



Supplementary Materials: Supervised projection of high-dimensional genome-wide expression on SOM transcriptome

Maria Nikoghosyan, Henry Loeffler-Wirth, Suren Davidavyan, Hans Binder and Arsen Arakelyan

Text S1. Parameters used for generation of simulated datasets using 'madsim' R package.

Simulated datasets were generated using 'madsim' R package with following command and parameters:

```
fparams = data.frame(m1=m1,m2=m2,shape2=4,lb=4,ub=14,pde=0.025,sym=0.5)
dparams = data.frame(lambda1=0.13, lambda2=2, muminde=1, sdde=0.5)
dat = madsim(mdata = NULL, n = 10000, ratio = 0, fparams, dparams, sdn = 0.7, rseed = 50).
```

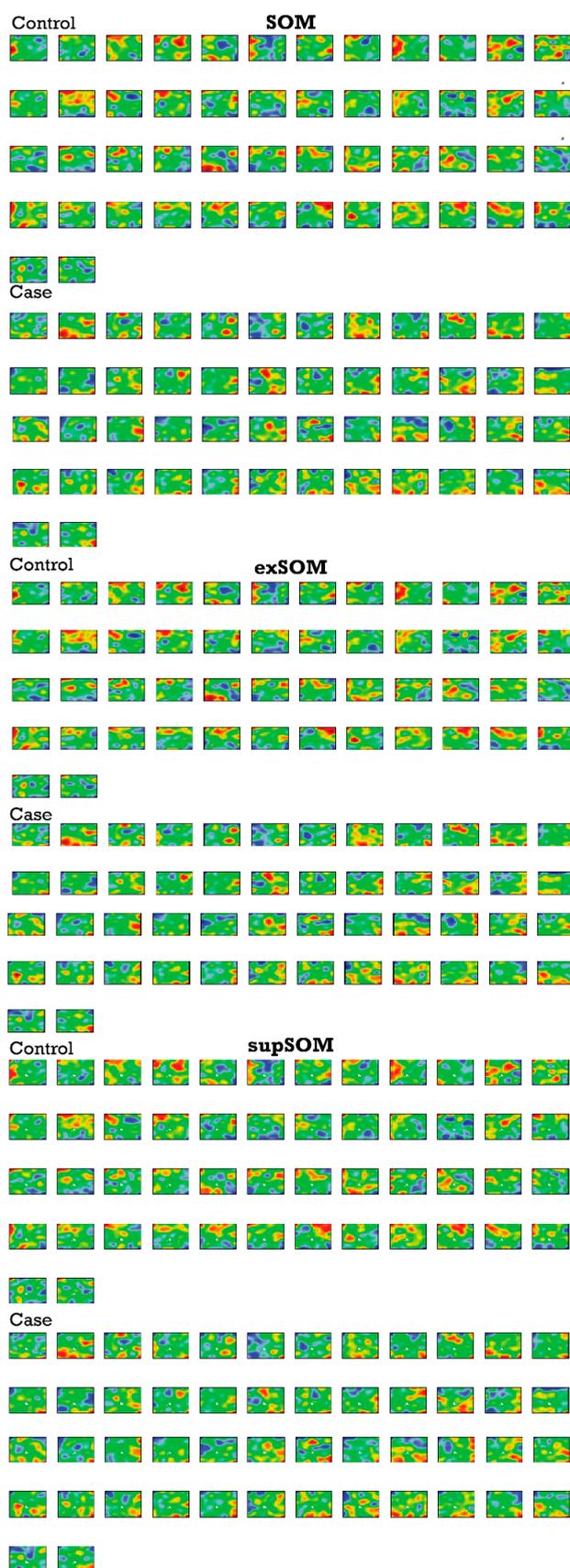


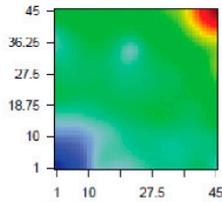
Figure S1. Complete portrayal of simulated dataset with SOM, exSOM, and supSOM.

SOM difference portraits

baseline vs placebo

Global Summary

%DE = 0.25
 # genes with $fdr < 0.2 = 551$ (227 + / 324 -)
 # genes with $fdr < 0.1 = 0$ (0 + / 0 -)
 # genes with $fdr < 0.05 = 0$ (0 + / 0 -)
 # genes with $fdr < 0.01 = 0$ (0 + / 0 -)



Differentially expressed genes

Rank ID	log(FC)	fdr p-value	Description Metagene
---------	---------	-------------	----------------------

Overexpressed

1	1560019_at	0.05 2e-07 0.2	16 x 18 DLGAP1 antisense RNA 2
2	1570482_at	0.02 3e-05 0.2	11 x 38
3	ERMN	0.02 4e-05 0.2	31 x 36 ermin
4	240232_at	0.07 4e-05 0.2	38 x 42
5	SLFN5	0.08 5e-05 0.2	37 x 44 schlafen family member 5
6	NETO2	0.08 6e-05 0.2	36 x 40 neuropilin and tolloid like 2
7	PITPNC1	0.05 8e-05 0.2	37 x 38 phosphatidylinositol transfer protein cytoplasmic 1
8	LAMC2	0.05 8e-05 0.2	8 x 37 laminin subunit gamma 2
9	NETO2	0.07 8e-05 0.2	36 x 41 neuropilin and tolloid like 2
10	240137_at	0.05 9e-05 0.2	37 x 40

Rank ID	log(FC)	fdr p-value	Description Metagene
---------	---------	-------------	----------------------

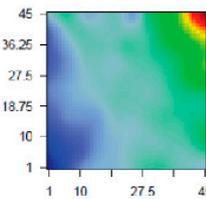
Underexpressed

1	ZDHHC1	-0.03 2e-05 0.2	34 x 7 zinc finger DHHC-type containing 1
2	SIDT2	-0.03 2e-05 0.2	10 x 9 SID1 transmembrane family member 2
3	1562799_at	-0.02 3e-05 0.2	26 x 26
4	CNGA1	-0.04 5e-05 0.2	9 x 10 cyclic nucleotide gated channel subunit alpha 1
5	MARCHF3	-0.03 5e-05 0.2	17 x 8 membrane associated ring-CH-type finger 3
6	GATAD1	-0.04 6e-05 0.2	21 x 32 GATA zinc finger domain containing 1
7	ADPRHL1	-0.03 6e-05 0.2	8 x 12 ADP-ribosylhydrolase like 1
8	ATG10	-0.02 6e-05 0.2	21 x 3 autophagy related 10
9	OR2F1	-0.02 8e-05 0.2	19 x 4 olfactory receptor family 2 subfamily F member 1
10	MMP28	-0.06 1e-04 0.2	7 x 4 matrix metalloproteinase 28

baseline vs responder

Global Summary

%DE = 0.5
 # genes with $fdr < 0.2 = 17842$ (7633 + / 10209 -)
 # genes with $fdr < 0.1 = 13973$ (6031 + / 7942 -)
 # genes with $fdr < 0.05 = 11095$ (4819 + / 6276 -)
 # genes with $fdr < 0.01 = 6968$ (3155 + / 3813 -)



Differentially expressed genes

Rank ID	log(FC)	fdr p-value	Description Metagene
---------	---------	-------------	----------------------

Overexpressed

1	S100A9	0.41 7e-14 1e-09	45 x 45 S100 calcium binding protein A9
2	214370_at	0.34 7e-14 1e-09	45 x 45
3	SOD2	0.21 1e-13 1e-08	45 x 45 superoxide dismutase 2
4	SOD2	0.24 6e-13 2e-08	45 x 45 superoxide dismutase 2
5	CXCL8	0.43 2e-12 2e-08	45 x 45 C-X-C motif chemokine ligand 8
6	NAMPT	0.17 3e-12 2e-08	45 x 43 nicotinamide phosphoribosyltransferase pseudogene 1
7	216565_x_at	0.1 3e-12 2e-08	45 x 36 interferon induced transmembrane protein pseudogene
8	TNFRSF6B	0.21 3e-12 2e-08	44 x 45 RTEL1-TNFRSF6B readthrough (NMD candidate)
9	S100A8	0.53 4e-12 2e-08	45 x 45 S100 calcium binding protein A8
10	DOC2B	0.19 4e-12 2e-08	45 x 29 double C2 domain beta

Rank ID	log(FC)	fdr p-value	Description Metagene
---------	---------	-------------	----------------------

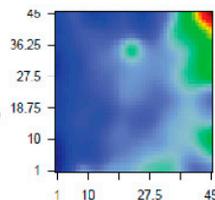
Underexpressed

1	MRPS25	-0.07 8e-12 2e-08	1 x 9 mitochondrial ribosomal protein S25
2	HNRNPDL	-0.06 5e-11 5e-08	1 x 31 heterogeneous nuclear ribonucleoprotein D like
3	MCCC1	-0.06 8e-11 1e-07	1 x 11 methylcrotonoyl-CoA carboxylase 1
4	EFCAB14	-0.05 1e-10 1e-07	9 x 9 EF-hand calcium binding domain 14
5	CCDC115	-0.06 1e-10 1e-07	1 x 12 coiled-coil domain containing 115
6	SSPN	-0.07 1e-10 1e-07	1 x 31 sarcospan
7	VWA8	-0.07 2e-10 1e-07	2 x 9 von Willebrand factor A domain containing 8
8	ANKH	-0.07 2e-10 1e-07	3 x 8 ANKH inorganic pyrophosphate transport regulator
9	DCAF4	-0.05 3e-10 2e-07	2 x 33 DDB1 and CUL4 associated factor 4
10	PLA2G12A	-0.08 4e-10 3e-07	1 x 7 phospholipase A2 group X1A

placebo vs responder

Global Summary

%DE = 0.34
 # genes with $fdr < 0.2 = 3991$ (1758 + / 2233 -)
 # genes with $fdr < 0.1 = 976$ (343 + / 633 -)
 # genes with $fdr < 0.05 = 77$ (18 + / 59 -)
 # genes with $fdr < 0.01 = 0$ (0 + / 0 -)



Differentially expressed genes

Rank ID	log(FC)	fdr p-value	Description Metagene
---------	---------	-------------	----------------------

Overexpressed

1	MYOZ3	0.02 9e-07 0.01	28 x 1 myozenin 3
2	NEUROD2	0.02 4e-06 0.02	40 x 17 neuronal differentiation 2
3	MAPK8IP2	0.03 8e-06 0.02	28 x 1 mitogen-activated protein kinase 8 interacting protein 2
4	1570575_at	0.02 8e-06 0.02	23 x 27
5	231595_at	0.03 1e-05 0.02	38 x 28 MRV1 antisense RNA 1
6	1565805_at	0.02 1e-05 0.02	39 x 17
7	MBP	0.05 1e-05 0.03	23 x 34 myelin basic protein
8	P4HA2	0.05 2e-05 0.03	42 x 10 prolyl 4-hydroxylase subunit alpha 2
9	JAK3	0.03 3e-05 0.03	30 x 1 Janus kinase 3
10	MBP	0.07 3e-05 0.03	22 x 35 myelin basic protein

Rank ID	log(FC)	fdr p-value	Description Metagene
---------	---------	-------------	----------------------

Underexpressed

1	EIF3J	-0.03 4e-07 0.01	2 x 21 eukaryotic translation initiation factor 3 subunit J
2	GCFC2	-0.05 2e-06 0.01	1 x 29 GC-rich sequence DNA-binding factor 2
3	ARIH2	-0.05 2e-06 0.01	1 x 19 ariadne RBR E3 ubiquitin protein ligase 2
4	242607_at	-0.05 2e-06 0.01	14 x 45
5	PNPT1	-0.05 2e-06 0.02	2 x 21 polyribonucleotide nucleotidyltransferase 1
6	1560019_at	-0.05 4e-06 0.02	16 x 18 DLGAP1 antisense RNA 2
7	POLR3E	-0.05 4e-06 0.02	1 x 31 RNA polymerase III subunit E
8	RRN3	-0.06 5e-06 0.02	1 x 30 RRN3 homolog, RNA polymerase I transcription factor
9	PUS7	-0.06 5e-06 0.02	1 x 21 pseudouridine synthase 7
10	GFM1	-0.05 7e-06 0.02	1 x 11 G elongation factor mitochondrial 1

Figure S2. Pairwise differential gene expression in primary SOM IBD dataset (GSE23597).

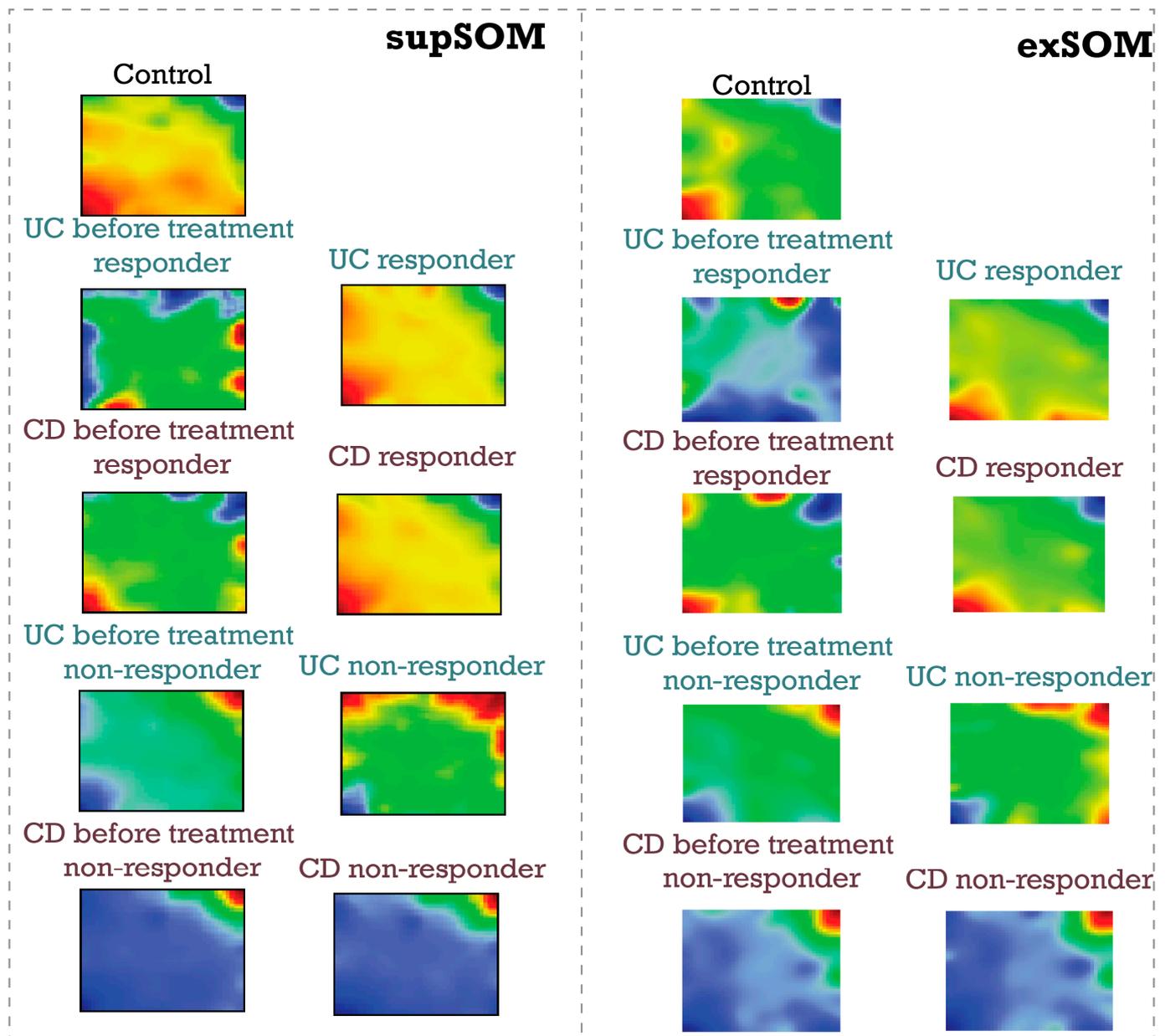


Figure S3. Comparison of supSOM and exSOM portraits in secondary IBD dataset (GSE16879).

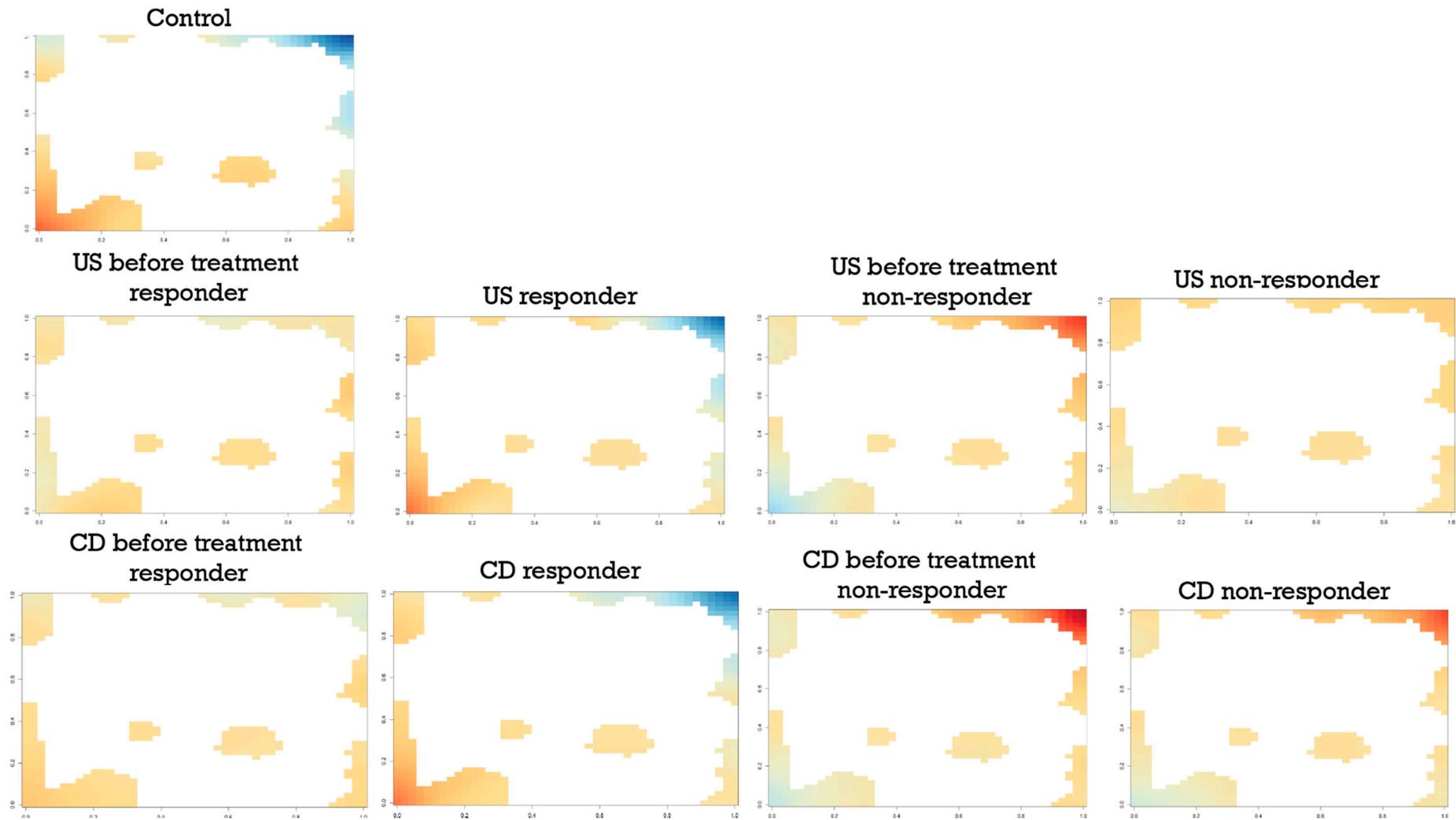
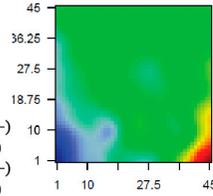


Figure S5. Differential expression landscape in IDB responders vs. non-responders in secondary dataset (GSE16879). Orange color indicates upregulation, blue color indicates down-regulation, white indicates region of invariant gene expression.

Normal vs grade_1

%DE = 0.55

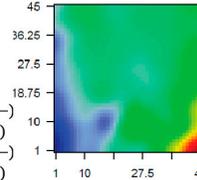
genes with fdr < 0.2 = 19871 (10684 + / 9187 -)
 # genes with fdr < 0.1 = 14551 (7476 + / 7075 -)
 # genes with fdr < 0.05 = 11160 (5546 + / 5614 -)
 # genes with fdr < 0.01 = 6333 (3094 + / 3239 -)



Normal vs grade_2

%DE = 0.52

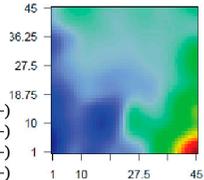
genes with fdr < 0.2 = 19207 (8776 + / 10431 -)
 # genes with fdr < 0.1 = 15572 (6874 + / 8698 -)
 # genes with fdr < 0.05 = 13587 (5848 + / 7739 -)
 # genes with fdr < 0.01 = 9743 (3980 + / 5763 -)



Normal vs grade_3

%DE = 0.57

genes with fdr < 0.2 = 21870 (9877 + / 11993 -)
 # genes with fdr < 0.1 = 18014 (7750 + / 10264 -)
 # genes with fdr < 0.05 = 15288 (6337 + / 8951 -)
 # genes with fdr < 0.01 = 10728 (4157 + / 6571 -)



Up-regulated in control

Rank ID	log(FC)	fdr	Description
	p-value	Metagene	
1	SH3KBP1	2.09 2e-12 2e-08	45 x 10 SH3 domain containing kinase binding protein 1
2	TMEM132C	2.72 6e-12 2e-08	43 x 1 transmembrane protein 132C
3	PLPP7	1.33 6e-12 8e-08	45 x 15 phospholipid phosphatase 7 (inactive)
4	GABARAPL1	1.04 2e-11 8e-08	45 x 28 GABA type A receptor associated protein like 1
5	HBEGF	1.27 4e-11 1e-07	45 x 12 heparin binding EGF like growth factor
6	MLXIPL	1.26 5e-11 1e-07	45 x 29 MLX interacting protein like
7	SH3KBP1	1.83 8e-11 1e-07	45 x 10 SH3 domain containing kinase binding protein 1
8	FAM89A	3.62 8e-11 1e-07	45 x 1 family with sequence similarity 89 member A
9	BIN1	1.27 9e-11 1e-07	45 x 28 bridging integrator 1
10	SGK2	1.42 9e-11 1e-07	45 x 14 serum/glucocorticoid regulated kinase 2

Up-regulated in control

Rank ID	log(FC)	fdr	Description
	p-value	Metagene	
1	MPC1	1.22 4e-18 5e-14	44 x 14 mitochondrial pyruvate carrier 1
2	STAT5A	1.44 2e-17 5e-13	45 x 12 signal transducer and activator of transcription 5A
3	CZIB	1.05 1e-16 9e-13	44 x 15 CXXC motif containing zinc binding protein
4	MXD1	0.98 2e-16 9e-13	45 x 24 MAX dimerization protein 1
5	ADH5	1.13 3e-16 1e-12	44 x 12 alcohol dehydrogenase 5 (class III), chi polypeptide
6	SMIM20	1.71 5e-16 2e-12	37 x 1 small integral membrane protein 20
7	MIR195	1.1 8e-16 3e-12	45 x 30 novel transcript
8	ARHGAP24	1.5 1e-15 4e-12	45 x 12 Rho GTPase activating protein 24
9	BIN1	1.12 2e-15 6e-12	45 x 28 bridging integrator 1
10	POMP	0.87 1e-14 3e-11	44 x 27 proteasome maturation protein

Up-regulated in control

Rank ID	log(FC)	fdr	Description
	p-value	Metagene	
1	MPC1	1.19 2e-18 1e-14	44 x 14 mitochondrial pyruvate carrier 1
2	CZIB	1.12 3e-18 6e-14	44 x 15 CXXC motif containing zinc binding protein
3	STAT5A	1.49 6e-18 2e-13	45 x 12 signal transducer and activator of transcription 5A
4	MTIF3	1.33 3e-17 2e-13	43 x 13 mitochondrial translational initiation factor 3
5	MIR195	1.19 7e-17 2e-13	45 x 30 novel transcript
6	CEP126	1.67 4e-16 8e-13	38 x 10 centrosomal protein 126
7	SMIM20	1.59 4e-16 1e-12	37 x 1 small integral membrane protein 20
8	PHLDB1	0.99 7e-16 1e-12	45 x 32 pleckstrin homology like domain family B member 1
9	ARHGAP24	1.61 8e-16 1e-12	45 x 12 Rho GTPase activating protein 24
10	ADH5	1.06 1e-15 1e-12	44 x 12 alcohol dehydrogenase 5 (class III), chi polypeptide

Up-regulated in grade I

Rank ID	log(FC)	fdr	Description
	p-value	Metagene	
1	226808_at	-1.89 7e-13 2e-08	3 x 7
2	244625_at	-1.89 1e-12 2e-08	2 x 7
3	238299_at	-1.7 3e-12 2e-08	1 x 13
4	XPO1	-2.32 4e-12 2e-08	1 x 7 exportin 1
5	238593_at	-3.27 5e-12 2e-08	1 x 7
6	232653_at	-3.58 1e-11 8e-08	1 x 2
7	LRRC1	-2.48 2e-11 8e-08	1 x 4 leucine rich repeat containing 1
8	TRPS1	-3.04 2e-11 8e-08	1 x 3 transcriptional repressor GATA binding 1
9	243963_at	-1.68 2e-11 8e-08	4 x 6
10	PAICS	-1.37 3e-11 8e-08	15 x 8 phosphoribosylaminoimidazole carboxylase

Up-regulated in grade II

Rank ID	log(FC)	fdr	Description
	p-value	Metagene	
1	1559491_at	-2.58 2e-20 6e-15	1 x 6
2	COL10A1	-3.26 4e-19 6e-15	1 x 3 collagen type X alpha 1 chain
3	COL10A1	-2.97 5e-19 5e-14	1 x 3 collagen type X alpha 1 chain
4	235538_at	-2.19 4e-18 5e-14	1 x 4
5	221973_at	-2.56 6e-18 5e-14	1 x 6 novel transcript
6	243329_at	-2.19 7e-18 3e-13	1 x 7
7	INHBA	-3.15 5e-17 5e-13	1 x 4 inhibin subunit beta A
8	COL11A1	-4.84 7e-17 5e-13	1 x 4 collagen type XI alpha 1 chain
9	RNF213	-2.49 8e-17 6e-13	1 x 8 ring finger protein 213
10	DENN1B	-2.6 1e-16 6e-13	1 x 6 DENN domain containing 1B

Up-regulated in grade III

Rank ID	log(FC)	fdr	Description
	p-value	Metagene	
1	242890_at	-1.85 1e-21 1e-15	15 x 9
2	221973_at	-2.19 5e-20 1e-15	1 x 6 novel transcript
3	MPV17	-0.94 1e-19 1e-15	15 x 6 mitochondrial inner membrane protein MPV17
4	TET3	-2.13 2e-19 1e-15	15 x 10 tet methylcytosine dioxygenase 3
5	1559491_at	-2.04 2e-19 2e-15	1 x 6
6	KNOP1	-1.52 3e-19 2e-15	15 x 7 lysine rich nucleolar protein 1
7	BAIAP2L1	-2.15 4e-19 1e-14	16 x 10 BAR/IMD domain containing adaptor protein 2 like 1
8	C6orf132	-1.74 2e-18 1e-14	1 x 5 chromosome 6 open reading frame 132
9	CGN	-2.4 2e-18 1e-14	1 x 3 cingulin
10	PSRC1	-1.86 3e-18 1e-14	16 x 10 proline and serine rich coiled-coil 1

Figure S6. Grade-dependent change of differential expression genes in breast cancer.

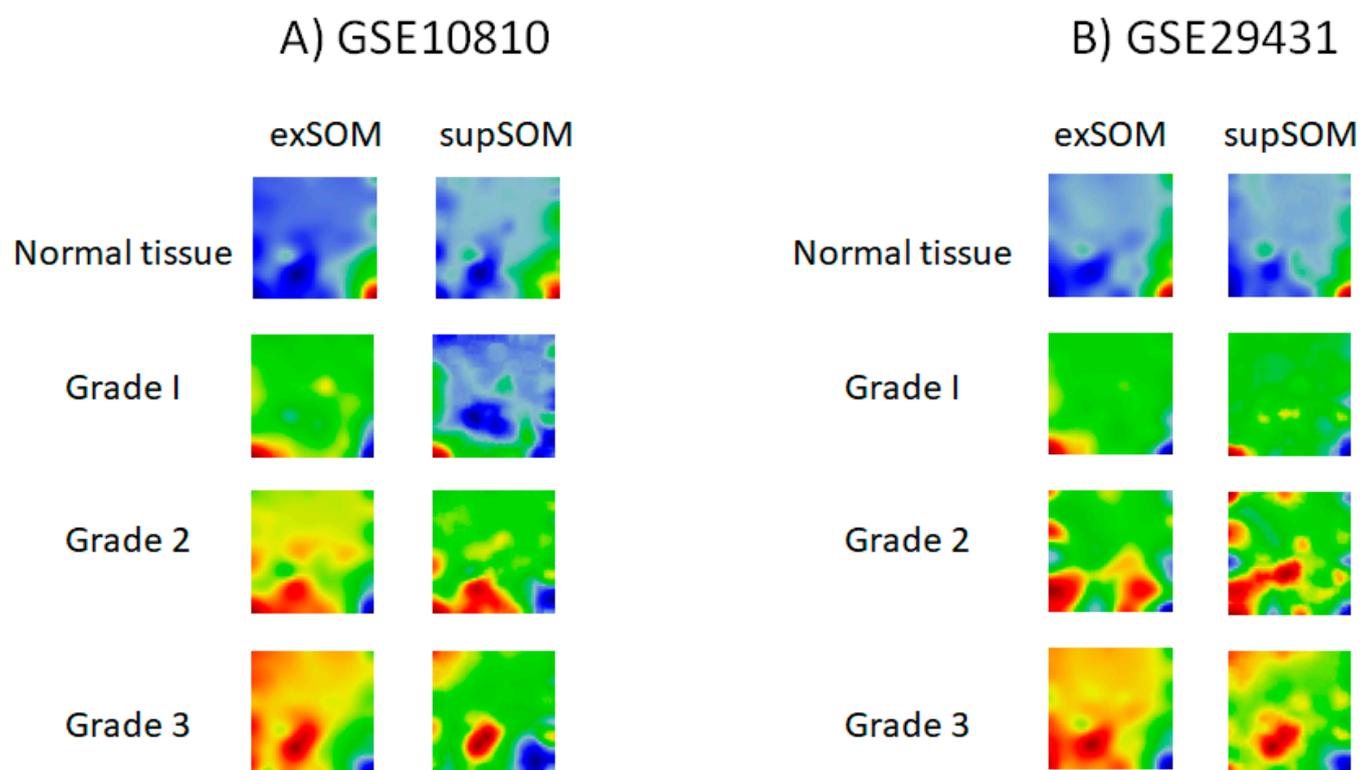


Figure S7. Comparison of exSOM and supSOM portraits in secondary breast cancer datasets. A) GSE10810, B) GSE29431.