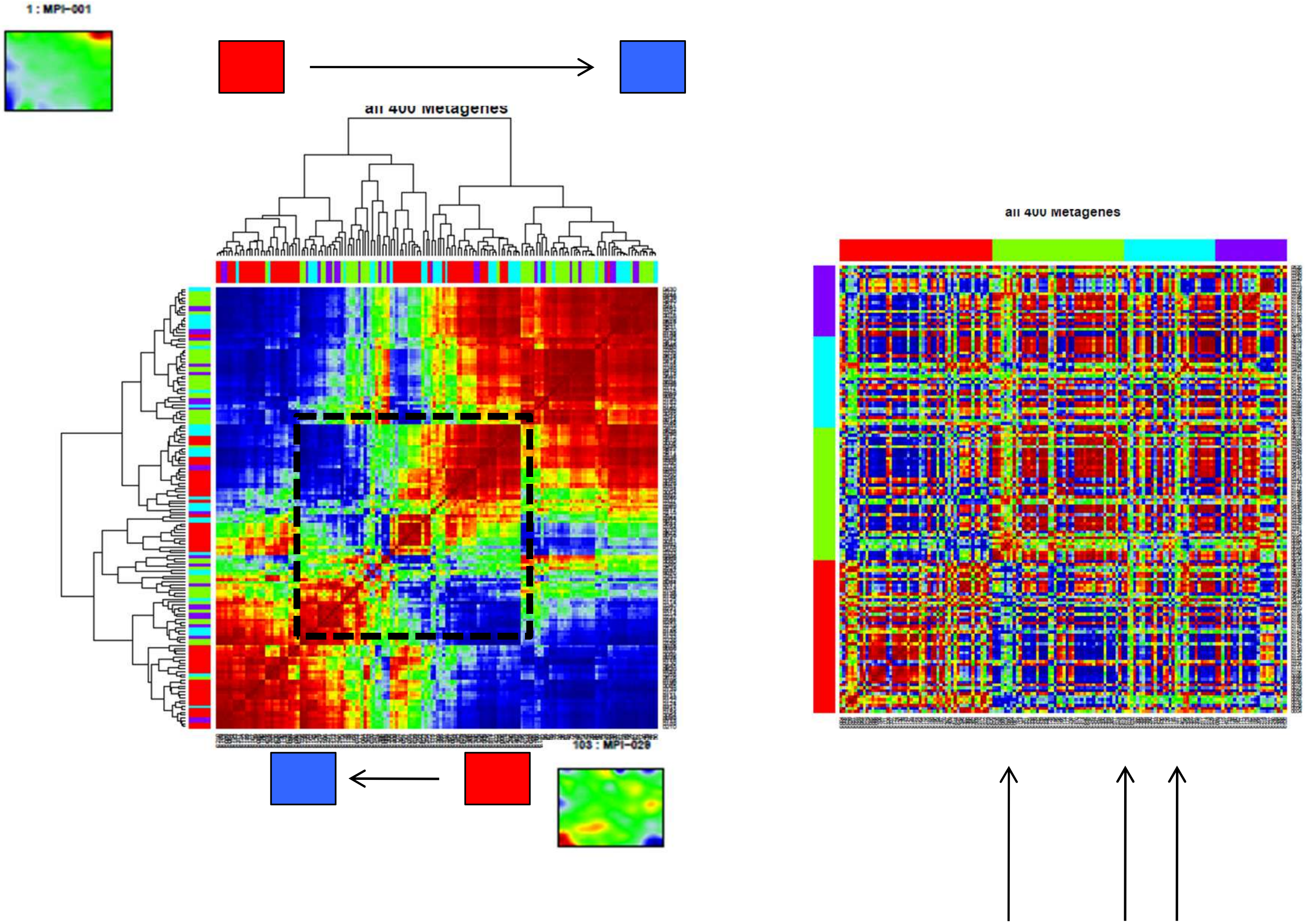
Preliminary I: miRNA





Preliminary II: miRNA expression pattern reproduces GMF-subtypes

Preliminary II: confusion about the miRNA data

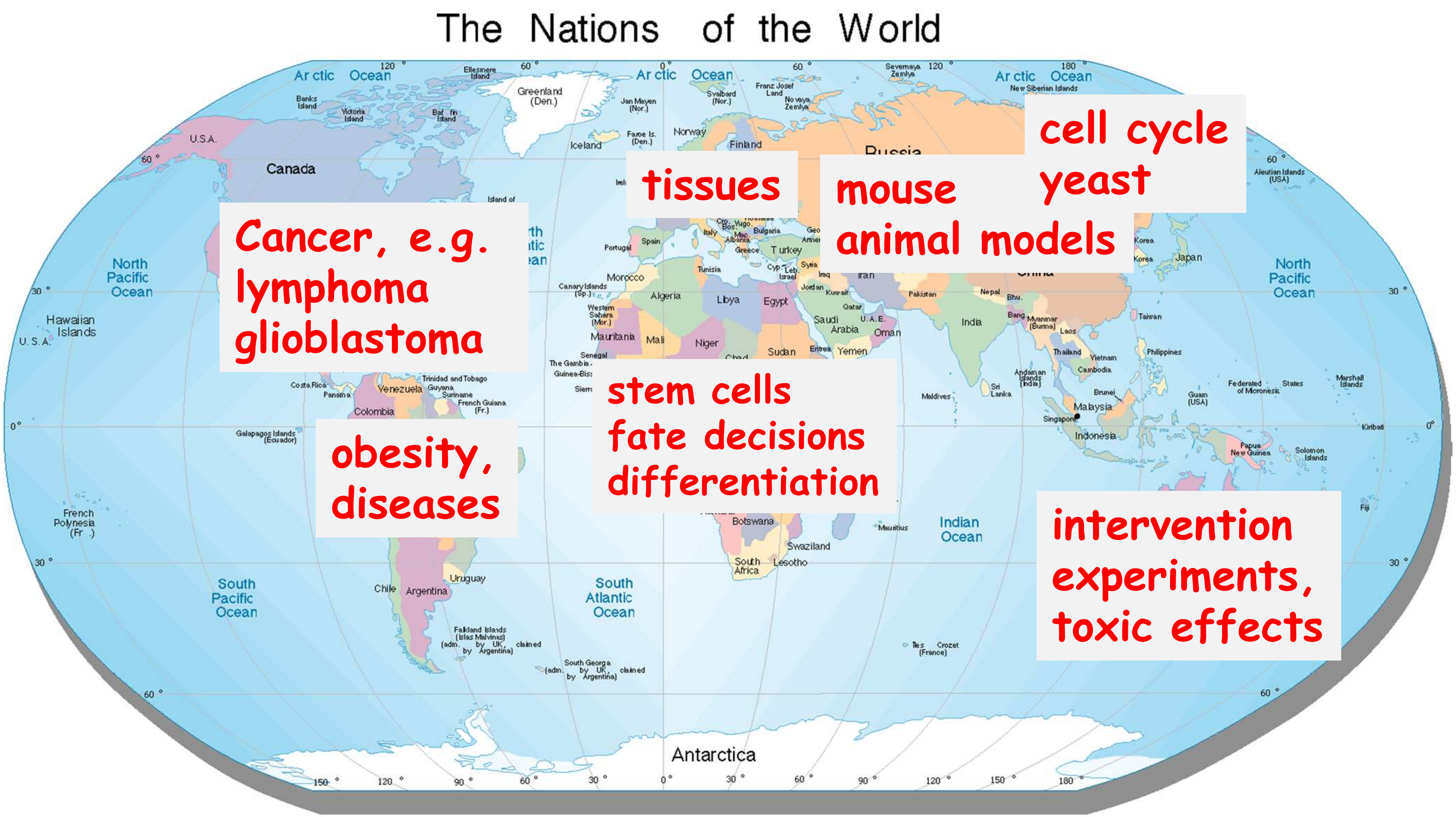


stripes

SOM and GMF

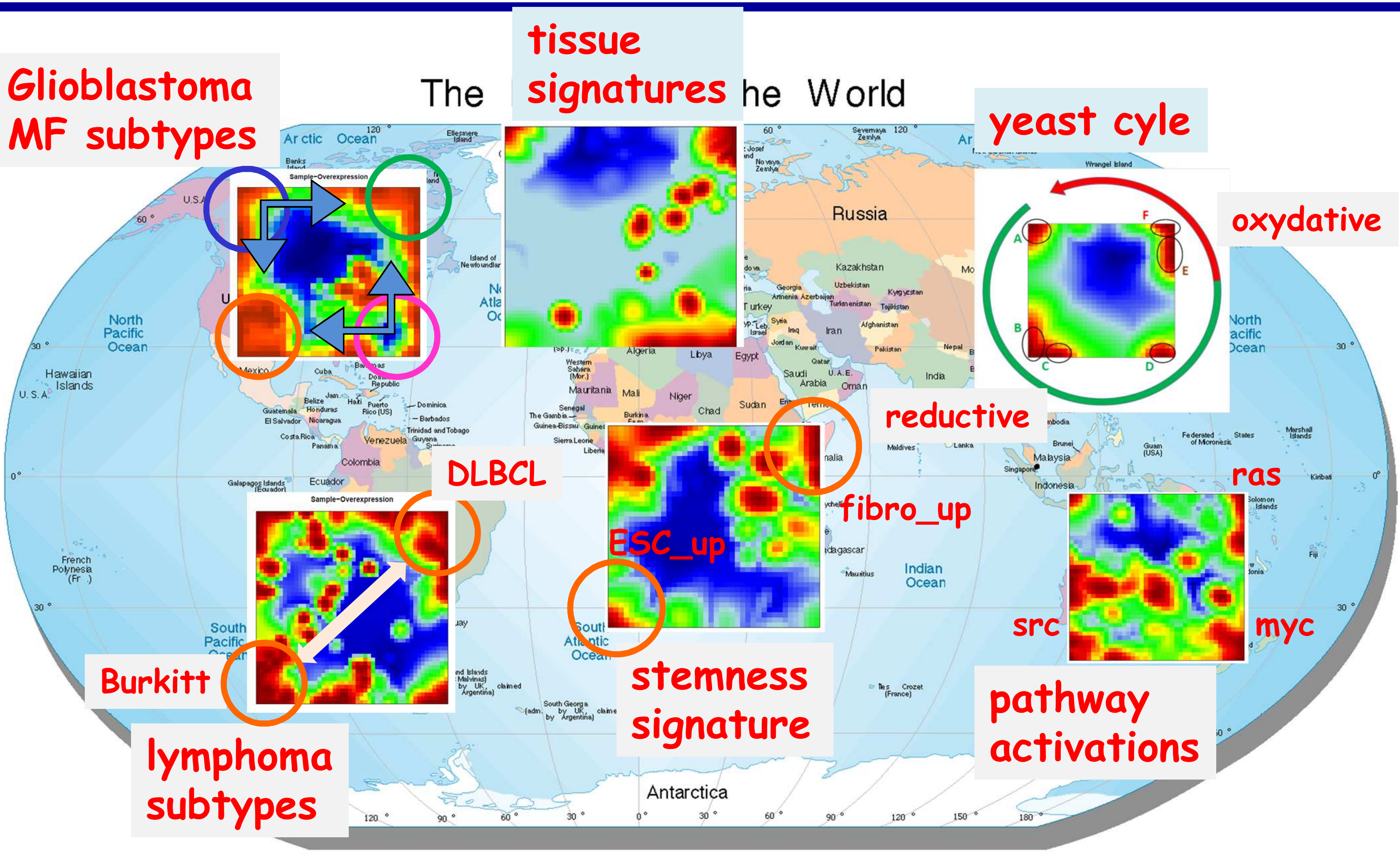
1. 4 subforms are well identified
2. Assignment of related mol. functions via spot-enrichment
3. Redundancy with other cancers, brain dysfunctions
4. Similarity relations: Mes-PNr and Cl-Nr mutually orthogonal expr. Changes; two ‚paths' between Mes and PNr via Cl. or Nr.
5. Fuzziness of expression: high for ‚intermediates'
6. integration of miRNA, CNV in progress

The transcriptome world



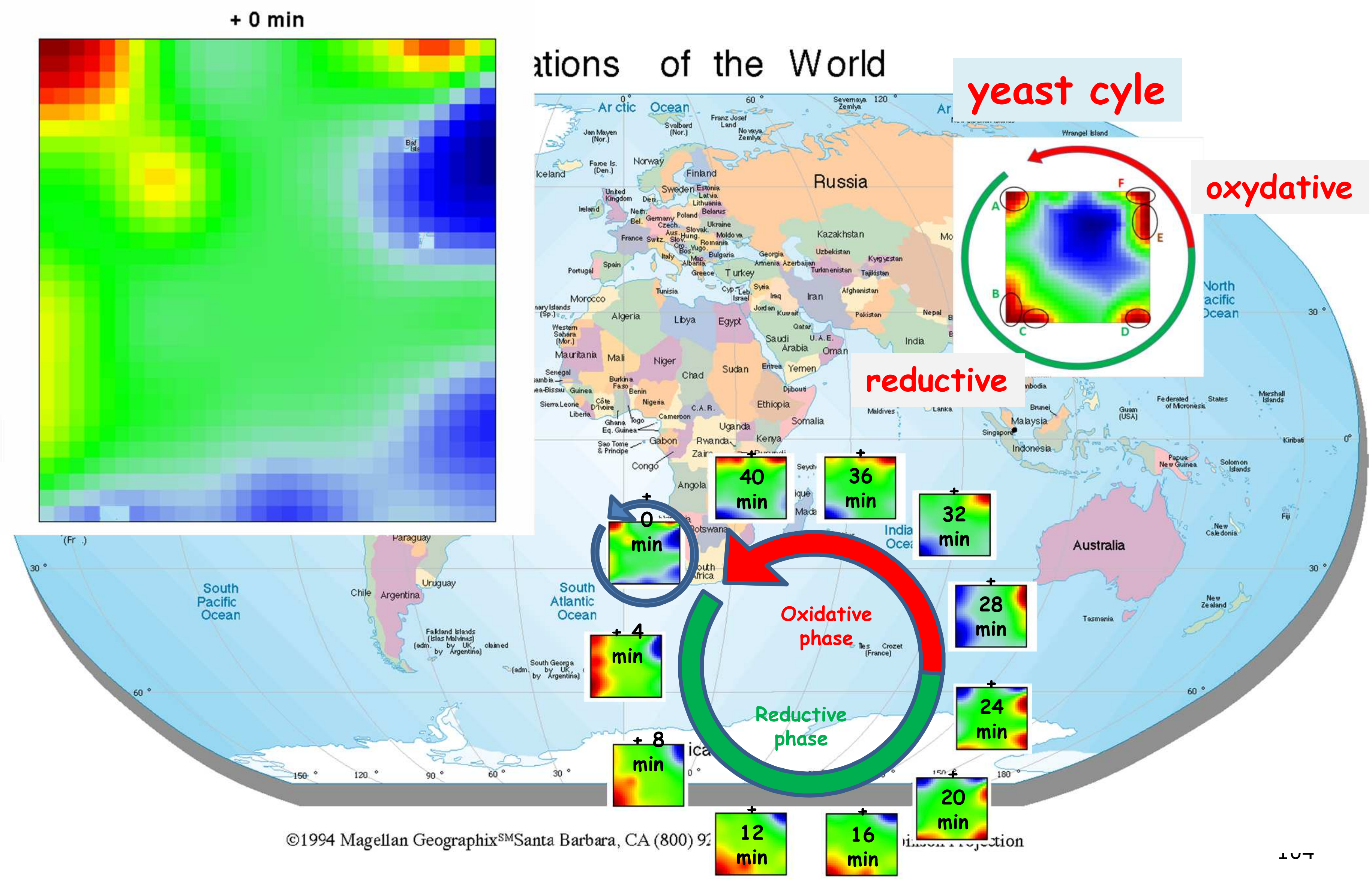
©1994 Magellan Geographix℠ Santa Barbara, CA (800) 929-4MAP Robinson Projection

The transcriptome world: expression signatures

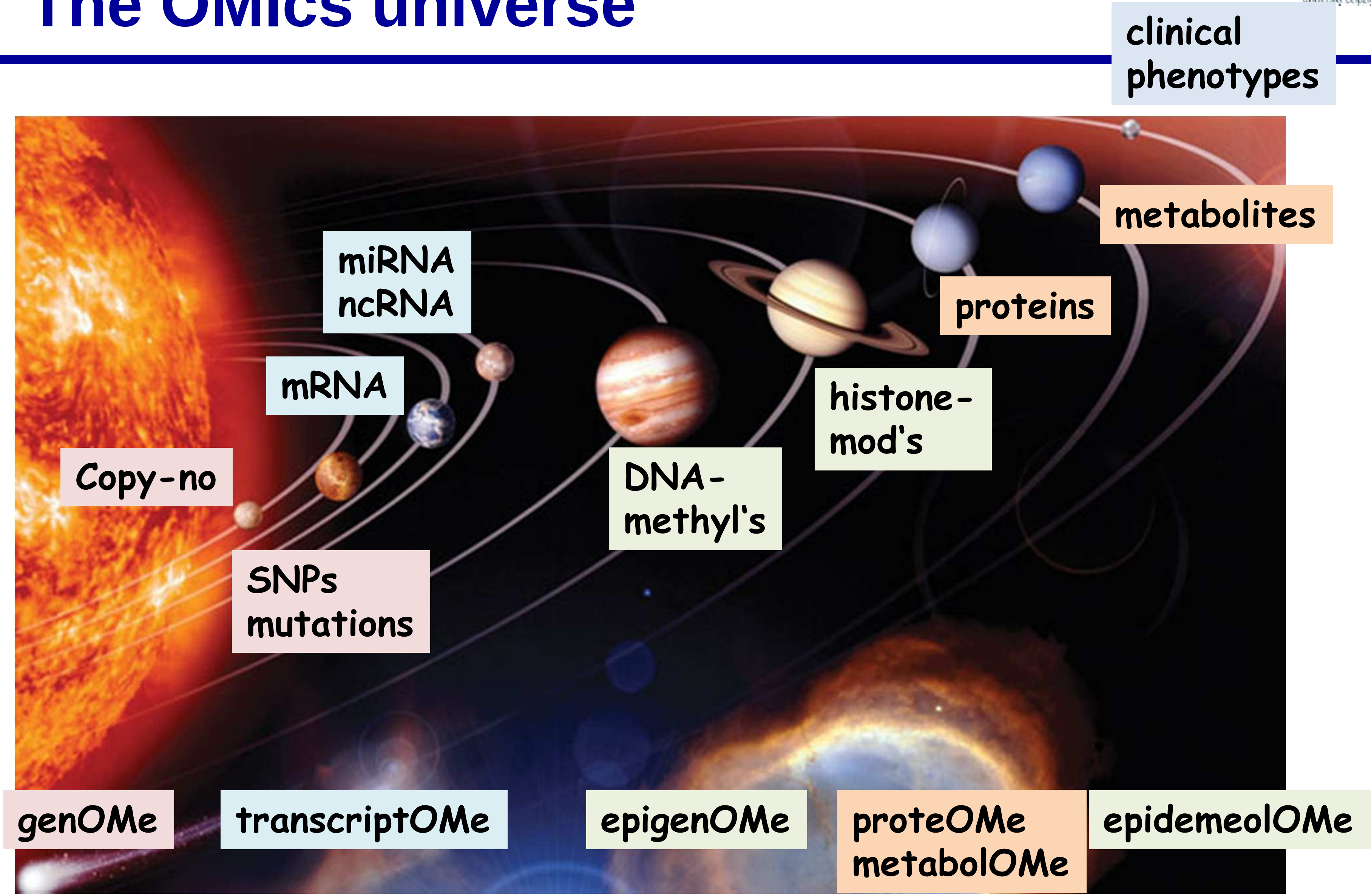


©1994 Magellan Geographix℠ Santa Barbara, CA (800) 929-4MAP Robinson Projection

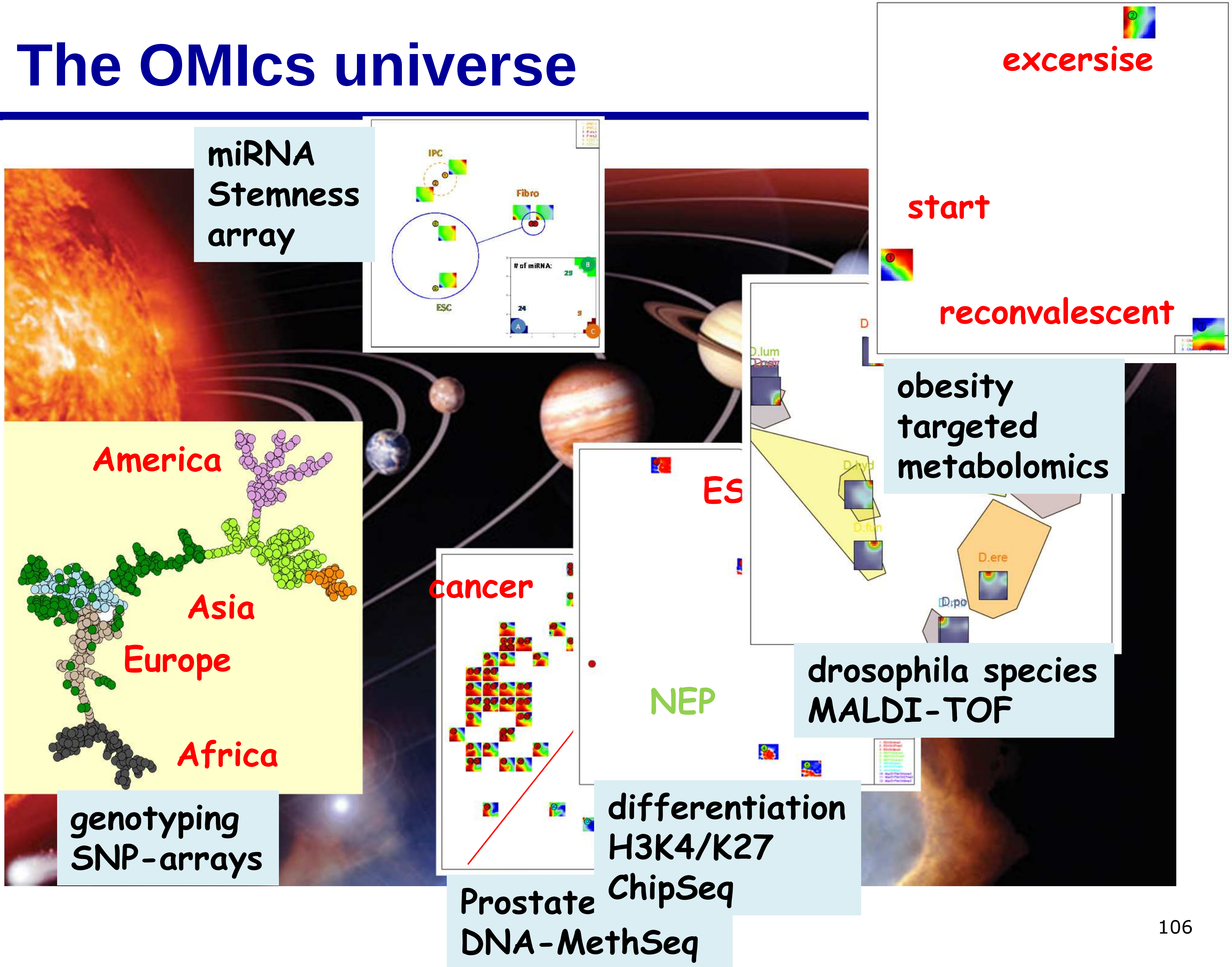
The transcriptome world: expression signatures



The OMIcs universe

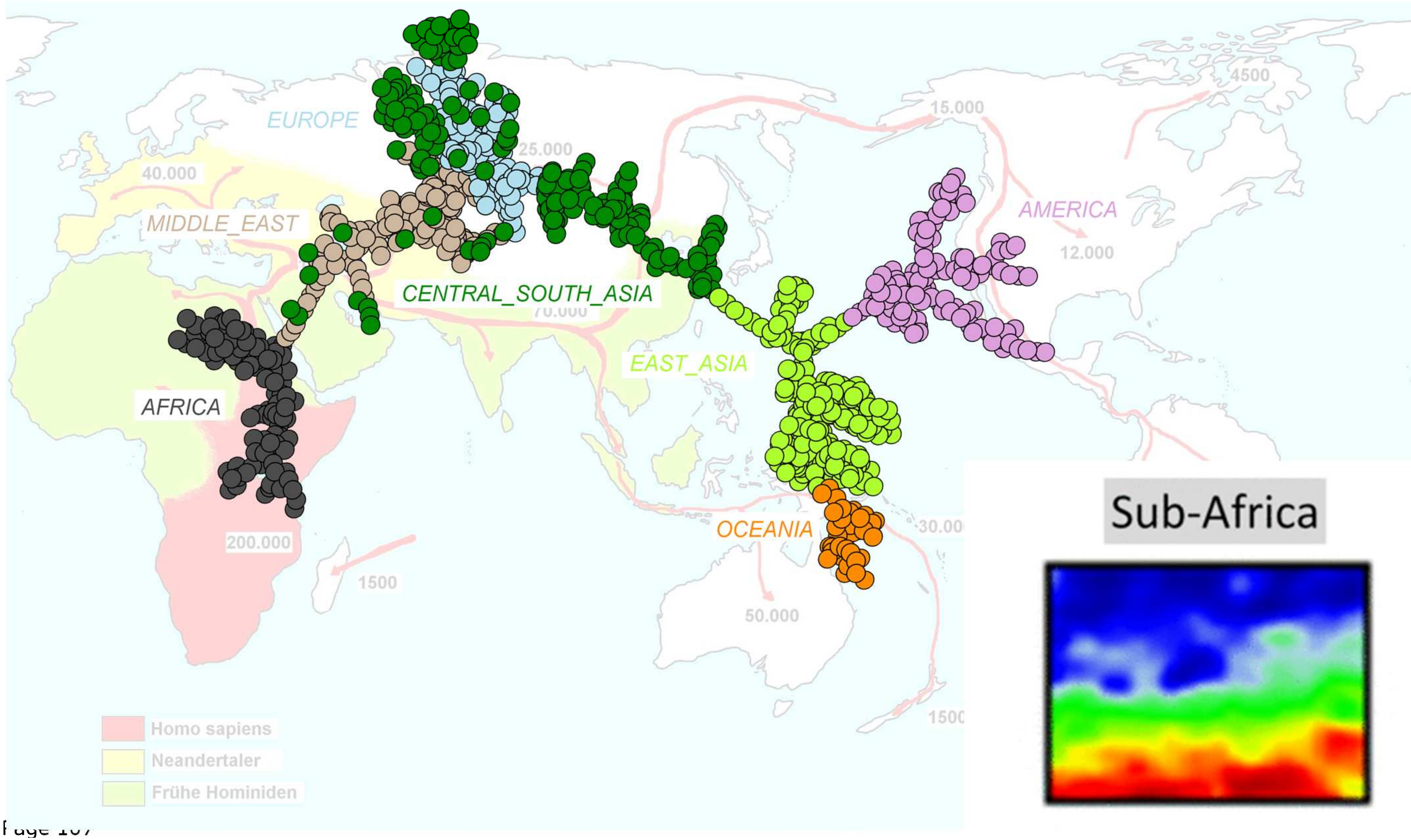


The OMIcs universe

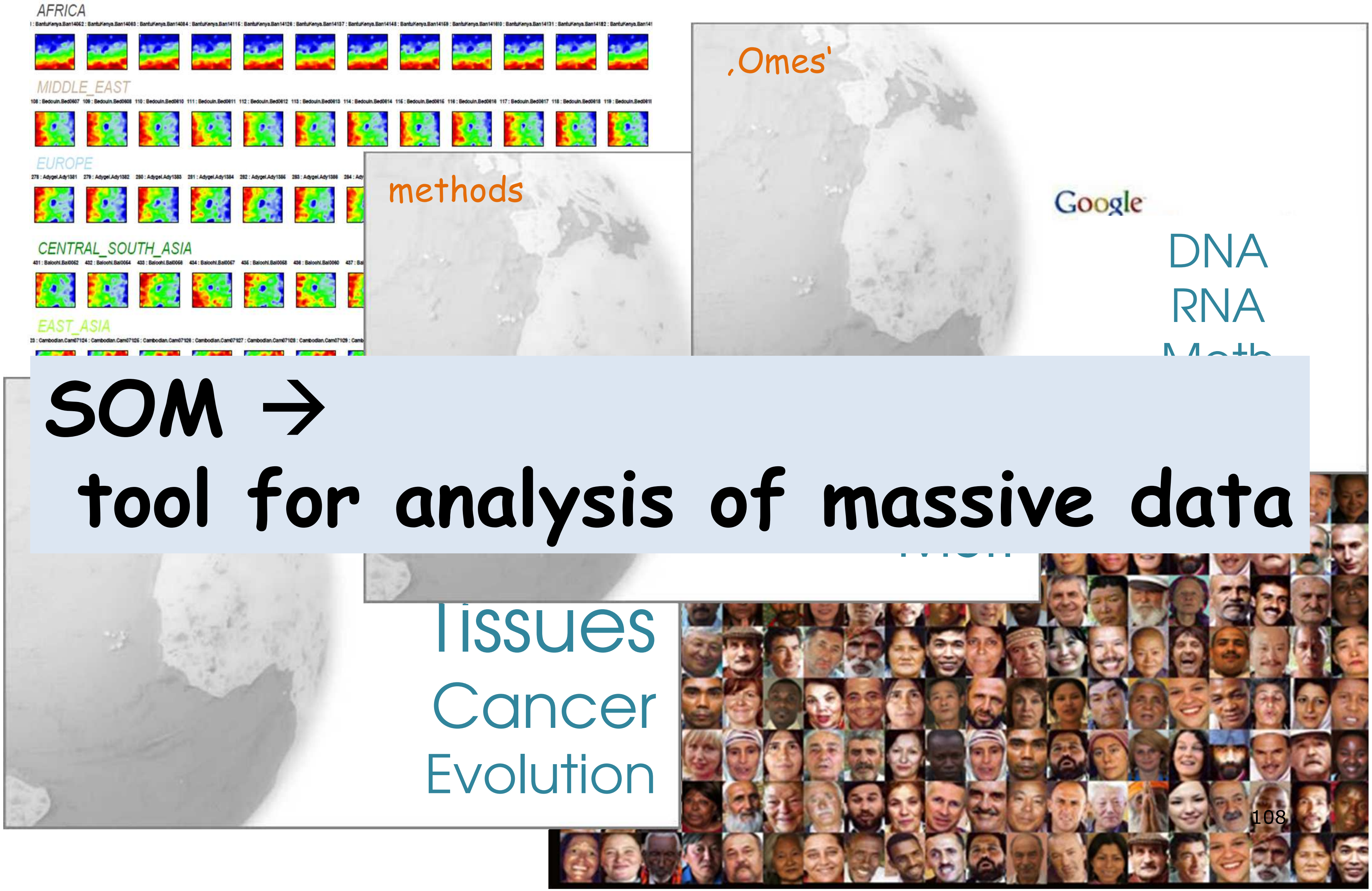


Human Genotype Atlas

Mapping the early human migrations



Genomic and/or molecular phenotypicic portraits



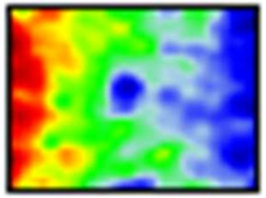
SOM →

tool for analysis of massive data

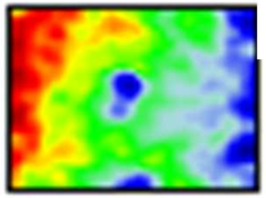
Thank you !

Thanks to

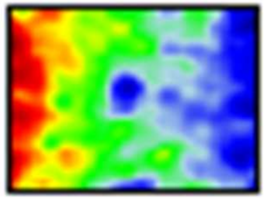
Henry Wirth



Mario Fasold



Lydia Hopp



Edith Willscher