

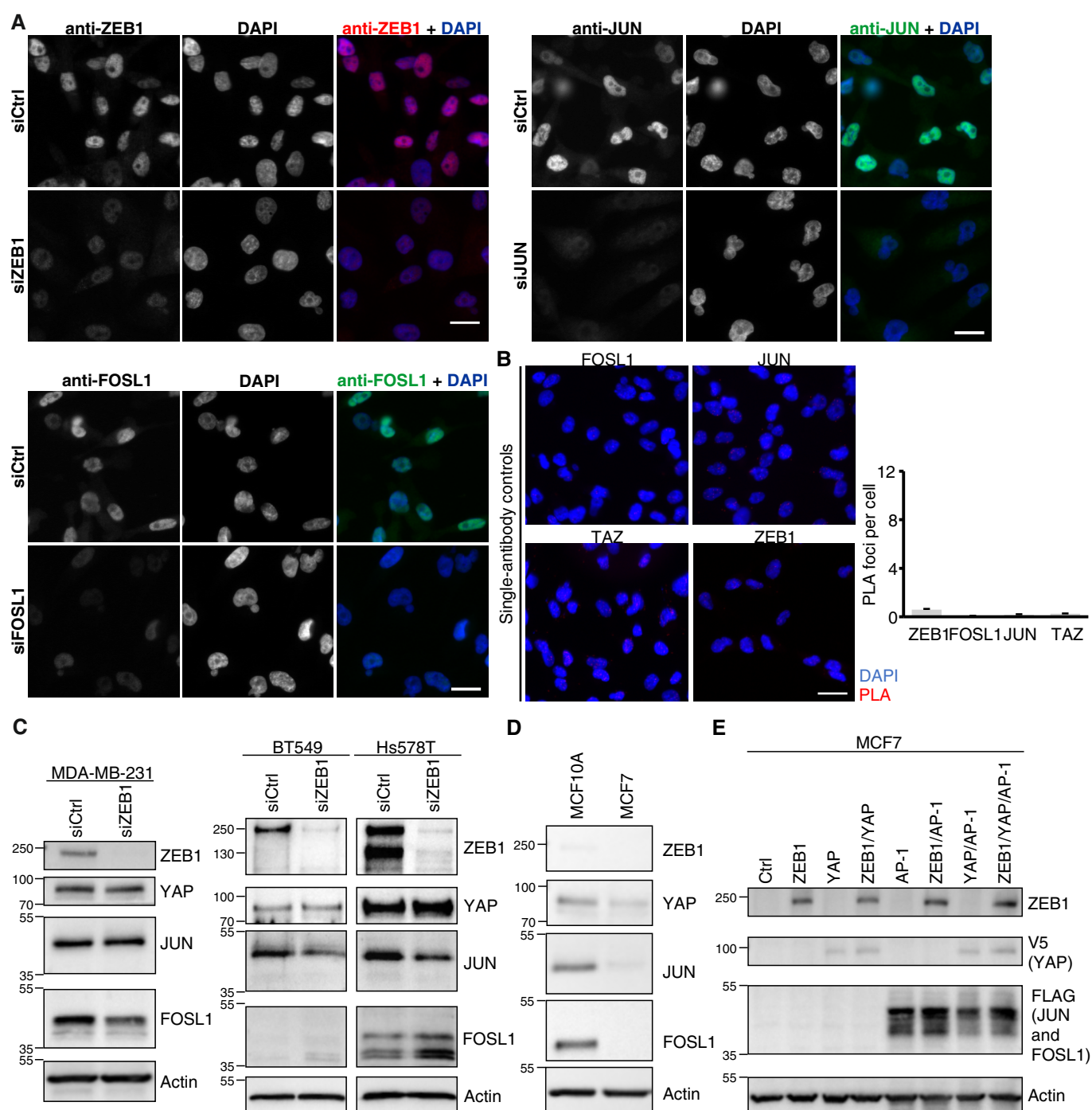
Appendix

Genome-wide cooperation of the EMT-activator ZEB1 with YAP and AP-1 factors in breast cancer

Nora Feldker^{1,*}, Fulvia Ferrazzi^{2,3,4,*}, Harald Schuhwerk¹, Sebastian A. Widholz^{1,5}, Kerstin Guenther⁶, Isabell Frisch¹, Kathrin Jakob¹, Julia Kleemann¹, Dania Riegel⁷, Ulrike Bönisch⁸, Sören Lukassen^{2,9}, Rebecca L. Eccles¹, Christian Schmidl⁷, Marc P. Stemmler¹, Thomas Brabletz^{1,10,#,*} and Simone Brabletz^{1,#,*}

Table of content

Appendix figure S1. Controls.	2
Appendix table S1. HOMER known motif analysis on all ZEB1 peaks.	4
Appendix table S2. HOMER <i>de novo</i> motif analysis on all ZEB1 peaks.	7
Appendix table S3. HOMER known motif analysis on ZEB1only peak set.	8
Appendix table S4. HOMER known motif analysis on ZEB1/YAP/JUN peak set.	9
Appendix table S5. <i>de novo</i> HOMER motif analysis on ZEB1 peak subsets.	10
Appendix table S6. HOMER known motif analysis on ZEB1 repressed regions (identified by ATAC-seq).	11
Appendix table S7. HOMER known motif analysis on ZEB1 activated regions (identified by ATAC-seq).	12
Appendix table S8. <i>de novo</i> HOMER motif analysis on ZEB1 peaks associated with active or repressed genomic regions.	13
Appendix table S9. ZEB1/YAP/JUN activated target genes positively correlate with ZEB1 expression in breast cancer cell lines.	14
Appendix table S10. ZEB1/YAP/JUN activated target genes positively correlate with ZEB1 expression in human breast cancer tumours.	15
Appendix table S11. Oligonucleotides used for cloning.	16
Appendix table S12. Oligonucleotides used as EMSA probes.	16
Appendix table S13. Oligonucleotides used for qPCR.	17



Appendix figure S1. Controls. (continued on next page)

(A) Immunofluorescence staining of ZEB1 and the AP-1 factors JUN and FOSL1 in control MDA-MB-231 cells and after transient siRNA mediated knockdown. Scale bar = 20 μ m.

(B) *In situ* proximity ligation assay (PLA) of MDA-MB-231 cells which were incubated with FOSL1, JUN, TAZ or ZEB1 antibodies alone showed no unspecific signal generation by individual antibodies only. In the left panel representative microscopic images are shown. Scale bar = 20 μ m. Quantification of the PLA is shown on the right. n=2; shown is one representative experiment with means $\pm$ s.e.m. of at least 900 cells per condition.

Appendix figure S1. Controls. (continued)

(C) Transient siRNA mediated knockdown of ZEB1 in MDA-MB-231, BT549 and Hs578T cells. Western blot analysis shows strong reduction of ZEB1 protein levels but no effect on JUN, YAP or FOSL1.

(D) Expression of ZEB1, YAP and the AP-1 factors JUN and FOSL1 is almost absent in MCF7 breast cancer cells. MCF10A cells express YAP and the two AP-1 factors, but not ZEB1.

(E) Control western blot for overexpression of ZEB1, V5 tagged YAP and FLAG tagged JUN/FOSL1 or different combinations of those factors in MCF7 cells.

Appendix table S1. HOMER known motif analysis on all ZEB1 peaks.

Table showing the 100 top enriched motifs (out of the HOMER motif database) identified in 200 bp regions centred on the ZEB1 peak summits. This database is maintained as part of the HOMER software and mostly based on the analysis of public ChIP-seq datasets. Motifs were grouped into those assumed to derive from the presence of the ZEB1 (blue), TEAD4 (red) or AP-1 (green) consensus motifs. Figure 1A shows one representative motif for each group and the rank of the group. The motifs named ZEB1, TEAD4 and AP-1 (in bold) were chosen as representative motif versions because they are present with the highest percentage in the ZEB1 peaks.

rank	motif	name/ChIP-seq dataset/software	p-value	log p-value	q-value (Benjamini)	# of binding sites with motif	% of binding sites with motif	# of background sequences with motif	% of background sequences with motif
1		ZEB1 (Zf)/PDAC-ZEB1-ChIP-Seq(GSE64557)/Homer	1e-1789	-4.119e+03	0.0000	8575.0	67.96%	10686.2	28.88%
2		Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer	1e-1181	-2.721e+03	0.0000	2487.0	19.71%	1136.7	3.07%
3		Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer	1e-1141	-2.628e+03	0.0000	2655.0	21.04%	1377.3	3.72%
4		JunB(bZIP)/DendriticCells-JunB-ChIP-Seq(GSE36099)/Homer	1e-1129	-2.600e+03	0.0000	2447.0	19.39%	1157.6	3.13%
5		BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer	1e-1111	-2.558e+03	0.0000	2576.0	20.42%	1327.3	3.59%
6		Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer	1e-1107	-2.550e+03	0.0000	2339.0	18.54%	1068.0	2.89%
7		E2A(bHLH)/near_PU.1/Bcell-PU.1-ChIP-Seq(GSE21512)/Homer	1e-1063	-2.450e+03	0.0000	6967.0	55.22%	9552.4	25.82%
8		AP-1 (bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-1016	-2.341e+03	0.0000	2708.0	21.46%	1633.6	4.42%
9		Fosl2(bZIP)/3T3L1-Fosl2-ChIP-Seq(GSE56872)/Homer	1e-997	-2.298e+03	0.0000	1935.0	15.34%	778.7	2.10%
10		Jun-AP1(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	1e-886	-2.041e+03	0.0000	1612.0	12.78%	586.7	1.59%
11		Bach2(bZIP)/OCILy7-Bach2-ChIP-Seq(GSE44420)/Homer	1e-427	-9.852e+02	0.0000	998.0	7.91%	490.6	1.33%
12		Slug(Zf)/Mesoderm-Snai2-ChIP-Seq(GSE61475)/Homer	1e-369	-8.504e+02	0.0000	3097.0	24.55%	4218.8	11.40%
13		Ptf1a(bHLH)/Panc01-Ptf1a-ChIP-Seq(GSE47459)/Homer	1e-363	-8.371e+02	0.0000	6981.0	55.33%	13835.1	37.39%
14		HEB(bHLH)/mES-Heb-ChIP-Seq(GSE53233)/Homer	1e-351	-8.088e+02	0.0000	6074.0	48.14%	11471.0	31.00%
15		E2A(bHLH)/proBcell-E2A-ChIP-Seq(GSE21978)/Homer	1e-339	-7.812e+02	0.0000	5251.0	41.62%	9433.4	25.50%
16		TEAD (TEA)/Fibroblast-PU.1-ChIP-Seq(Unpublished)/Homer	1e-332	-7.665e+02	0.0000	1599.0	12.67%	1531.4	4.14%
17		TEAD4 (TEA)/Tropoblast-Tea 4 -ChIP-Seq(GSE37350)/Homer	1e-317	-7.307e+02	0.0000	2112.0	16.74%	2500.2	6.76%
18		TEAD2 (TEA)/Py2T-Tea 2 -ChIP-Seq(GSE55709)/Homer	1e-278	-6.403e+02	0.0000	1454.0	11.52%	1467.3	3.97%
19		Ascl1(bHLH)/NeuralTubes-Ascl1-ChIP-Seq(GSE55840)/Homer	1e-212	-4.902e+02	0.0000	4393.0	34.82%	8360.0	22.59%
20		NF-E2(bZIP)/K562-NFE2-ChIP-Seq(GSE31477)/Homer	1e-182	-4.204e+02	0.0000	341.0	2.70%	124.1	0.34%
21		Bach1(bZIP)/K562-Bach1-ChIP-Seq(GSE31477)/Homer	1e-163	-3.767e+02	0.0000	289.0	2.29%	97.3	0.26%
22		Nrf2(bZIP)/Lymphoblast-Nrf2-ChIP-Seq(GSE37589)/Homer	1e-146	-3.365e+02	0.0000	285.0	2.26%	110.7	0.30%
23		Tbx5(T-box)/HL1-Tbx5 biotin-ChIP-Seq(GSE21529)/Homer	1e-143	-3.310e+02	0.0000	6166.0	48.87%	13939.9	37.68%
24		MafK(bZIP)/C2C12-MafK-ChIP-Seq(GSE36030)/Homer	1e-103	-2.384e+02	0.0000	648.0	5.14%	717.1	1.94%
25		CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer	1e-92	-2.126e+02	0.0000	541.0	4.29%	575.6	1.56%
26		BORIS(Zf)/K562-CTCF-L-ChIP-Seq(GSE32465)/Homer	1e-60	-1.404e+02	0.0000	859.0	6.81%	1373.6	3.71%

Appendix table S1. HOMER known motif analysis on all ZEB1 peaks. (continued)

rank	motif	name/ChIP-seq dataset/software	p-value	log p-value	q-value (Benjamini)	# of binding sites with motif	% of binding sites with motif	# of background sequences with motif	% of background sequences with motif
27		Tbet(T-box)/CD8-Tbet-ChIP-Seq(GSE33802)/Homer	1e-46	-1.073e+02	0.0000	1466.0	11.62%	2935.7	7.93%
28		MafA(bZIP)/Islet-MafA-ChIP-Seq(GSE30298)/Homer	1e-43	-9.983e+01	0.0000	1319.0	10.45%	2618.4	7.08%
29		NPAS2(bHLH)/Liver-NPAS2-ChIP-Seq(GSE39860)/Homer	1e-43	-9.979e+01	0.0000	2060.0	16.33%	4481.3	12.11%
30		RUNX1(Runt)/HPC7-Runx1-ChIP-Seq(GSE22178)/Homer	1e-42	-9.875e+01	0.0000	1085.0	8.60%	2061.0	5.57%
31		Eomes(T-box)/H9-Eomes-ChIP-Seq(GSE26097)/Homer	1e-42	-9.712e+01	0.0000	2538.0	20.12%	5749.8	15.54%
32		RUNX1(Runt)/Jurkat-RUNX1-ChIP-Seq(GSE29180)/Homer	1e-40	-9.288e+01	0.0000	1372.0	10.87%	2789.5	7.54%
33		MyoG(bHLH)/C2C12-MyoG-ChIP-Seq(GSE36024)/Homer	1e-39	-9.125e+01	0.0000	2223.0	17.62%	4970.4	13.43%
34		Flt1(ETS)/CD8-FLI-ChIP-Seq(GSE20898)/Homer	1e-37	-8.709e+01	0.0000	2293.0	18.17%	5188.8	14.02%
35		Bcl6(Zf)/Liver-Bcl6-ChIP-Seq(GSE31578)/Homer	1e-36	-8.517e+01	0.0000	1689.0	13.39%	3637.6	9.83%
36		KLF5(Zf)/LoVo-KLF5-ChIP-Seq(GSE49402)/Homer	1e-33	-7.717e+01	0.0000	4004.0	31.73%	9936.8	26.86%
37		ERG(ETS)/VCaP-ERG-ChIP-Seq(GSE14097)/Homer	1e-32	-7.371e+01	0.0000	2748.0	21.78%	6526.8	17.64%
38		NF1-halfsite(CTF)/LNCaP-NF1-ChIP-Seq(Unpublished)/Homer	1e-31	-7.323e+01	0.0000	2926.0	23.19%	7013.9	18.96%
39		MafB(bZIP)/BMM-MafB-ChIP-Seq(GSE75722)/Homer	1e-29	-6.899e+01	0.0000	607.0	4.81%	1085.6	2.93%
40		EWS:ERG-fusion(ETS)/CADO_ES1-EWS:ERG-ChIP-Seq(SRA014231)/Homer	1e-29	-6.794e+01	0.0000	969.0	7.68%	1950.4	5.27%
41		Twist(bHLH)/HMLE-TWIST1-ChIP-Seq(Chang_et_al)/Homer	1e-28	-6.583e+01	0.0000	355.0	2.81%	542.3	1.47%
42		Elv2(ETS)/ES-ER71-ChIP-Seq(GSE59402)/Homer(0.967)	1e-26	-6.117e+01	0.0000	1641.0	13.01%	3704.9	10.01%
43		RUNX2(Runt)/PCa-RUNX2-ChIP-Seq(GSE33889)/Homer	1e-26	-6.078e+01	0.0000	1064.0	8.43%	2231.4	6.03%
44		ETS1(ETS)/Jurkat-ETS1-ChIP-Seq(GSE17954)/Homer	1e-25	-5.939e+01	0.0000	1928.0	15.28%	4475.9	12.10%
45		NF1(CTF)/LNCaP-NF1-ChIP-Seq(Unpublished)/Homer	1e-24	-5.687e+01	0.0000	853.0	6.76%	1734.0	4.69%
46		Tlx?(NR)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer	1e-23	-5.504e+01	0.0000	810.0	6.42%	1640.6	4.43%
47		RUNX-AML(Runt)/CD4+-PolII-ChIP-Seq(Barski_et_al)/Homer	1e-23	-5.485e+01	0.0000	986.0	7.81%	2077.9	5.62%
48		Sp5(Zf)/mES-Sp5.Flag-ChIP-Seq(GSE72989)/Homer	1e-21	-4.856e+01	0.0000	3183.0	25.23%	8016.3	21.67%
49		KLF6(Zf)/PDAC-KLF6-ChIP-Seq(GSE64557)/Homer	1e-18	-4.355e+01	0.0000	3190.0	25.28%	8107.8	21.91%
50		ETV1(ETS)/GIST48-ETV1-ChIP-Seq(GSE22441)/Homer	1e-18	-4.327e+01	0.0000	2466.0	19.55%	6109.0	16.51%
51		Tgif2(Homeobox)/mES-Tgif2-ChIP-Seq(SE55404)/Homer	1e-18	-4.263e+01	0.0000	3655.0	28.97%	9423.9	25.47%
52		KLF3(Zf)/MEF-Klf3-ChIP-Seq(GSE44748)/Homer	1e-18	-4.239e+01	0.0000	1728.0	13.70%	4119.5	11.13%
53		NFAT:AP1(RHD,bZIP)/Jurkat-NFATC1-ChIP-Seq(Jolma_et_al)/Homer	1e-18	-4.225e+01	0.0000	187.0	1.48%	266.0	0.72%
54		Foxa2(Forkhead)/Liver-Foxa2-ChIPSeq(GSE25694)/Homer	1e-17	-3.967e+01	0.0000	638.0	5.06%	1316.1	3.56%
55		Tbr1(T-box)/Cortex-Tbr1-ChIP-Seq(GSE71384)/Homer	1e-17	-3.960e+01	0.0000	1446.0	11.46%	3398.6	9.19%

Appendix table S1. HOMER known motif analysis on all ZEB1 peaks. (continued)

rank	motif	name/ChIP-seq dataset/software	p-value	log p-value	q-value (Benjamini)	# of binding sites with motif	% of binding sites with motif	# of background sequences with motif	% of background sequences with motif
56		Foxa3(Forkhead)/Liver-Foxa3-ChIPSeq(GSE77670)/Homer	1e-17	-3.946e+01	0.0000	265.0	2.10%	439.0	1.19%
57		NFkB-p65-Rel(RHD)/ThioMac-LPSEexpression(GSE23622)/Homer	1e-17	-3.930e+01	0.0000	115.0	0.91%	135.7	0.37%
58		Pdx1(Homeobox)/Islet-Pdx1-ChIPSeq(SRA008281)/Homer	1e-15	-3.662e+01	0.0000	671.0	5.32%	1419.4	3.84%
59		Klf4(Zf)/mES-Klf4-ChIP-Seq(GSE11431)/Homer	1e-15	-3.616e+01	0.0000	1173.0	9.30%	2713.2	7.33%
60		ELF3(ETS)/PDAC-ELF3-ChIP-Seq(GSE64557)/Homer	1e-15	-3.544e+01	0.0000	998.0	7.91%	2263.7	6.12%
61		EHF(ETS)/LoVo-EHF-ChIP-Seq(GSE49402)/Homer	1e-15	-3.530e+01	0.0000	1878.0	14.88%	4609.0	12.46%
62		PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-15	-3.505e+01	0.0000	758.0	6.01%	1650.3	4.46%
63		Ets1-distal(ETS)/CD4+-PolII-ChIPSeq(Barski_et_al)/Homer	1e-14	-3.405e+01	0.0000	451.0	3.57%	895.2	2.42%
64		Stat3(Stat)/mES-Stat3-ChIP-Seq(GSE11431)/Homer	1e-14	-3.344e+01	0.0000	787.0	6.24%	1737.4	4.70%
65		Gata6(Zf)/HUG1N-GATA6-ChIP-Seq(GSE51936)/Homer	1e-14	-3.231e+01	0.0000	661.0	5.24%	1426.1	3.85%
66		EKLF(Zf)/Erythrocyte-Klf1-ChIP-Seq(GSE20478)/Homer	1e-13	-3.168e+01	0.0000	536.0	4.25%	1118.5	3.02%
67		Gata4(Zf)/Heart-Gata4-ChIP-Seq(GSE35151)/Homer	1e-13	-3.111e+01	0.0000	741.0	5.87%	1639.3	4.43%
68		GABPA(ETS)/Jurkat-GABPA-ChIPSeq(GSE17954)/Homer	1e-13	-3.088e+01	0.0000	1703.0	13.50%	4193.0	11.33%
69		EWS:FLI1-fusion(ETS)/SK_N_MC-EWS:FLI1-ChIP-Seq(SRA014231)/Homer	1e-13	-3.049e+01	0.0000	1030.0	8.16%	2395.7	6.47%
70		Rbpj1(7)/Panc1-Rbpj1-ChIP-Seq(GSE47459)/Homer	1e-13	-3.044e+01	0.0000	1923.0	15.24%	4798.3	12.97%
71		STAT4(Stat)/CD4-Stat4-ChIP-Seq(GSE22104)/Homer	1e-12	-2.953e+01	0.0000	963.0	7.63%	2229.9	6.03%
72		SPDEF(ETS)/VCaP-SPDEF-ChIPSeq(SRA014231)/Homer	1e-12	-2.934e+01	0.0000	1576.0	12.49%	3868.1	10.45%
73		Gata2(Zf)/K562-GATA2-ChIP-Seq(GSE18829)/Homer	1e-12	-2.884e+01	0.0000	503.0	3.99%	1056.8	2.86%
74		Gata1(Zf)/K562-GATA1-ChIP-Seq(GSE18829)/Homer	1e-12	-2.830e+01	0.0000	468.0	3.71%	973.1	2.63%
75		FOXA1(Forkhead)/LNCAP-FOXA1-ChIPSeq(GSE27824)/Homer	1e-12	-2.786e+01	0.0000	803.0	6.36%	1828.2	4.94%
76		Fox:Ebox(Forkhead,bHLH)/Panc1-Foxa2-ChIPSeq(GSE47459)/Homer	1e-11	-2.693e+01	0.0000	880.0	6.97%	2038.0	5.51%
77		FOXM1(Forkhead)/MCF7-FOXM1-ChIPSeq(GSE72977)/Homer	1e-10	-2.494e+01	0.0000	711.0	5.64%	1618.0	4.37%
78		Six1(Homeobox)/Myoblast-Six1-ChIPChip(GSE20150)/Homer	1e-10	-2.488e+01	0.0000	251.0	1.99%	470.8	1.27%
79		Tcf12(bHLH)/GM12878-Tcf12-ChIPSeq(GSE32465)/Homer	1e-10	-2.475e+01	0.0000	1941.0	15.38%	4934.9	13.34%
80		FOXA1(Forkhead)/MCF7-FOXA1-ChIPSeq(GSE26831)/Homer	1e-10	-2.339e+01	0.0000	642.0	5.09%	1453.7	3.93%
81		STAT6(Stat)/Macrophage-Stat6-ChIPSeq(GSE38377)/Homer	1e-9	-2.265e+01	0.0000	467.0	3.70%	1012.7	2.74%
82		GATA3(Zf)/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	1e-9	-2.145e+01	0.0000	1083.0	8.58%	2641.7	7.14%
83		Elk1(ETS)/Hela-Elk1-ChIP-Seq(GSE31477)/Homer	1e-9	-2.126e+01	0.0000	1353.0	10.72%	3375.4	9.12%
84		ETS:RUNX(ETS,Runt)/Jurkat-RUNX1-ChIPSeq(GSE17954)/Homer	1e-9	-2.116e+01	0.0000	249.0	1.97%	485.6	1.31%

Appendix table S1. HOMER known motif analysis on all ZEB1 peaks. (continued)

rank	motif	name/ChIP-seq dataset/software	p-value	log p-value	q-value (Benjamini)	# of binding sites with motif	% of binding sites with motif	# of background sequences with motif	% of background sequences with motif
85		Elk4(ETS)/Hela-Elk4-ChIP-Seq(GSE31477)/Homer	1e-9	-2.105e+01	0.0000	1328.0	10.53%	3310.5	8.95%
86		Sp1(Zf)/Promoter/Homer	1e-8	-2.030e+01	0.0000	1260.0	9.99%	3136.7	8.48%
87		Myf5(bHLH)/GM-Myf5-ChIP-Seq(GSE24852)/Homer	1e-8	-1.989e+01	0.0000	1203.0	9.53%	2987.8	8.08%
88		ELF5(ETS)/T47D-ELF5-ChIP-Seq(GSE30407)/Homer	1e-8	-1.963e+01	0.0000	1048.0	8.31%	2571.0	6.95%
89		STAT1(Stat)/HelaS3-STAT1-ChIP-Seq(GSE12346)/Homer	1e-8	-1.880e+01	0.0000	275.0	2.18%	562.9	1.52%
90		RBPJ:Ebox(?,bHLH)/Panc1-Rbpj1-ChIPSeq(GSE47459)/Homer	1e-7	-1.832e+01	0.0000	561.0	4.45%	1292.6	3.49%
91		STAT6(Stat)/CD4-Stat6-ChIP-Seq(GSE22104)/Homer	1e-7	-1.811e+01	0.0000	436.0	3.46%	971.2	2.62%
92		STAT5(Stat)/mCD4--Stat5-ChIP-Seq(GSE12346)/Homer	1e-7	-1.795e+01	0.0000	314.0	2.49%	664.3	1.80%
93		Maz(Zf)/HepG2-Maz-ChIP-Seq(GSE31477)/Homer	1e-7	-1.718e+01	0.0000	3623.0	28.72%	9832.8	26.58%
94		KLF10(Zf)/HEK293-KLF10.GFP-ChIPSeq(GSE58341)/Homer	1e-7	-1.706e+01	0.0000	1295.0	10.26%	3282.9	8.87%
95		Tcf3(HMG)/mES-Tcf3-ChIP-Seq(GSE11724)/Homer	1e-7	-1.689e+01	0.0000	202.0	1.60%	397.1	1.07%
96		Stat3+il21(Stat)/CD4-Stat3-ChIP-Seq(GSE19198)/Homer	1e-7	-1.668e+01	0.0000	871.0	6.90%	2134.2	5.77%
97		Six2(Homeobox)/NephronProgenitor-Six2-ChIPSeq(GSE39837)/Homer	1e-6	-1.560e+01	0.0000	861.0	6.82%	2122.0	5.74%
98		OCT:OCT-short(POU,Homeobox)/NPC-OCT6-ChIP-Seq(GSE43916)/Homer	1e-6	-1.557e+01	0.0000	394.0	3.12%	886.8	2.40%
99		Foxo3(Forkhead)/U2OS-Foxo3-ChIP-Seq(EMTAB-2701)/Homer	1e-6	-1.488e+01	0.0000	477.0	3.78%	1108.3	3.00%
100		Sox10(HMG)/SciaticNerve-Sox3-ChIPSeq(GSE35132)/Homer	1e-6	-1.434e+01	0.0000	1408.0	11.16%	3642.8	9.85%

Appendix table S2. HOMER *de novo* motif analysis on all ZEB1 peaks.

The three top enriched DNA-binding motifs identified by HOMER *de novo* motif analysis on 200 bp regions centred on the ZEB1 peak summits. The three identified motifs were compared to all known DNA-binding motifs present in the HOMER motif database and a similarity score was assigned. The HOMER motif database is maintained as part of the HOMER software and mostly based on the analysis of public ChIP-seq datasets. The similarity score indicates the percentage of identity between the *de novo* motif and individual known motifs.

rank	consensus sequence of the identified motif	p value	binding sites	background	known motifs identified by transcription factor ChIP-seq with a similarity score >0.9
1		1e-2194	82.5%	38.9%	ZEB1
2		1e-1175	20.4%	3.3%	FOSL1, JUNB, BATF, FOSL2, ATF3, JUN, AP-1
3		1e-317	19.6%	7.3%	TEAD, TEAD4, TEAD2

Appendix table S3. HOMER known motif analysis on ZEB1only peak set.

Table showing the 25 top enriched motifs (out of the HOMER motif database) identified in 200 bp regions centred on the ZEB1only peak summits. This database is maintained as part of the HOMER software and mostly based on the analysis of public ChIP-seq datasets. Motifs were grouped into those assumed to derive from the presence of the ZEB1 (blue), TEAD4 (red) or AP-1 (green) consensus motifs. Figure 5A shows one representative motif for each group and the rank of the group. To ensure comparability, the same representative motif versions were chosen as for figure 1A, ZEB1, TEAD4 and AP-1 (in bold).The two motifs of CTCF and its paralogue BORIS were also grouped together and the group was assigned rank two.

rank	motif	name/ChIP-seq dataset/software	p-value	log p-value	q-value (Benjamini)	# of binding sites with motif	% of binding sites with motif	# of background sequences with motif	% of background sequences with motif
1		ZEB1(Zf)/PDAC-ZEB1-ChIP-Seq(GSE64557)/Homer	1e-1432	-3.298e+03	0.0000	4692.0	78.69%	12116.9	28.05%
2		E2A(bHLH),near_PU.1/Bcell-PU.1-ChIP-Seq(GSE21512)/Homer	1e-899	-2.070e+03	0.0000	3843.0	64.45%	10769.0	24.93%
3		Slug(Zf)/Mesoderm-Snai2-ChIP-Seq(GSE61475)/Homer	1e-329	-7.591e+02	0.0000	1788.0	29.98%	4891.6	11.32%
4		Ptf1a(bHLH)/Panc1-Ptf1a-ChIP-Seq(GSE47459)/Homer	1e-316	-7.284e+02	0.0000	3621.0	60.72%	15725.5	36.40%
5		HEB(bHLH)/mES-Heb-ChIP-Seq(GSE53233)/Homer	1e-305	-7.043e+02	0.0000	3248.0	54.47%	13405.2	31.03%
6		E2A(bHLH)/proBcell-E2A-ChIP-Seq(GSE21978)/Homer	1e-300	-6.914e+02	0.0000	2851.0	47.81%	10999.6	25.46%
7		Ascl1(bHLH)/NeuralTubes-Ascl1-ChIP-Seq(GSE55840)/Homer	1e-186	-4.296e+02	0.0000	2380.0	39.91%	9918.7	22.96%
8		CTCF(Zf)/CD4-CTCF-ChIP-Seq(Barski_et_al)/Homer	1e-128	-2.967e+02	0.0000	431.0	7.23%	758.2	1.76%
9		Tbx5(T-box)/HL1-Tbx5.biotin-ChIP-Seq(GSE21529)/Homer	1e-96	-2.226e+02	0.0000	2814.0	47.19%	14709.0	34.05%
10		BORIS(Zf)/K562-CTCF-ChIP-Seq(GSE32465)/Homer	1e-81	-1.867e+02	0.0000	663.0	11.12%	2126.6	4.92%
11		NPAS2(bHLH)/Liver-NPAS2-ChIP-Seq(GSE39860)/Homer	1e-48	-1.109e+02	0.0000	1065.0	17.86%	4923.2	11.40%
12		Eomes(T-box)/H9-Eomes-ChIP-Seq(GSE26097)/Homer	1e-37	-8.526e+01	0.0000	1042.0	17.47%	5097.7	11.80%
13		MyoG(bHLH)/C2C12-MyoG-ChIP-Seq(GSE36024)/Homer	1e-27	-6.263e+01	0.0000	1131.0	18.97%	5988.7	13.86%
14		Twist(bHLH)/HMLE-TWIST1-ChIP-Seq(Chang_et_al)/Homer	1e-25	-5.951e+01	0.0000	183.0	3.07%	542.4	1.26%
15		TEAD4(TEA)/Tropoblast-Tea4-ChIP-Seq(GSE37350)/Homer	1e-23	-5.376e+01	0.0000	486.0	8.15%	2183.4	5.05%
16		TEAD2(TEA)/Py2T-Tea2-ChIP-Seq(GSE55709)/Homer	1e-22	-5.087e+01	0.0000	303.0	5.08%	1195.0	2.77%
17		Tbet(T-box)/CD8-Tbet-ChIP-Seq(GSE33802)/Homer	1e-20	-4.771e+01	0.0000	535.0	8.97%	2541.2	5.88%
18		TEAD(TEA)/Fibroblast-PU.1-ChIP-Seq(Unpublished)/Homer	1e-17	-4.035e+01	0.0000	276.0	4.63%	1142.2	2.64%
19		Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer	1e-16	-3.829e+01	0.0000	212.0	3.56%	817.1	1.89%
20		Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer	1e-16	-3.828e+01	0.0000	251.0	4.21%	1025.0	2.37%
21		BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer	1e-14	-3.401e+01	0.0000	232.0	3.89%	961.7	2.23%
22		JunB(bZIP)/DendriticCells-JunB-ChIP-Seq(GSE36099)/Homer	1e-14	-3.338e+01	0.0000	211.0	3.54%	853.8	1.98%
23		Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer	1e-13	-3.066e+01	0.0000	195.0	3.27%	791.2	1.83%
24		AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-12	-2.773e+01	0.0000	262.0	4.39%	1195.7	2.77%
25		Bach1(bZIP)/K562-Bach1-ChIP-Seq(GSE31477)/Homer	1e-10	-2.363e+01	0.0000	32.0	0.54%	56.5	0.13%

Appendix table S4. HOMER known motif analysis on ZEB1/YAP/JUN peak set.

Table showing the 25 top enriched motifs (out of the HOMER motif database) identified in 200 bp regions centred on the ZEB1/YAP/JUN peak summits. This database is maintained as part of the HOMER software and mostly based on the analysis of public ChIP-seq datasets. Motifs were grouped into those assumed to derive from the presence of the ZEB1 (blue), TEAD4 (red) or AP-1 (green) consensus motifs. Figure 5A shows one representative motif for each group and the rank of the group. To ensure comparability, the same representative motif versions were chosen as for figure 1A, ZEB1, TEAD4 and AP-1 (in bold). The three RUNX motifs present in the list were also grouped together and the group was assigned rank three.

rank	motif	name/ChIP-seq dataset/software	p-value	log p-value	q-value (Benjamini)	# of binding sites with motif	% of binding sites with motif	# of background sequences with motif	% of background sequences with motif
1		Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer	1e-613	-1.412e+03	0.0000	896.0	44.96%	2257.7	4.77%
2		Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer	1e-599	-1.381e+03	0.0000	956.0	47.97%	2828.6	5.98%
3		BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer	1e-572	-1.317e+03	0.0000	919.0	46.11%	2717.2	5.75%
4		Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer	1e-566	-1.304e+03	0.0000	831.0	41.70%	2064.6	4.37%
5		JunB(bZIP)/DendriticCells-Junb-ChIP-Seq(GSE36099)/Homer	1e-561	-1.294e+03	0.0000	865.0	43.40%	2350.9	4.97%
6		AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-557	-1.284e+03	0.0000	975.0	48.92%	3330.5	7.04%
7		Fosl2(bZIP)/3T3L1-Fosl2-ChIP-Seq(GSE56872)/Homer	1e-474	-1.092e+03	0.0000	659.0	33.07%	1416.6	3.00%
8		Jun-AP1(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	1e-413	-9.514e+02	0.0000	543.0	27.25%	1020.7	2.16%
9		TEAD4(TEA)/Tropoblast-Tea4-ChIP-Seq(GSE37350)/Homer	1e-357	-8.231e+02	0.0000	874.0	43.85%	4395.9	9.30%
10		TEAD(TEA)/Fibroblast-PU.1-ChIP-Seq(Unpublished)/Homer	1e-317	-7.311e+02	0.0000	744.0	37.33%	3442.0	7.28%
11		TEAD2(TEA)/Py2T-Tea2-ChIP-Seq(GSE5709)/Homer	1e-283	-6.527e+02	0.0000	644.0	32.31%	2799.7	5.92%
12		Bach2(bZIP)/OCILy7-Bach2-ChIP-Seq(GSE44420)/Homer	1e-191	-4.412e+02	0.0000	318.0	15.96%	838.9	1.77%
13		NF-E2(bZIP)/K562-NFE2-ChIP-Seq(GSE31477)/Homer	1e-84	-1.952e+02	0.0000	118.0	5.92%	225.6	0.48%
14		Nrf2(bZIP)/Lymphoblast-Nrf2-ChIP-Seq(GSE37589)/Homer	1e-70	-1.631e+02	0.0000	101.0	5.07%	201.2	0.43%
15		Bach1(bZIP)/K562-Bach1-ChIP-Seq(GSE31477)/Homer	1e-59	-1.366e+02	0.0000	87.0	4.37%	182.8	0.39%
16		MatK(bZIP)/C2C12-MaTK-ChIP-Seq(GSE36030)/Homer	1e-49	-1.149e+02	0.0000	173.0	8.68%	1048.0	2.22%
17		RUNX(Runt)/HPC7-Runx1-ChIP-Seq(GSE22178)/Homer	1e-43	-1.009e+02	0.0000	308.0	15.45%	3065.2	6.48%
18		RUNX-AML(Runt)/CD4+-Poli-ChIP-Seq(Barski_et_al.)/Homer	1e-38	-8.791e+01	0.0000	300.0	15.05%	3156.7	6.68%
19		RUNX2(Runt)/PCa-RUNX2-ChIP-Seq(GSE33889)/Homer	1e-35	-8.087e+01	0.0000	322.0	16.16%	3648.2	7.72%
20		ZEB1(Zf)/PDAC-ZEB1-ChIP-Seq(GSE64557)/Homer	1e-34	-8.031e+01	0.0000	778.0	39.04%	12422.6	26.27%
21		RUNX1(Runt)/Jurkat-RUNX1-ChIP-Seq(GSE29180)/Homer	1e-32	-7.473e+01	0.0000	363.0	18.21%	4484.5	9.48%
22		MafA(bZIP)/Islet-MafA-ChIP-Seq(GSE30298)/Homer	1e-27	-6.237e+01	0.0000	301.0	15.10%	3687.7	7.80%
23		MatB(bZIP)/BMM-MaB-ChIP-Seq(GSE75722)/Homer	1e-22	-5.250e+01	0.0000	173.0	8.68%	1772.8	3.75%
24		Bcl6(Zf)/Liver-Bcl6-ChIP-Seq(GSE31578)/Homer	1e-22	-5.187e+01	0.0000	411.0	20.62%	6001.4	12.69%
25		ETS1(ETS)/Jurkat-ETS1-ChIP-Seq(GSE17954)/Homer	1e-21	-4.978e+01	0.0000	355.0	17.81%	5002.5	10.58%

Appendix table S5. *de novo* HOMER motif analysis on ZEB1 peak subsets.

DNA sequence motifs identified within ZEB1-only **(A)** and ZEB1/YAP/JUN peaks **(B)** by the HOMER *de novo* motif search algorithm in 200 bp regions centred on ZEB1 peak summit positions. Motifs matching the ZEB1, TEAD or AP-1 consensus motifs are shown. Their respective rank in the *de novo* motif analysis is indicated in column one. The consensus sequences of the identified motifs were compared to all known DNA-binding motifs in the HOMER motif database and a similarity score was assigned. The motif with the highest similarity score and the motif shown in figure 5 are listed (set in parenthesis if different from the one with the highest similarity score).

A

de novo HOMER motif analysis: ZEB1-only

rank	consensus sequence of the identified motif	p value	binding sites	background	best matched motif (motif shown in main figure)	similarity score
1		1e-1498	86.0%	34.2%	ZEB1	0.97
7		1e-24	2.5%	1.0%	TEAD3(TEAD4)	0.91(0.90)
10		1e-15	3.4%	1.8%	FOS::JUN(AP-1)	0.94(0.89)

B

de novo HOMER motif analysis: ZEB1/YAP/JUN

rank	consensus sequence of the identified motif	p value	binding sites	background	best matched motif (motif shown in main figure)	similarity score
1		1e-607	46.9%	5.5%	FRA-1 (AP-1)	0.99 (0.97)
2		1e-374	36.2%	5.5%	TEAD3 (TEAD4)	0.99 (0.98)
3		1e-82	51.5%	30.7%	ZEB1	0.84

Appendix table S6. HOMER known motif analysis on ZEB1 repressed regions (identified by ATAC-seq).

Table showing the 25 top enriched motifs (out of the HOMER motif database) identified in 200 bp regions centred on ZEB1 peak summits which overlap with genomic regions which are repressed in the presence of ZEB1. This database is maintained as part of the HOMER software and mostly based on the analysis of public ChIP-seq datasets. Motifs were grouped into those assumed to derive from the presence of the ZEB1 (blue), TEAD4 (red) or AP-1 (green) consensus motifs. Figure 6A shows one representative motif for each group and the rank of the group. To ensure comparability, the same representative motif versions were chosen as for figure 1A, ZEB1, TEAD4 and AP-1 (in bold). The two motifs of CTCF and its paralogue BORIS were also grouped together and the group was assigned rank three.

rank	motif	name/ChIP-seq dataset/software	p-value	log p-value	q-value (Benjamini)	# of binding sites with motif	% of binding sites with motif	# of background sequences with motif	% of background sequences with motif
1		ZEB1(Zf)/PDAC-ZEB1-ChIP-Seq(GSE64557)/Homer	1e-912	-2.100e+03	0.0000	3011.0	78.13%	12717.0	27.89%
2		E2A(bHLH).near_PU.1/Bcell-PU.1-ChIP-Seq(GSE21512)/Homer	1e-554	-1.276e+03	0.0000	2440.0	63.31%	11310.9	24.80%
3		Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer	1e-269	-6.199e+02	0.0000	592.0	15.36%	1134.4	2.49%
4		JunB(bZIP)/DendriticCells-Junb-ChIP-Seq(GSE36099)/Homer	1e-254	-5.853e+02	0.0000	584.0	15.15%	1178.6	2.58%
5		Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer	1e-252	-5.805e+02	0.0000	551.0	14.30%	1043.6	2.29%
6		Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer	1e-247	-5.693e+02	0.0000	619.0	16.06%	1382.0	3.03%
7		BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer	1e-247	-5.688e+02	0.0000	610.0	15.83%	1340.4	2.94%
8		Fosl2(bZIP)/3T3L1-Fosl2-ChIP-Seq(GSE56872)/Homer	1e-246	-5.681e+02	0.0000	485.0	12.58%	796.6	1.75%
9		Jun-AP1(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	1e-226	-5.206e+02	0.0000	413.0	10.72%	608.2	1.33%
10		AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-209	-4.830e+02	0.0000	615.0	15.96%	1616.1	3.54%
11		Ptf1a(bHLH)/Panc1-Ptf1a-ChIP-Seq(GSE47459)/Homer	1e-182	-4.192e+02	0.0000	2270.0	58.90%	16425.4	36.02%
12		HEB(bHLH)/mES-Heb-ChIP-Seq(GSE53233)/Homer	1e-172	-3.981e+02	0.0000	2004.0	52.00%	13795.3	30.25%
13		Slug(Zf)/Mesoderm-Snai2-ChIP-Seq(GSE61475)/Homer	1e-164	-3.797e+02	0.0000	1049.0	27.22%	5090.1	11.16%
14		E2A(bHLH)/proBcell-E2A-ChIP-Seq(GSE21978)/Homer	1e-161	-3.724e+02	0.0000	1736.0	45.04%	11345.2	24.88%
15		CTCF(Zf)/CD4+CTCF-ChIP-Seq(Barski_et_al)/Homer	1e-123	-2.853e+02	0.0000	325.0	8.43%	743.0	1.63%
16		Bach2(bZIP)/OCI-Ly7-Bach2-ChIP-Seq(GSE44420)/Homer	1e-118	-2.733e+02	0.0000	263.0	6.82%	493.6	1.08%
17		Ascl1(bHLH)/NeuralTubes-Ascl1-ChIP-Seq(GSE55840)/Homer	1e-108	-2.487e+02	0.0000	1469.0	38.12%	10164.5	22.29%
18		BORIS(Zf)/K562-CTCF-ChIP-Seq(GSE32465)/Homer	1e-74	-1.707e+02	0.0000	451.0	11.70%	2031.3	4.45%
19		Tbx5(T-box)/HL1-Tbx5.biotin-ChIP-Seq(GSE21529)/Homer	1e-62	-1.435e+02	0.0000	1861.0	48.29%	16025.7	35.14%
20		TEAD4(TEA)/Tropoblast-Tea4-ChIP-Seq(GSE37350)/Homer	1e-51	-1.174e+02	0.0000	446.0	11.57%	2423.1	5.31%
21		NF-E2(bZIP)/K562-NFE2-ChIP-Seq(GSE31477)/Homer	1e-48	-1.108e+02	0.0000	89.0	2.31%	132.4	0.29%
22		Bach1(bZIP)/K562-Bach1-ChIP-Seq(GSE31477)/Homer	1e-47	-1.094e+02	0.0000	76.0	1.97%	92.0	0.20%
23		TEAD2(TEA)/Py2-Tead2-ChIP-Seq(GSE5709)/Homer	1e-41	-9.626e+01	0.0000	287.0	7.45%	1372.8	3.01%
24		TEAD(TEA)/Fibroblast-PU.1-ChIP-Seq(Unpublished)/Homer	1e-40	-9.327e+01	0.0000	286.0	7.42%	1388.5	3.04%
25		Nrf2(bZIP)/Lymphoblast-Nrf2-ChIP-Seq(GSE37589)/Homer	1e-38	-8.772e+01	0.0000	69.0	1.79%	100.2	0.22%

Appendix table S7. HOMER known motif analysis on ZEB1 activated regions (identified by ATAC-seq).

Table showing the 25 top enriched motifs (out of the HOMER motif database) identified in 200 bp regions centred on ZEB1 peak summits which overlap with genomic regions which are activated in the presence of ZEB1. This database is maintained as part of the HOMER software and mostly based on the analysis of public ChIP-seq datasets. Motifs were grouped into those assumed to derive from the presence of the ZEB1 (blue), TEAD4 (red) or AP-1 (green) consensus motifs. Figure 6A shows one representative motif for each group and the rank of the group. To ensure comparability, the same representative motif versions were chosen as for figure 1A, ZEB1, TEAD4 and AP-1 (in bold).

rank	motif	name/ChIP-seq dataset/software	p-value	log p-value	q-value (Benjamini)	# of binding sites with motif	% of binding sites with motif	# of background sequences with motif	% of background sequences with motif
1		Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer	1e-316	-7.292e+02	0.0000	685.0	29.49%	2359.5	5.02%
2		JunB(bZIP)/DendriticCells-Junb-ChIP-Seq(GSE36099)/Homer	1e-310	-7.151e+02	0.0000	624.0	26.86%	1933.7	4.11%
3		Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer	1e-307	-7.080e+02	0.0000	617.0	26.56%	1905.2	4.05%
4		BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer	1e-297	-6.851e+02	0.0000	652.0	28.07%	2265.4	4.82%
5		Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer	1e-295	-6.799e+02	0.0000	581.0	25.01%	1735.9	3.69%
6		AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-295	-6.795e+02	0.0000	706.0	30.39%	2748.1	5.84%
7		FosI2(bZIP)/3T3L1-FosI2-ChIP-Seq(GSE56872)/Homer	1e-254	-5.865e+02	0.0000	462.0	19.89%	1204.8	2.56%
8		Jun-AP1(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	1e-211	-4.861e+02	0.0000	367.0	15.80%	883.5	1.88%
9		TEAD(TEA)/Fibroblast-PU.1-ChIP-Seq(Unpublished)/Homer	1e-157	-3.630e+02	0.0000	520.0	22.38%	2687.2	5.71%
10		TEAD4(TEA)/Tropoblast-Tead4-ChIP-Seq(GSE37350)/Homer	1e-145	-3.347e+02	0.0000	602.0	25.91%	3794.6	8.07%
11		ZEB1(Zf)/PDAC-ZEB1-ChIP-Seq(GSE64557)/Homer	1e-129	-2.973e+02	0.0000	1239.0	53.34%	13747.8	29.23%
12		TEAD2(TEA)/Py2T-Tead2-ChIP-Seq(GSE55709)/Homer	1e-121	-2.790e+02	0.0000	428.0	18.42%	2310.1	4.91%
13		Bach2(bZIP)/OCILy7-Bach2-ChIP-Seq(GSE44420)/Homer	1e-113	-2.622e+02	0.0000	233.0	10.03%	687.7	1.46%
14		E2A(bHLH),near_PU.1/Bcell-PU.1-ChIP-Seq(GSE21512)/Homer	1e-64	-1.476e+02	0.0000	981.0	42.23%	12213.2	25.97%
15		RUNX(Runt)/HPC7-Runx1-ChIP-Seq(GSE22178)/Homer	1e-52	-1.203e+02	0.0000	356.0	15.33%	2966.4	6.31%
16		NF-E2(bZIP)/K562-NFE2-ChIP-Seq(GSE31477)/Homer	1e-47	-1.104e+02	0.0000	82.0	3.53%	187.1	0.40%
17		Bach1(bZIP)/K562-Bach1-ChIP-Seq(GSE31477)/Homer	e-47	-1.089e+02	0.0000	72.0	3.10%	137.6	0.29%
18		Nrf2(bZIP)/Lymphoblast-Nrf2-ChIP-Seq(GSE37589)/Homer	1e-46	-1.076e+02	0.0000	74.0	3.19%	150.5	0.32%
19		RUNX2(Runt)/PCa-RUNX2-ChIP-Seq(GSE33889)/Homer	1e-40	-9.400e+01	0.0000	377.0	16.23%	3649.0	7.76%
20		RUNX1(Runt)/Jurkat-RUNX1-ChIP-Seq(GSE29180)/Homer	1e-38	-8.830e+01	0.0000	421.0	18.12%	4396.5	9.35%
21		RUNX-AML(Runt)/CD4+-PolII-ChIP-Seq(Barski_et_al.)/Homer	1e-37	-8.568e+01	0.0000	337.0	14.51%	3222.8	6.85%
22		ETS1(ETS)/Jurkat-ETS1-ChIP-Seq(GSE17954)/Homer	1e-35	-8.135e+01	0.0000	463.0	19.93%	5181.0	11.02%
23		ERG(ETS)/VcAP-ERG-ChIP-Seq(GSE14097)/Homer	1e-31	-7.256e+01	0.0000	632.0	27.21%	8165.5	17.36%
24		Flt1(ETS)/CD8-Flt1-ChIP-Seq(GSE20898)/Homer	1e-30	-7.081e+01	0.0000	473.0	20.36%	5586.5	11.88%
25		Etv2(ETS)/ES-ER71-ChIP-Seq(GSE59402)/Homer(0.967)	1e-30	-6.920e+01	0.0000	406.0	17.48%	4569.7	9.72%

Appendix table S8. *de novo* HOMER motif analysis on ZEB1 peaks associated with active or repressed genomic regions.

DNA sequence motifs identified within ZEB1 peaks located in repressed **(A)** and activated **(B)** genomic regions by the HOMER *de novo* motif search algorithm in 200 bp regions centred on ZEB1 peak summit positions. Motifs matching the ZEB1, TEAD or AP-1 consensus motifs are shown. Their respective rank in the *de novo* motif analysis is indicated in column one. The consensus sequences of the identified motifs were compared to all known DNA-binding motifs in the HOMER motif database and a similarity score was assigned. The motif with the highest similarity score and the motif shown in figure 6 (set in parenthesis if different from the one with the highest similarity score) are listed.

A *de novo* HOMER motif analysis: ATACseq - ZEB1 repressed

rank	consensus sequence of the identified motif	p value	binding sites	background	best matched motif (motif shown in main figure)	similarity score
1		1e-974	87.0%	35.0%	ZEB1	0.96
2		1e-265	14.4%	2.2%	FRA-1 (AP-1)	0.99 (0.96)
4		1e-58	6.7%	2.1%	TEAD3 (TEAD4)	0.95 (0.93)

B *de novo* HOMER motif analysis: ATACseq - ZEB1 activated

rank	consensus sequence of the identified motif	p value	binding sites	background	best matched motif (motif shown in main figure)	similarity score
1		1e-326	29.1%	4.7%	cJUN (AP-1)	0.99 (0.98)
2		1e-196	80.3%	50.3%	ZEB1	0.93
3		1e-192	34.5%	11.2%	TEAD4	0.97

Appendix table S9. ZEB1/YAP/JUN activated target genes positively correlate with ZEB1 expression in breast cancer cell lines.

Spearman's rank correlation of mRNA expression between ZEB1 and the top 23 ZEB1/YAP/JUN activated target genes as well as four representative ZEB1-only repressed target genes in all CCLE breast cancer cell lines. The activated targets are all genes which had overlapping ZEB1, YAP and JUN peaks in their promoter region and a log fold change (logFC) < -2 in ZEB1 knock-down compared to control MDA-MB-231 cells in our microarray analysis, thus they are strongly activated in the presence of ZEB1. Known epithelial genes were chosen as repressed targets.

	gene set	gene	spearman's rank correlation	p-value	significance
1	ZEB1/YAP/JUN activated	NRP1	0.59	1.41E-06	***
2	ZEB1/YAP/JUN activated	CTGF	0.53	2.26E-05	***
3	ZEB1/YAP/JUN activated	THBS1	0.44	7.27E-04	**
4	ZEB1/YAP/JUN activated	CYR61	0.59	1.46E-06	***
5	ZEB1/YAP/JUN activated	MAP1B	0.75	3.95E-11	***
6	ZEB1/YAP/JUN activated	ADAMTS6	0.76	1.28E-11	***
7	ZEB1/YAP/JUN activated	TGFBI	0.65	6.42E-08	***
8	ZEB1/YAP/JUN activated	ANKRD1	0.55	1.19E-05	***
9	ZEB1/YAP/JUN activated	FRMD6	0.71	1.12E-09	***
10	ZEB1/YAP/JUN activated	LATS2	0.47	2.71E-04	**
11	ZEB1/YAP/JUN activated	CSF2	0.47	2.85E-04	**
12	ZEB1/YAP/JUN activated	RHOBTB3	0.20	1.47E-01	
13	ZEB1/YAP/JUN activated	EDN1	-0.05	7.25E-01	
14	ZEB1/YAP/JUN activated	S100A10	0.36	6.99E-03	**
15	ZEB1/YAP/JUN activated	CAP2	0.01	9.13E-01	
16	ZEB1/YAP/JUN activated	TRIM34	-0.14	2.91E-01	
17	ZEB1/YAP/JUN activated	SLC20A1	0.37	5.64E-03	**
18	ZEB1/YAP/JUN activated	ARHGAP24	0.36	5.68E-03	**
19	ZEB1/YAP/JUN activated	SAMD4A	0.56	8.54E-06	***
20	ZEB1/YAP/JUN activated	LRRRC8A	0.37	4.75E-03	**
21	ZEB1/YAP/JUN activated	PTPN14	0.56	6.43E-06	***
22	ZEB1/YAP/JUN activated	AJUBA	0.22	9.88E-02	
23	ZEB1/YAP/JUN activated	EHD1	0.08	5.57E-01	
24	ZEB1-only repressed	CDH1	-0.43	9.58E-04	***
25	ZEB1-only repressed	LLGL2	-0.67	1.50E-08	***
26	ZEB1-only repressed	CLDN7	-0.56	6.17E-06	***
27	ZEB1-only repressed	CLDN3	-0.71	1.12E-09	***

Appendix table S10. ZEB1/YAP/JUN activated target genes positively correlate with ZEB1 expression in human breast cancer tumours.

Spearman's rank correlation of mRNA expression between ZEB1 and the top 23 ZEB1/YAP/JUN activated target genes as well as four representative ZEB1-only repressed target genes in 188 breast cancer patient samples (GSE18229) . The activated targets are all genes which have overlapping ZEB1, YAP and JUN peaks in their promoter region and a log fold change (logFC) < -2 in ZEB1 knock-down compared to control MDA-MB-231 cells in our microarray analysis, thus they are strongly activated in the presence of ZEB1. Known epithelial genes were chosen as repressed targets.

	gene set	gene	Spearman's rank correlation	p-value	significance
1	ZEB1/YAP/JUN activated	NRP1	0.34	2.27E-06	***
2	ZEB1/YAP/JUN activated	CTGF	0.42	1.34E-09	***
3	ZEB1/YAP/JUN activated	THBS1	0.31	1.42E-05	***
4	ZEB1/YAP/JUN activated	CYR61	0.44	3.53E-10	***
5	ZEB1/YAP/JUN activated	MAP1B	0.38	6.22E-08	***
6	ZEB1/YAP/JUN activated	ADAMTS6	0.35	1.18E-06	***
7	ZEB1/YAP/JUN activated	TGFB1	0.33	4.94E-06	***
8	ZEB1/YAP/JUN activated	ANKRD1	-0.07	3.12E-01	
9	ZEB1/YAP/JUN activated	FRMD6	0.51	4.12E-14	***
10	ZEB1/YAP/JUN activated	LATS2	0.62	2.43E-21	***
11	ZEB1/YAP/JUN activated	CSF2	-0.14	4.87E-02	*
12	ZEB1/YAP/JUN activated	RHOBTB3	0.20	6.26E-03	*
13	ZEB1/YAP/JUN activated	EDN1	0.13	8.66E-02	
14	ZEB1/YAP/JUN activated	S100A10	0.23	1.57E-03	*
15	ZEB1/YAP/JUN activated	CAP2	0.09	2.18E-01	
16	ZEB1/YAP/JUN activated	TRIM34	0.26	2.95E-04	**
17	ZEB1/YAP/JUN activated	SLC20A1	0.17	2.35E-02	*
18	ZEB1/YAP/JUN activated	ARHGAP24	0.39	2.86E-08	***
19	ZEB1/YAP/JUN activated	SAMD4A	no probe in array		
20	ZEB1/YAP/JUN activated	LRRC8A	0.23	1.79E-03	*
21	ZEB1/YAP/JUN activated	PTPN14	-0.09	2.49E-01	
22	ZEB1/YAP/JUN activated	AJUBA	0.15	4.43E-02	*
23	ZEB1/YAP/JUN activated	EHD1	0.00	9.83E-01	
24	ZEB1-only repressed	CDH1	-0.22	2.71E-03	*
25	ZEB1-only repressed	LLGL2	-0.14	5.74E-02	
26	ZEB1-only repressed	CLDN7	-0.35	1.21E-06	***
27	ZEB1-only repressed	CLDN3	-0.29	6.11E-05	***

Appendix table S11. Oligonucleotides used for cloning.

Restriction enzyme sites are underlined.

name	sequence (5' - 3')
hJUN-FLAG_fw	ATAATGAATTCATGACTGCAAAGATGGAAACGAC
hJUN-FLAG_rev	ATAATTCTAGATTACTTGTCGTCATCGTCTTTGTAGTCAAATGTTTGCAACTGCTGCGTTAGC
hFOSL1-FLAG_fw	ATAATGAATTCATGTTCCGAGACTTCGGGGAACC
hFOSL1-FLAG_rev	ATAATTCTAGATTACTTGTCGTCATCGTCTTTGTAGTCCAAAGCGAGGAGGGTTGGAGAGCC
hDOCK9_enh_fw	ATAATCTCGAGATGGGTGGCACCTGCACAATG
hDOCK9_enh_rev	ATAATAAGCTTTTAGTGTTTGTTGAGTCAATGAATC
hANKRD1_prom-425_fw	ATAATCTCGAGAGTGCATTACTGAATGCTTTCA
hANKRD1_prom+120_rev	ATAATAAGCTTCCGAAGTAACACTCCCACCC
hANKRD1_prom+73_rev	ATAATAAGCTTGAATTTTCATTCCAGACTCA
hANKRD1_prom_mutZbox_fw	GTCTGGAATGAAAATTTACCCGCCTCTGAGTTGGCTCC
hANKRD1_prom_mutZvox_rev	GGAGCCAACTCAGAGGCGGGTAAATTTTCATTCCAGAC

Appendix table S12. Oligonucleotides used as EMSA probes.

DNA-oligonucleotides were used as double-stranded probes.

name	sequence (5' - 3')
hLLGL2_E2/3_s	CTGCGTCCAGGTGCGCGCAGGTGAGGCCGG
hLLGL2_E2/3_as	CCGGCCTCACCTGCGCGCACCTGGACGCAG
hANKRD1_AP1_s	CATTTCTTCTGATTACATATTC
hANKRD1_AP1_as	GAATATGTGAATCAGGAAGAAATG
hANKRD1_MCAT_s	GATGACTTGCAATTGCTGAGCGATGTG
hANKRD1_MCAT_as	CACATCGCTCAGCAATGCAAGTCATC

Appendix table S13. Oligonucleotides used for qPCR.

name	sense (5' - 3')	anti-sense (5' - 3')
hGAPDH	agccacatcgctcagacac	gccaatacgaccaaattcc
hZEB1	aactgctgggaggatgacac	tcctgctcatctgcctga
hYAP1	atcccagcacagcaaattct	tggatttgagtcccacat
hFOSL1	gggcatgttccgagacttc	tggcaccagggtgaactt
hCDH1	ttgacgccgagagctacac	gtcgaccgggtgcaattt
hANKRD1	cagtggaaaagcgagaaaca	ttcaagcttgatcttgttctagt
hCTGF	cctgcaggctagagaagcag	tggagatttgggagtacgg
hTHBS1	caatgccacagttcctgatg	tggagaccagccatcgtc
hCYR61	aagaaacccggatttgtgag	gctgcatttctgcccttt
hTGFB1	cgagtgtgtcctggatatg	cccaggggtctcgtaaagggt
hNRP1	taccctgagaatgggtggac	cgtagaaaagcgcagaag
hEDN1	tcctgtgtgttggtgctg	gagctcagcgctaagactg
hS100A10	aaggagttccctggattttg	cactggtccaggctcctcat
hMAP1B	attcctgggcaaactggtct	aaacgtcacttcggtgatga
hEDH1	caacgccttcctcaacaggt	gtcgatgatgctgatgctgt
hDOCK9	gacgaggaggtcgacaaaga	ttggtgatcccacccttct