Supplementary Information

Knopfova L et al.

Transcription factor c-Myb inhibits breast cancer lung metastasis by suppression of tumor cell seeding

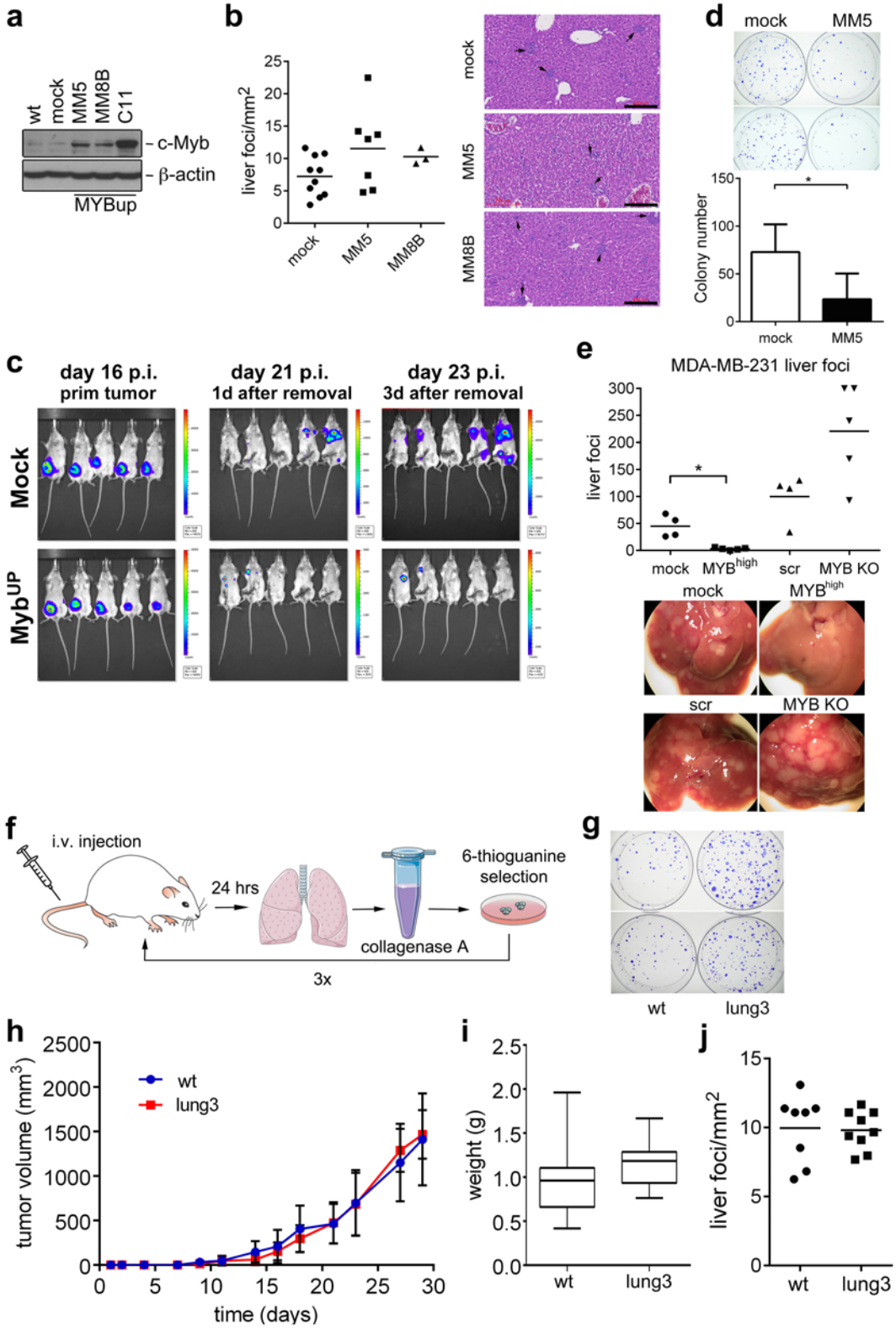
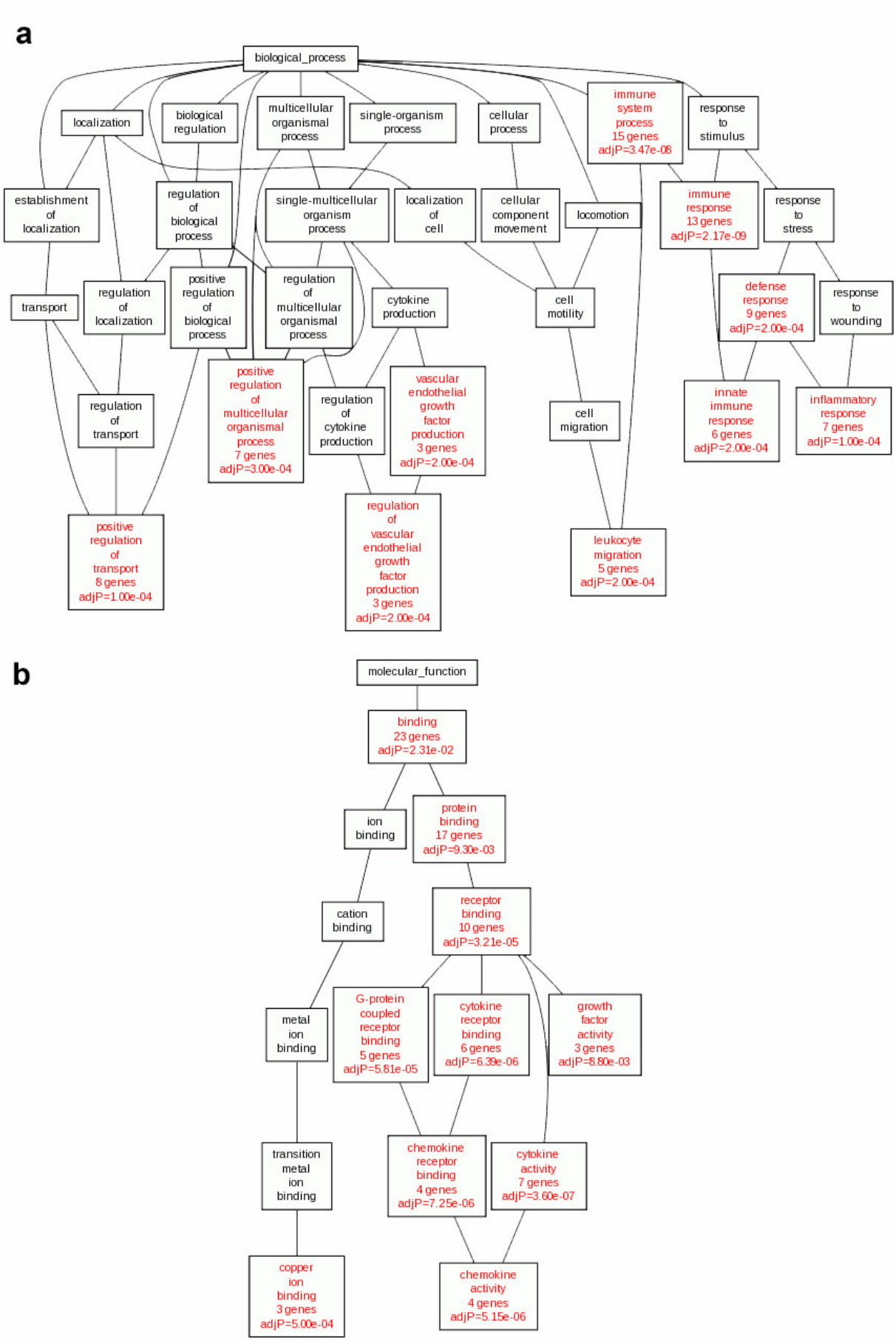


Figure S1. c-Myb expression inhibits lung metastasis and has no effect on liver metastasis. (a) Immunoblot analysis of c-Myb expression in 4T1 cells: wt- wild type, mock, *Myb*-overexpressing MM5, MM8B, and C11. **(b)** Quantification of liver foci (left panel) and representative H&E staining of liver sections 28 days after mf.p. injection of 4T1 cells (mock and MYB^{high} MM5, MM8B) (2 independent experiments). Scale bar =200μm (right panel). **(c)** *Myb*-overexpressing (MYB^{UP}) and mock-transfected 4T1-12B cells (expressing luciferase) were injected into m.f.p. Tumor formation and metastasis was monitored by bioluminescence imaging. 20 days p.i. tumors were removed, animals were then imaged at day 1 and 3 post

tumor removal. **(d)** Lung seeding of 4T1 cells, mock and MYB^{high} clone MM5, 24 hours after i.v. injection represented as number of 6-thioguanine-resistant colonies from digested lungs (n=3); representative images of cell colonies and quantification after 10 days. **(e)** Metastasis of MDA-MB-231 cells upon mammary fat pad injection into NSG mice and terminated at day 42, using MDA-MB-231 cells: transfected with empty vector (mock), overexpressing *MYB* (MYB^{high}), transfected with control gRNA (Scr); and deficient in *MYB* expression (MYB KO). Quantification of liver metastasis with representative tissue images. **(f)** Scheme showing the selection of lung3 cells from parental 4T1 wt cell line. **(g)** Lung seeding of 4T1 cells, wt and lung3 subline, representative images of 6-thioguanine-resistant colonies from digested lungs cell colonies after 10 days. **(h)** Tumor growth was monitored for 28 days p.i. of 4T1 wt and lung3 cells by measuring tumor length (l)/width (w) by caliper, tumor volume was calculated using a formula $\pi*l*w^2/2$. **(i)** Tumor weight was measured at termination. Data are median, $75^{th}/25^{th}$ percentile ± min/max. **(j)** Number of liver metastatic foci in wt and lung3-injected mice (2 independent experiments).


a
biological_process
localization
biological regulation
multicellular organismal process
single-organism process
cellular process
immune system process 15 genes adjP=3.47e-08
response to stimulus
establishment of localization
regulation of biological process
single-multicellular organism process
localization of cell
cellular component movement
locomotion
immune response 13 genes adjP=2.17e-09
response to stress
transport
regulation of localization
positive regulation of biological process
regulation of multicellular organismal process
cytokine production
cell motility
defense response 9 genes adjP=2.00e-04
response to wounding
regulation of transport
positive regulation of multicellular organismal process 7 genes adjP=3.00e-04
regulation of cytokine production
vascular endothelial growth factor production 3 genes adjP=2.00e-04
cell migration
innate immune response 6 genes adjP=2.00e-04
inflammatory response 7 genes adjP=1.00e-04
positive regulation of transport 8 genes adjP=1.00e-04
regulation of vascular endothelial growth factor production 3 genes adjP=2.00e-04
leukocyte migration 5 genes adjP=2.00e-04
b
molecular_function
binding 23 genes adjP=2.31e-02
ion binding
protein binding 17 genes adjP=9.30e-03
cation binding
receptor binding 10 genes adjP=3.21e-05
metal ion binding
G-protein coupled receptor binding 5 genes adjP=5.81e-05
cytokine receptor binding 6 genes adjP=6.39e-06
growth factor activity 3 genes adjP=8.80e-03
transition metal ion binding
chemokine receptor binding 4 genes adjP=7.25e-06
cytokine activity 7 genes adjP=3.60e-07
copper ion binding 3 genes adjP=5.00e-04
chemokine activity 4 genes adjP=5.15e-06

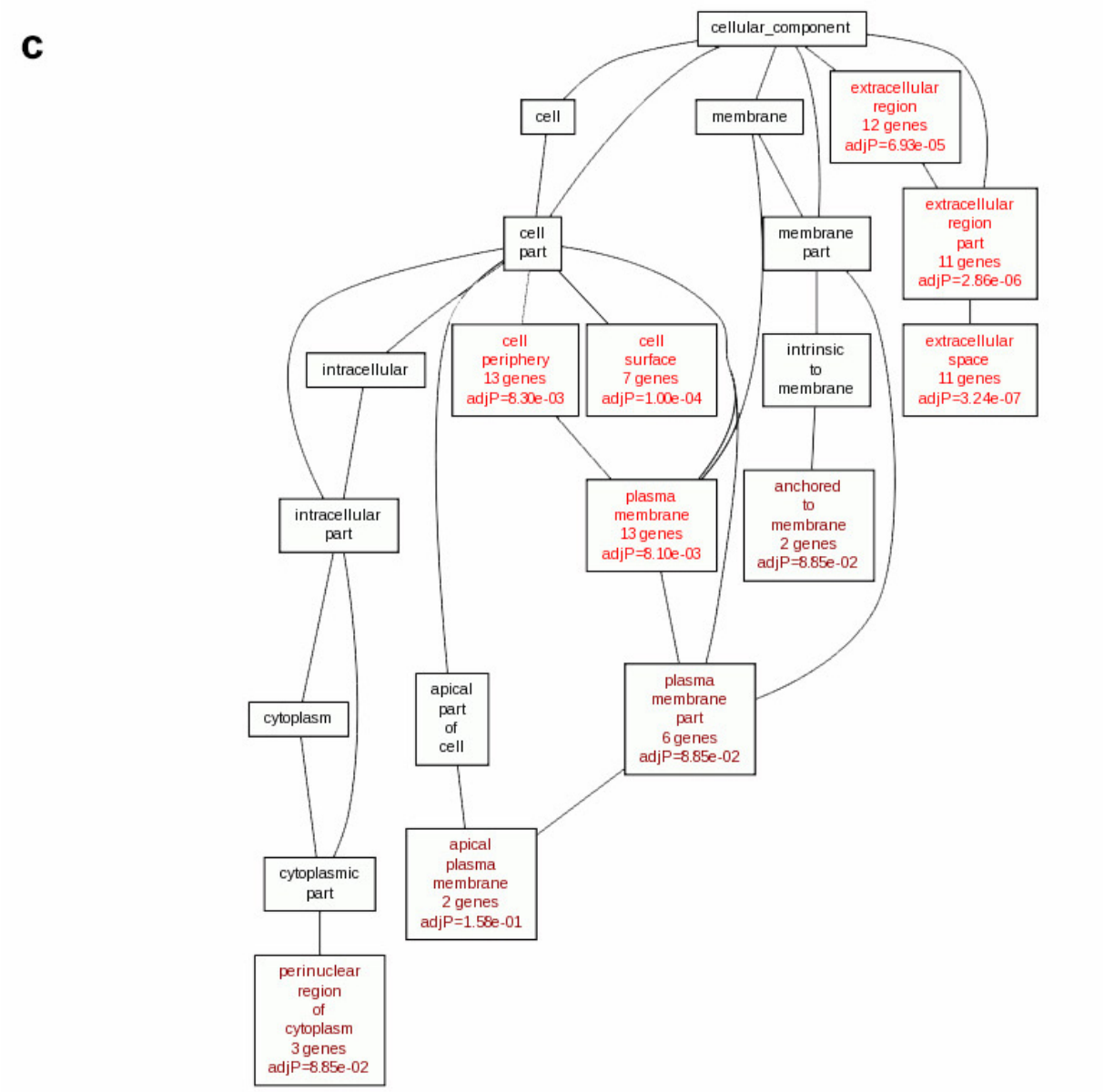


Figure S2. Graph-based visualization of GO enrichment analysis for 35 genes highly expressed in lung3 and suppressed in MYB^{high} cells. GO enrichment analysis was performed using WebGestalt (WEB-based GEne SeT AnaLysis Toolkit). Significantly enriched GO categories under Biological Process **(a)**, Molecular Function **(b)**, and Cellular Component **(c)** are shown with three separate Directed Acyclic Graphs (DAGs). Each GO category is a node in the DAG. The enriched GO categories in the top 10 that have a p value < 0.05 are colored red. GO categories in the top 10 that have a p value > 0.05 are colored brown, and the black ones are the parents of the top 10 categories.

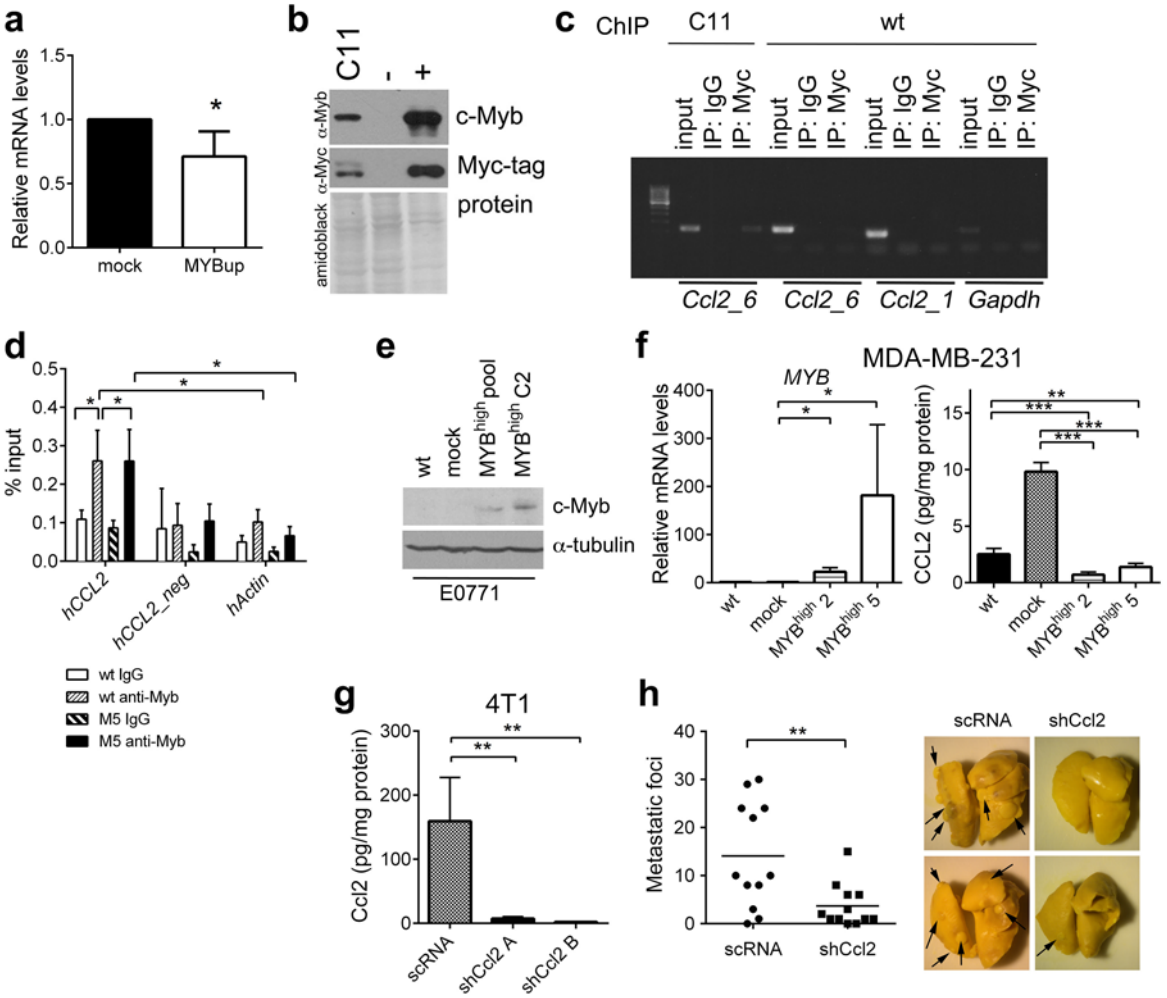


Figure S3. **c-Myb directly modulates Ccl2 expression and Ccl2 knock-down mimics the inhibitory effect of *Myb*-overexpression on lung metastasis. (a)** Ccl2 is down-regulated in 4T1 cells upon transient overexpression of *Myb*. mRNA levels were determined by qPCR 48 hours after transfection with expression plasmid harboring *Myb* cDNA (MYBup). Values were normalized to *Gapdh* and displayed as relative to mock-transfected cells (n=5). **(b)** Myc-tagged Myb expression in stable clone of 4T1 cells (C11) by imunoblot, + positive control, - negative control (right panel). **(c)** PCR was performed with immunoprecipitated chromatin fragments from C11 and wt cells. Two regions in the *Ccl2* promoter harboring the MBSs (*Ccl2*_1, 100 bp; *Ccl2*_6, 142 bp) were PCR amplified. Amplification of *Gapdh*_TSS (124 bp) was used as a specific control. Non-immunoprecipitated chromatin was used as input. **(d)** ChIP assays were performed with anti-Myb antibody or non-specific rabbit IgG in MDA-MB-231 cells transfected with human *MYB* (M5) and wt cells (n=3). Primers spanning two different MBSs (*hCCL2*) and region without MBSs (*hCCL2_neg*) were used for qPCR amplification and data are shown as percentage of input, unpaired t-test. *hActin* was used as a negative control. **(e)** Immunoblot analysis of c-Myb expression in E0771.LMB cells stably transfected with *Myb* (MYB^{high}). **(f)** *MYB*-overexpression in MDA-MB-231 breast cancer cells. mRNA levels (left panel) were determined by qPCR in two independent clones. Values were normalized to *GAPDH* and displayed as relative to wt cells. CCL2 levels (right panel) in the medium of respective *MYB*-overexpressing MDA-MB-231 clones (n=3). **(g)** Ccl2 protein levels in the conditioned media of 2 independent clones of 4T1 cells transduced with pLKO.1-puro harboring 2 different *Ccl2* specific shRNA (shCcl2 A and B) and non-targeting shRNA control (scRNA) were normalized to total protein content (n = 3, unpaired t-test). **(h)** Number of metastatic foci in lungs of tumor bearing mice 28 days after m.f.p. injection of

4T1 scRNA and shCcl2 B cells (2 independent experiments). Representative images of lungs fixed in Bouin's solution. *; $p<0.05$; **; $p<0.01$; ***; $p<0.001$.

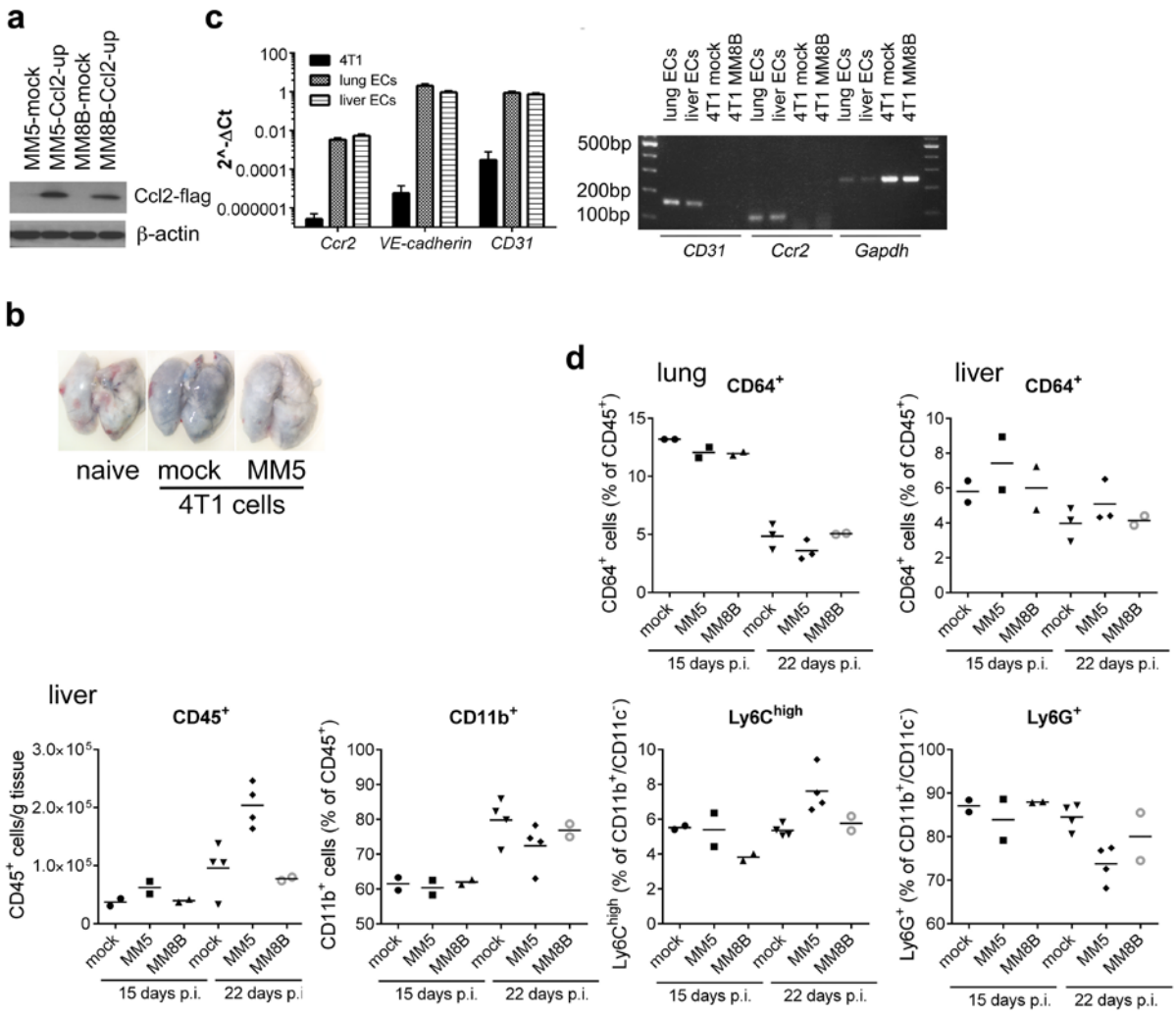


Figure S4. Endothelial *Ccr2* mediated tissue specific effect. (a) Immunoblot detection of exogenous Ccl2 protein using Ab specific for flag epitope 24 hours after transient transfection of *Myb*-overexpressing cells (MM5 and MM8B), left. **(b)** Macroscopic images of lungs from mice: untreated (naïve) or 24 hours p.i. with 4T1 mock and MM5 cells, treated with Evans blue. **(c)** Expression levels of *Ccr2*, *VE-cadherin*, and *CD31* mRNA 4T1 tumor cells and in primary ECs isolated from mouse lung and liver tissues using a positive immuno-magnetic selection (anti-CD31) were analyzed by qPCR, normalized to *Gapdh* and displayed as 2^-ΔCt (n=2) (left panel). *CD31*, *Ccr2*, and *Gapdh* mRNA in endothelial and tumor cells (4T1 mock and MM8B) as detected by RT-PCR (right panel). **(d)** Flow cytometry analysis of macrophage-like cells ($CD45^+CD64^+$), total leukocytes ($CD45^+$), myeloid cells ($CD45^+CD11b^+$), granulocytes ($CD45^+CD11b^+CD11c^-Ly6G^+$), and inflammatory monocytes ($CD45^+CD11b^+CD11c^-Ly6G^-Ly6C^{hi}$) recruited to lungs and liver of mice m.f.p.-injected with 4T1 mock or MYB^{high} (MM5 and MM8B) cells 15 and 22 days post injection (p.i.).

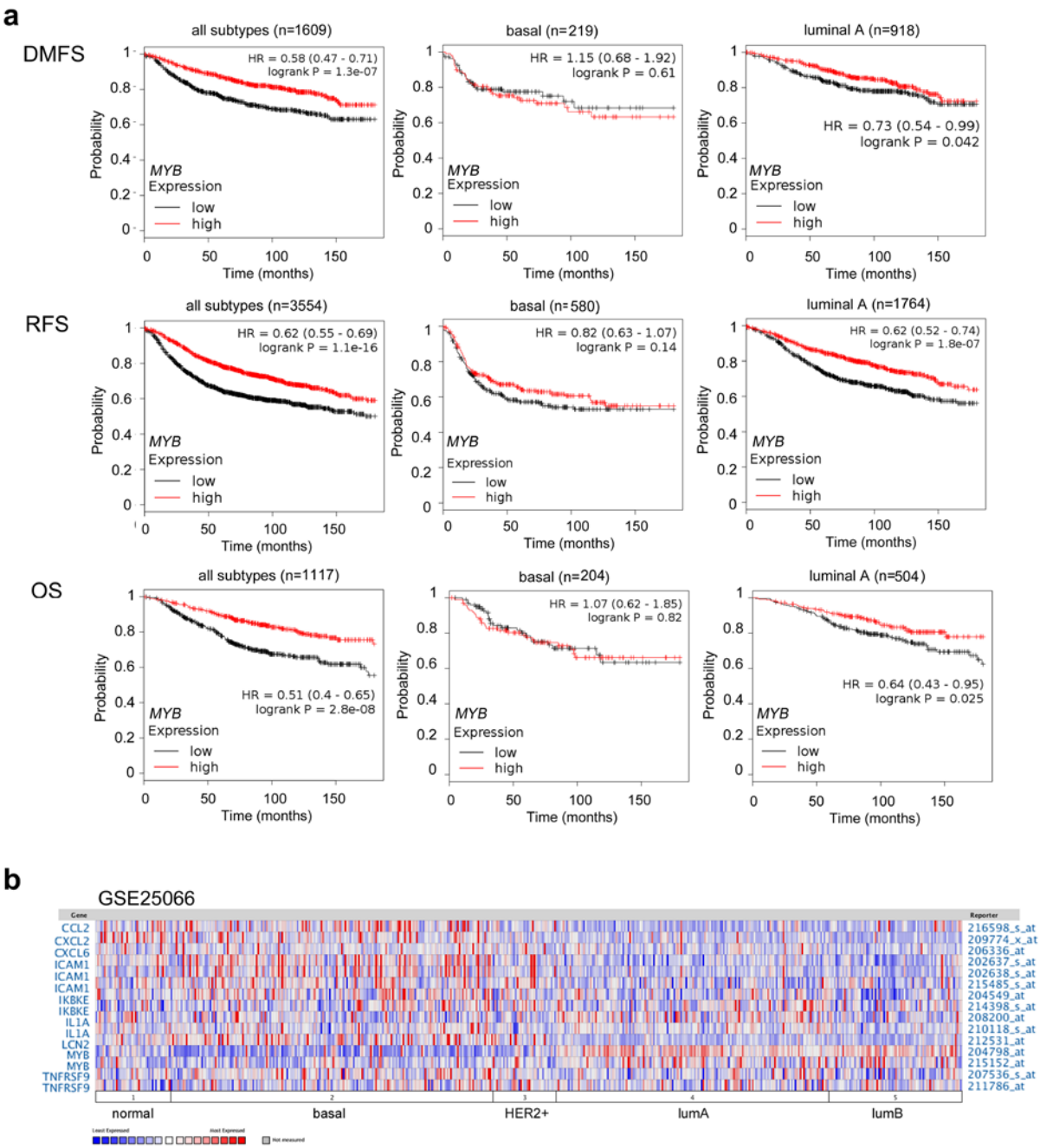


Figure S5. **Survival analysis. (a)** Meta-analyses of BCs patients available on KMplot.com representing the probability of distant metastasis free survival (DMFS), relapse-free survival (RFS) and overall survival (OS) in BCs stratified according to *MYB* expression. The log-rank test *P* value reflects the significance of the correlation between *MYB* high and longer survival outcome. Patients split by median, follow up threshold 15 years. KM plotter 2014 version (n=4142). **(b)** Heat map showing expression levels of *MYB* and identified signature genes in human BCs (GSE25066). Red and blue indicate high and low mRNA expression levels, respectively.

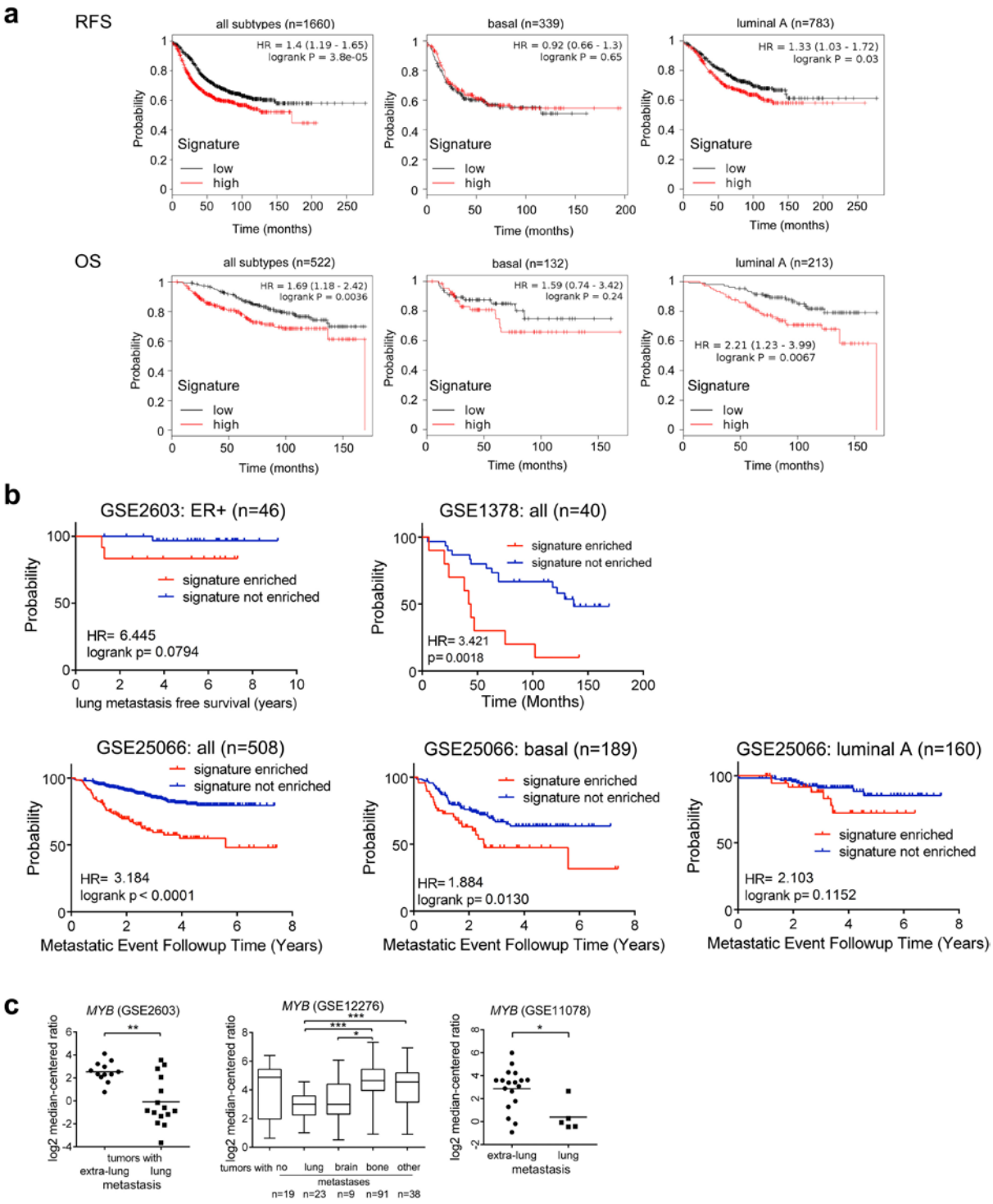


Figure S6. Kaplan-Meier survival plots. (**a)** Kaplan–Meier plots from meta-analyses (kmplot.com) (1) representing the probability of distant metastasis free survival (DMFS), relapse-free survival (RFS) and overall survival (OS) in BCs stratified according to the expression status of the signature genes (*CCL2*, *CXCL1*, *CXCL2*, *CXCL6*, *CXCL16*, *ICAM1*, *IL1A*, *TNFRSF9*, *LCN2*, *IKBKE)* and *MYB* (inversed). The log-rank test *P* value reflects the significance of the correlation between gene set high and shorter survival outcome. **(b)** Kaplan–Meier plots showing metastasis-free survival in additional cohorts of BC patients (GSE2603 ER+ only, GSE1378, GSE25066) based on signature enrichment as determined by z-score. Log-rank test analysis. **(c)** Expression of *MYB* mRNA as determined by microarray in BCs with site-specific metastases (GSE2603; GSE12276, patients with single metastases) and in metastatic deposits of BCs in lung and other organs (GSE11078).

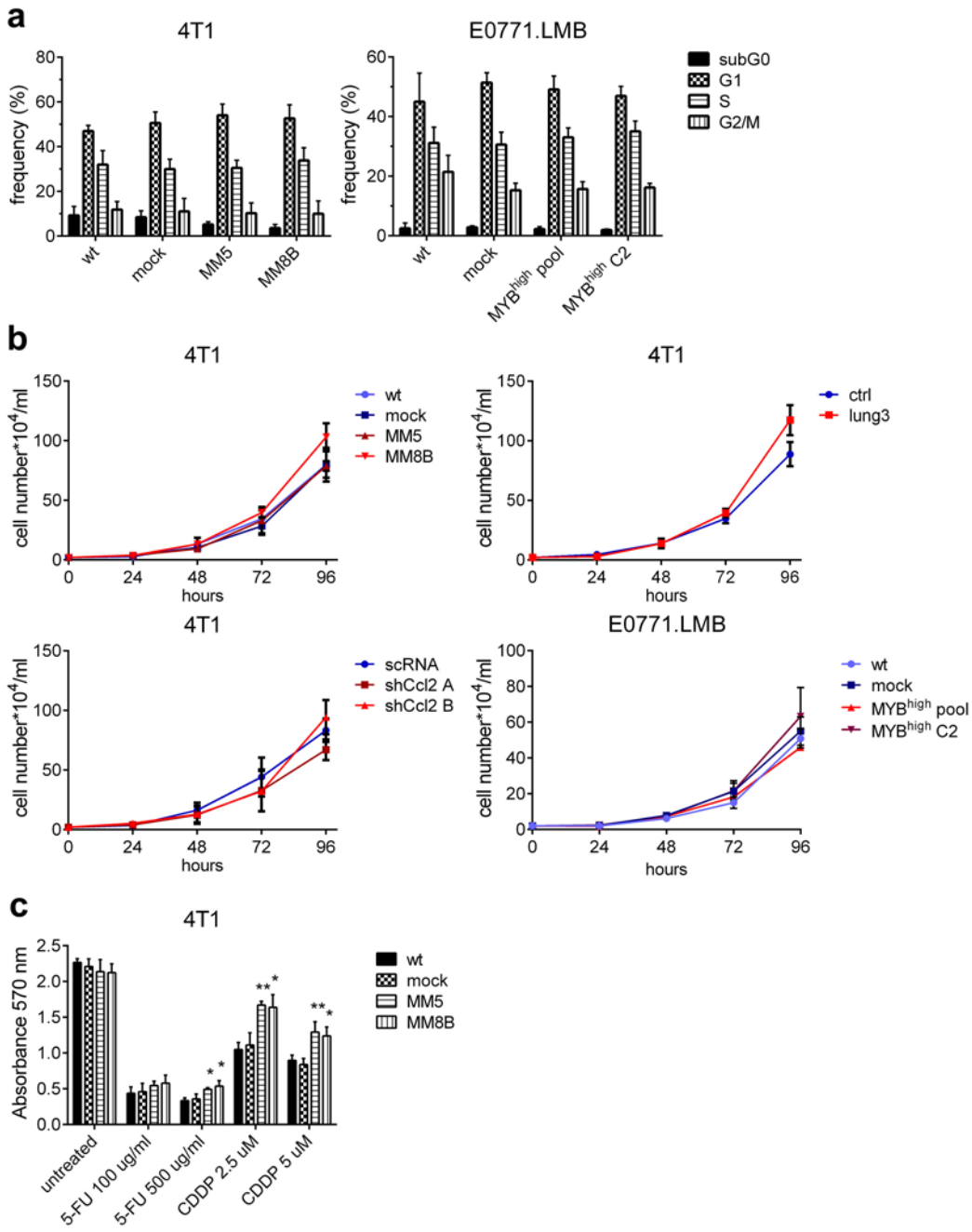


Figure S7. Characterization of *Myb*-overexpressing murine cell lines 4T1 and E0771.LMB. a) Cell cycle distribution in 4T1 and E0771.LMB MYB^{high} cells (n=3). **b)** Proliferation of 4T1 and E0771.LMB MYB^{high} cells, lung3 and Ccl2 knock-down sublines of 4T1 cells *in vitro*. Number of viable cells as determined by CASY cell counter (Roche) at the indicated time points (n=3). **c)** Chemosensitivity of 4T1 MYB^{high} cells. 4T1 cells (wt, mock, MYB^{high} MM5 and MM8B) were treated with 5-fluorouracil (5-FU) and cisplatin (CDDP) at the indicated concentrations for 48 hours and viable cells were quantified using crystal violet staining (n=3).

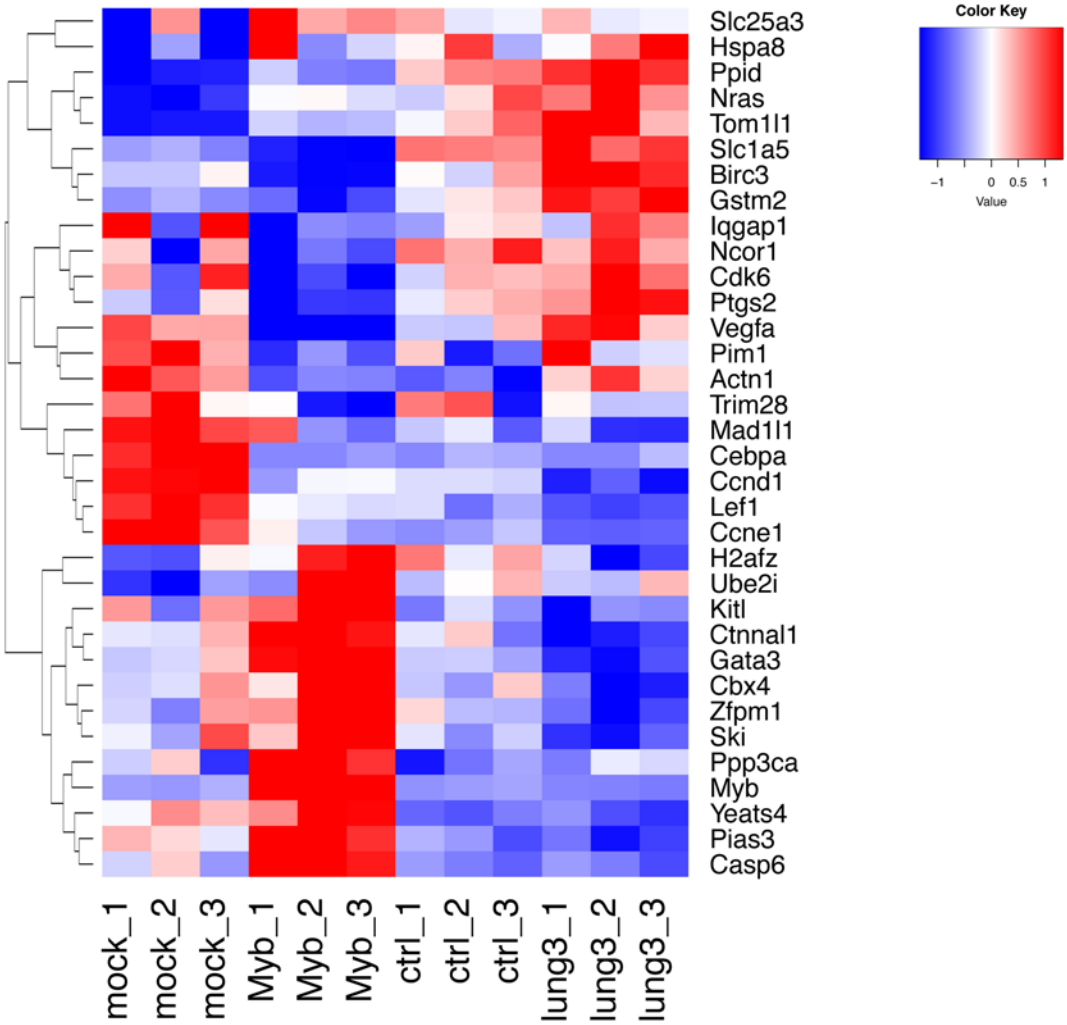


Figure S8. Testing Myb signature genes in 4T1MYB^{high}/lung3 cells. Heat map of Myb responsive genes (http://www.ncbi.nlm.nih.gov/biosystems/138073, and (2) differentially expressed ($p<0.05$) in 4T1MM5 cells compared to mock-transfected cells as determined by RNAseq.

Table S1: Primers used for qPCR:

	Primer sequences
Ccl2	Forward: 5' TTAACGCCCCACTCACCTGC 3'
	Reverse: 5' TTGGGTCAGCACAGACCTCTC 3'
Cxcl1	Forward: 5' ACCGAAGTCATAGCCACACTC 3'
	Reverse: 5' CTCCGTTACTTGGGGACACC 3'
Cxcl5	Forward: 5' CTGCGTTGTGTTTGCTTAAC 3'
	Reverse: 5' GACAGACCTCCTTCTGGTTT 3'
Cxcl16	Forward: 5' AACTCTGCAGGTTTGCAGCTC 3'
	Reverse: 5' GGGTCAGCAGGTCAAAGCTA 3'
Gapdh	Forward: 5' TCATGACCACAGTCCATGCC 3'
	Reverse: 5' ACTTGGCAGGTTTCTCCAGG 3'
Icam1	Forward: 5' CCCCGCAGGTCCAATTCACA 3'
	Reverse: 5' CCAAGCAGTCCGTCTCGTCC 3'
Ikbke	Forward: 5' GGCCCGAAACAAGAAATCCG 3'
	Reverse: 5' GCAGGACCTCAAACTCCCTC 3'
Il1a	Forward: 5' GCTTGAGTCGGCAAAGAAATCAA 3'
	Reverse: 5' ACGTTGCTGATACTGTCACCC 3'
Lcn2	Forward: 5' CCAGGCCCAGGACTCAACTC 3'
	Reverse: 5' CCTGCCCCGGAACTGATCG 3'
Myb	Forward: 5' CATTTGATGGGGTTTGGGCA 3'
	Reverse: 5' AGGATAGGGAACGTGACTGGA 3'
Tnfrsf9	Forward: 5' GCTGCCCTGAGATCGAAA 3'
	Reverse: 5' GAAAGTACCAGGCTGACAGTTA 3'

Primers used for ChIP:

	Primer sequences
Ccl2_1	Forward: 5' CACTTCCTGGAAACACCCGA 3'
	Reverse: 5' TGCTTGGTGCCAAGGAGTAG 3'
Ccl2_4	Forward: 5' AGCCAGACATTCCAGTTGGC 3'
	Reverse: 5' AGAAGCCTGTGTGTTAGCGT 3'
Ccl2_6	Forward: 5' AAACAGGGCAGAGAGCTACC 3'
	Reverse: 5' TCCAGCCTGGCTATCATCAC 3'
Gapdh_TSS	Forward: 5' TGAGCCTCCTCCAATTCAACC 3'
	Reverse: 5' CTGCAGCCTGGAAACCTGATA 3'
Gata3_TSS	Forward: 5' CCGAGGACATGGAGGTGACT 3'
	Reverse: 5' TCCATGTACGAATGGCCGAG 3'
hCCL2	Forward: 5' TTCCAGCAGTATGTCAGAGAGG 3'
	Reverse: 5' TGTAGTGCTGCAGAAGAAATGC 3'
hCCL2_neg	Forward: 5' CCCGGGGTAACTGAGGATTC 3'
	Reverse: 5' TAGGCTCTGGCACAAACCTG 3'
hActin	Forward: 5' AGTGTGGTCCTGCGACTTCTAAG 3'
	Reverse: 5' CCTGGGCTTGAGAGGTAGAGTGT 3'

Primer used for plasmid construction:

	Primer sequences
mCcl2	Forward: 5’ GAGCGAATTCCACCATGCAGGTCCCTGTCATGC 3’ Reverse: 5’ GAGCTCTAGAGTTCACTGTCACACTGGTCACTC 3’
Myb-stop	Forward: 5’ GAGCAAGCTTGCCATGGCCCGGAGAC 3’ Reverse: 5’ GAGCGGATCCCATGACCAGAGTTCGAGCTGAGA 3’
Myc-tag	Forward: 5’ GATCCGAACAAAAACTCATCTCAGAAGAGGATCTGTGATAGC 3’ Reverse: 5’ TCGAGCTATCACAGATCCTCTTCTGAGATGAGTTTTTGTTCG 3’

Oligont used for gRNA construction:

	Primer sequences
Myb	Forward: 5’ CACCGAAGTCTGGAAAGCGTCACTT 3’ Reverse: 5’ AAACAAGTGACGCTTTCCAGACTTC 3’
GFP	Forward: 5’ CACCGAACAGTCGCGTTTGCGACTG 3’ Reverse: 5’ AAACCAGTCGCAAACGCGACTGTTC 3’
seq	Forward: 5’ GGTTCAGACATGCAGGGGAA 3’ Reverse: 5’ AGGATGAGAACTCAACAGGCA 3’

Table S2: Total number of mapped reads per sample in RNA-seq.

RNA-seq Samples	Mapped Read Count Total
mock_1	25990118
mock_2	17385612
mock_3	18404018
Myb_1	15177941
Myb_2	16367893
Myb_3	22738254
ctrl_1	27277311
ctrl_2	30560072
ctrl_3	25331221
lung3_1	17051044
lung3_2	10959734
lung3_3	18106160

Table S3
Curtis Breast (EGAS00000000083)

	all subtypes (n=1985)			**basal (n=331)**			**luminal A (n=720)**			**luminal B (n=492)**			**HER2+ (n=240)**			**normal (n=202)**		
MYB / ILMN_1711894	r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)	
CCL2 / ILMN_1720048	**-0.4331**	< 0.0001	****	**-0.23**	< 0.0001	****	**-0.3094**	< 0.0001	****	**-0.2586**	< 0.0001	****	**-0.4767**	< 0.0001	****	**-0.3921**	< 0.0001	****
CXCL1 / ILMN_1787897	**-0.2709**	< 0.0001	****	-0.01935	0.7258	ns	-0.08715	0.0193	*	-0.05378	0.2337	ns	**-0.1829**	0.0045	**	-0.101	0.1525	ns
CXCL2 / ILMN_1682636	**-0.2673**	< 0.0001	****	-0.03169	0.5656	ns	**-0.1076**	0.0038	**	-0.06945	0.1239	ns	**-0.3614**	< 0.0001	****	**-0.2192**	0.0017	**
CXCL6 / ILMN_1779234	**-0.2045**	< 0.0001	****	0.1158	0.0352	*	-0.07693	0.039	*	**-0.1171**	0.0093	**	**-0.1438**	0.0259	*	**-0.1694**	0.0159	*
CXCL16 / ILMN_1672278	-0.06619	0.0032	**	-0.00468	0.9324	ns	**-0.1348**	0.0003	***	-0.04254	0.3464	ns	-0.1157	0.0737	ns	0.05108	0.4704	ns
IL1A / ILMN_1658483	**-0.1529**	< 0.0001	****	-0.05844	0.2891	ns	**-0.1274**	0.0006	***	**-0.1333**	0.0031	**	**-0.1731**	0.0072	**	-0.09537	0.177	ns
ICAM1 / ILMN_1812226	-0.09391	< 0.0001	****	**-0.1266**	0.0212	*	-0.05921	0.1124	ns	-0.03724	0.4098	ns	-0.07003	0.2799	ns	-0.1192	0.0912	ns
TNFRSF9 / ILMN_1813379	**-0.3785**	< 0.0001	****	**-0.1914**	0.0005	***	**-0.2333**	< 0.0001	****	**-0.1454**	0.0012	**	**-0.2017**	0.0017	**	**-0.2702**	0.0001	***
LCN2 / ILMN_1692223	**-0.3561**	< 0.0001	****	0.06508	0.2377	ns	**-0.1075**	0.0039	**	-0.05049	0.2637	ns	**-0.2426**	0.0002	***	-0.1038	0.1414	ns
IKBKE / ILMN_1755024	**-0.1733**	< 0.0001	****	-0.03553	0.5194	ns	0.059	0.1137	ns	0.05253	0.2458	ns	0.1883	0.0034	**	-0.08953	0.2051	ns

Table S3 cont.
Hatzis Breast (GSE25066)

	all subtypes (n=508)			**basal (n=189)**			**luminal A (n=160)**			**luminal B (n=78)**			**HER2+ (n=37)**			**normal (n=44)**		
MYB / 204798_at	r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)	
CCL2 / 216598_s_at	**-0.3114**	< 0.0001	****	**-0.1518**	0.037	*	**-0.1891**	0.0166	*	-0.02539	0.8254	ns	0.2853	0.0869	ns	-0.2254	0.1412	ns
CXCL1 / 204470_at	**-0.2792**	< 0.0001	****	-0.1142	0.1176	ns	**-0.2027**	0.0102	*	**-0.2296**	0.0431	*	-0.1756	0.2986	ns	0.02119	0.8914	ns
CXCL2 / 209774_x_at	**-0.2313**	< 0.0001	****	-0.1349	0.0641	ns	0.01379	0.8626	ns	-0.06971	0.5442	ns	-0.09483	0.5766	ns	0.01896	0.9028	ns
CXCL6 / 206336_at	**-0.1427**	0.0013	**	0.03469	0.6355	ns	**-0.1704**	0.0312	*	**-0.319**	0.0044	**	**-0.3725**	0.0232	*	-0.008293	0.9574	ns
ICAM1 / 202637_s_at	**-0.429**	< 0.0001	****	**-0.2909**	< 0.0001	****	**-0.2806**	0.0003	***	-0.1219	0.2879	ns	-0.1899	0.2603	ns	-0.1799	0.2425	ns
ICAM1 / 202638_s_at	**-0.285**	< 0.0001	****	**-0.2382**	0.001	***	0.06277	0.4304	ns	0.1146	0.3176	ns	-0.02906	0.8644	ns	-0.02766	0.8585	ns
ICAM1 / 215485_s_at	**-0.2894**	< 0.0001	****	**-0.2085**	0.004	**	**-0.198**	0.0121	*	-0.001536	0.9894	ns	-0.02489	0.8837	ns	**-0.5034**	0.0005	***
IKBKE / 204549_at	**-0.3324**	< 0.0001	****	-0.115	0.1151	ns	**-0.4609**	< 0.0001	****	**-0.3982**	0.0003	***	-0.1549	0.3599	ns	**-0.2983**	0.0492	*
IKBKE / 214398_s_at	-0.09471	0.0328	*	-0.01735	0.8127	ns	**-0.2154**	0.0062	**	**-0.2245**	0.0482	*	-0.05331	0.754	ns	-0.128	0.4078	ns
IL1A / 208200_at	-0.06317	0.1551	ns	-0.114	0.1182	ns	**-0.1971**	0.0125	*	**-0.4369**	< 0.0001	****	0.07732	0.6492	ns	0.1574	0.3075	ns
IL1A / 210118_s_at	**-0.1641**	0.0002	***	-0.1095	0.1338	ns	**-0.2509**	0.0014	**	**-0.3479**	0.0018	**	-0.2675	0.1095	ns	-0.2294	0.1341	ns
LCN2 / 212531_at	**-0.326**	< 0.0001	****	-0.06943	0.3424	ns	**-0.2996**	0.0001	***	-0.02057	0.8581	ns	-0.1639	0.3323	ns	0.0961	0.5349	ns
TNFRSF9 / 207536_s_at	**-0.2588**	< 0.0001	****	-0.08788	0.2292	ns	**-0.337**	< 0.0001	****	**-0.2515**	0.0263	*	-0.2209	0.1889	ns	-0.2201	0.1512	ns

Table S3 cont.

Glück Breast (GSE22358)

	all subtypes (n=152)			**basal (n=45)**			**luminal A (n=46)**			**luminal B (n=25)**			**HER2+ (n=21)**			**normal (n=15)**		
MYB / 16972	r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)	
CCL2 / 31182	**-0.4579**	< 0.0001	****	**-0.3637**	0.014	*	**-0.4158**	0.0041	**	**-0.6952**	0.0001	***	-0.03541	0.8789	ns	0.3317	0.2271	ns
CCL2 / 37637	**-0.4963**	< 0.0001	****	**-0.409**	0.0053	**	**-0.4377**	0.0024	**	**-0.7037**	< 0.0001	****	-0.102	0.66	ns	0.3955	0.1445	ns
CXCL1 / 2192	**-0.3485**	< 0.0001	****	-0.00939	0.9512	ns	**-0.444**	0.002	**	**-0.478**	0.0157	*	0.1213	0.6004	ns	0.2873	0.2992	ns
CXCL1 / 24518	**-0.3403**	< 0.0001	****	-0.0104	0.9459	ns	**-0.3266**	0.0267	*	**-0.4171**	0.038	*	0.1192	0.6069	ns	0.3382	0.2176	ns
CXCL1 / 35599	**-0.3264**	< 0.0001	****	-0.00253	0.9868	ns	**-0.4161**	0.004	**	**-0.5028**	0.0104	*	0.06089	0.7932	ns	0.195	0.4861	ns
CXCL16 / 6008	**-0.3647**	< 0.0001	****	-0.09895	0.5178	ns	**-0.3269**	0.0266	*	**-0.681**	0.0002	***	-0.2068	0.3684	ns	0.03536	0.9004	ns
CXCL2 / 22039	**-0.2879**	0.0003	***	0.003119	0.9838	ns	-0.2142	0.1528	ns	**-0.4507**	0.0238	*	0.1494	0.5181	ns	-0.1414	0.6152	ns
CXCL2 / 3722	**-0.3919**	< 0.0001	****	-0.1282	0.4014	ns	**-0.3427**	0.0197	*	**-0.4278**	0.0329	*	-0.1599	0.4886	ns	0.239	0.391	ns
CXCL6 / 25085	**-0.3128**	< 0.0001	****	-0.1817	0.2323	ns	**-0.3783**	0.0095	**	**-0.5142**	0.0086	**	-0.07318	0.7526	ns	0.1655	0.5555	ns
ICAM1 / 24842	**-0.4825**	< 0.0001	****	**-0.3899**	0.0081	**	**-0.2935**	0.0477	*	-0.1175	0.5759	ns	-0.3743	0.0946	ns	-0.2191	0.4328	ns
ICAM1 / 27689	**-0.5117**	< 0.0001	****	**-0.5134**	0.0003	***	**-0.4324**	0.0027	**	-0.3162	0.1236	ns	-0.349	0.121	ns	0.558	0.0307	*
ICAM1 / 7682	**-0.4438**	< 0.0001	****	**-0.4361**	0.0027	**	-0.1514	0.3151	ns	-0.1034	0.623	ns	-0.3724	0.0964	ns	0.2685	0.3333	ns
IKBKE / 18931	0.01187	0.8846	ns	-0.01317	0.9316	ns	0.43	0.0029	**	-0.227	0.2752	ns	0.1129	0.626	ns	0.1783	0.5248	ns
IL1A / 13203	**-0.1998**	0.0136	*	-0.1484	0.3306	ns	-0.14	0.3533	ns	-0.1381	0.5103	ns	-0.00203	0.993	ns	0.05968	0.8327	ns
IL1A / 17011	**-0.1784**	0.0279	*	-0.1743	0.2521	ns	-0.2701	0.0695	ns	-0.06358	0.7627	ns	0.07516	0.7461	ns	-0.03447	0.9029	ns
IL1A / 42653	-0.00251	0.9756	ns	-0.0927	0.5447	ns	0.05321	0.7285	ns	-0.1022	0.6269	ns	-0.07469	0.7476	ns	-0.02304	0.935	ns
LCN2 / 33737	**-0.1756**	0.0305	*	0.06941	0.6505	ns	-0.1928	0.1992	ns	0.1646	0.4318	ns	0.1757	0.4461	ns	0.1432	0.6107	ns
LCN2 / 7029	**-0.1934**	0.017	*	0.05617	0.714	ns	-0.06909	0.6482	ns	0.2578	0.2135	ns	-0.04479	0.8471	ns	0.2991	0.2788	ns
MYB / 24716	0.9938	< 0.0001	****	0.9942	< 0.0001	****	0.9905	< 0.0001	****	0.9882	< 0.0001	****	0.9883	< 0.0001	****	0.9789	< 0.0001	****
MYB / 39141	0.974	< 0.0001	****	0.9758	< 0.0001	****	0.9748	< 0.0001	****	0.9432	< 0.0001	****	0.9087	< 0.0001	****	0.9549	< 0.0001	****
TNFRSF9/4975	**-0.2707**	0.0008	***	-0.2342	0.1215	ns	**-0.3907**	0.0073	**	-0.1544	0.4612	ns	-0.2617	0.2519	ns	-0.3938	0.1636	ns

Table S3 cont.
Esserman Breast (GSE22226)

	all subtypes (n=129)			**basal (n=43)**			**luminal A (n=32)**			**luminal B (n=25)**			**HER2+ (n=21)**			**normal (n=8)**		
MYB / A_23_P31073	r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)		r	p (two-tailed)	
CCL2 / A_23_P89431	**-0.3138**	0.0003	***	-0.1994	0.1998	ns	-0.2046	0.2613	ns	**-0.431**	0.0315	*	-0.335	0.1377	ns	-0.04065	0.9239	ns
CXCL1 / A_23_P7144	**-0.3081**	0.0004	***	0.04465	0.7762	ns	0.02848	0.8771	ns	0.0977	0.6422	ns	-0.3862	0.0838	ns	-0.2994	0.4712	ns
CXCL16 / A_23_P38505	**-0.2381**	0.0068	**	-0.00799	0.96	ns	0.1228	0.503	ns	0.0447	0.832	ns	-0.2658	0.2441	ns	-0.0931	0.8264	ns
CXCL6 / A_23_P155755	**-0.2032**	0.0209	*	-0.2682	0.0821	ns	0.1427	0.4358	ns	-0.044	0.8357	ns	-0.2103	0.3603	ns	**-0.7098**	0.0486	*
ICAM1 / A_23_P153320	**-0.3414**	< 0.0001	****	-0.222	0.1524	ns	0.02051	0.9113	ns	-0.203	0.3304	ns	-0.4237	0.0556	ns	-0.5211	0.1854	ns
IKBKE / A_23_P887	-0.1476	0.0964	ns	-0.127	0.423	ns	0.1376	0.4526	ns	-0.073	0.7275	ns	-0.0206	0.9294	ns	**-0.8633**	0.0057	**
IL1A / A_23_P72096	-0.001919	0.9828	ns	-0.1041	0.5063	ns	0.137	0.4545	ns	0.1994	0.3392	ns	0.1805	0.4337	ns	-0.2579	0.5375	ns
LCN2 / A_23_P169437	**-0.3074**	0.0004	***	-0.1643	0.2924	ns	0.2021	0.2673	ns	-0.237	0.2549	ns	0.08325	0.7198	ns	**-0.768**	0.026	*
TNFRSF9 / A_23_P51936	-0.03086	0.7285	ns	0.1304	0.4045	ns	-0.03132	0.8649	ns	-0.312	0.1292	ns	0.2019	0.3802	ns	0.3258	0.431	ns

Table S3 cont.

Boss Breast (GSE12276)

	all subtypes (n=204)		
MYB / 204798_at	r	p (two-tailed)	
CCL2 / 216598_s_at	**-0.3441**	< 0.0001	****
CXCL1 / 204470_at	**-0.5058**	< 0.0001	****
CXCL16 / 223454_at	**-0.3616**	< 0.0001	****
CXCL2 / 209774_x_at	**-0.383**	< 0.0001	****
CXCL2 / 230101_at	0.07933	0.2594	ns
CXCL6 / 206336_at	**-0.252**	0.0003	***
ICAM1 / 202637_s_at	**-0.4202**	< 0.0001	****
ICAM1 / 202638_s_at	**-0.3988**	< 0.0001	****
ICAM1 / 215485_s_at	**-0.3747**	< 0.0001	****
IKBKE / 204549_at	-0.08999	0.2005	ns
IKBKE / 214398_s_at	**-0.1528**	0.0291	*
IL1A / 208200_at	0.09163	0.1924	ns
IL1A / 210118_s_at	-0.03529	0.6163	ns
LCN2 / 212531_at	**-0.4217**	< 0.0001	****
TNFRSF9 / 207536_s_at	**-0.1925**	0.0058	**
TNFRSF9 / 211786_at	**-0.1272**	0.0698	ns

Table S3 cont.

Kao Breast (GSE20685)

MYB / 204798_at	**all (n=327)**			**basal (n=43)**			**lum A (n=125)**		
	r	P (two-tailed)		r	P (two-tailed)		r	P (two-tailed)	
CCL2 / 216598_s_at	**-0.4116**	< 0.0001	****	0.08341	0.5949	ns	**-0.3871**	< 0.0001	****
CXCL1 / 204470_at	**-0.3483**	< 0.0001	****	0.1572	0.314	ns	-0.07594	0.4	ns
CXCL16 / 223454_at	**-0.4418**	< 0.0001	****	0.01759	0.9109	ns	**-0.2097**	0.0189	*
CXCL2 / 209774_x_at	**-0.2559**	< 0.0001	****	0.1153	0.4615	ns	-0.1004	0.2654	ns
CXCL2 / 230101_at	-0.01981	0.7212	ns	-0.08689	0.5796	ns	0.06825	0.4495	ns
CXCL6 / 206336_at	**-0.1687**	0.0022	**	0.2817	0.0672	ns	-0.002832	0.975	ns
ICAM1 / 202637_s_at	**-0.4574**	< 0.0001	****	-0.01102	0.9441	ns	**-0.1863**	0.0375	*
ICAM1 / 202638_s_at	**-0.3294**	< 0.0001	****	0.03342	0.8315	ns	-0.08209	0.3627	ns
ICAM1 / 215485_s_at	**-0.3456**	< 0.0001	****	-0.1026	0.5125	ns	-0.08075	0.3707	ns
IKBKE / 204549_at	**-0.2002**	0.0003	***	0.07222	0.6453	ns	0.0128	0.8874	ns
IKBKE / 214398_s_at	**-0.2003**	0.0003	***	0.02959	0.8506	ns	0.04673	0.6048	ns
IL1A / 208200_at	0.05595	0.3131	ns	0.1495	0.3386	ns	0.1212	0.1781	ns
IL1A / 210118_s_at	-0.07649	0.1676	ns	-0.07031	0.6542	ns	0.06359	0.4811	ns
LCN2 / 212531_at	**-0.3001**	< 0.0001	****	-0.07311	0.6413	ns	0.02289	0.8	ns
TNFRSF9 / 207536_s_at	**-0.3978**	< 0.0001	****	0.08965	0.5675	ns	**-0.2384**	0.0074	**
TNFRSF9 / 211786_at	**-0.1599**	0.0037	**	0.06945	0.6581	ns	-0.03279	0.7166	ns

Table S4

Category	Term	Count	%	PValue	Fold Enrichment	Bonferroni	Benjamini
GOTERM_BP_FAT	GO:0006955~immune response	12	34.285	9.738038E-10	11.937	5.268277E-7	5.268277E-7
GOTERM_MF_FAT	GO:0005125~cytokine activity	8	22.8571	1.520365E-08	24.607	1.50516E-6	1.50516E-6
GOTERM_MF_FAT	GO:0008009~chemokine activity	5	14.285	4.84164E-07	72.85	4.79311E-5	2.396584E-5
GOTERM_MF_FAT	GO:0042379~chemokine receptor binding	5	14.285	5.388794E-07	70.982	5.334765E-5	1.778287E-5
GOTERM_BP_FAT	GO:0006954~inflammatory response	7	20.0	5.356089E-06	14.577	0.002893	0.001448
GOTERM_CC_FAT	GO:0005615~extracellular space	8	22.857	0.000025	8.156	0.00129	0.00129
GOTERM_CC_FAT	GO:0044421~extracellular region part	9	25.7142	0.000044	6.058	0.002255	0.001128

GO terms enriched for 35 genes highly expressed in lung3 and suppressed in MYBup cells (fold change>1.5, $p<0.01$). GO enrichment (FDR < 0.05) for biological processes (BP), molecular function (MF) and cellular compartment (CC) was performed using DAVID. Table shows category, GO term, number of genes (count), %, *P* values for EASE score, fold enrichment, Bonferroni and Benjamini adjustment.

Table S5: GO enrichment analysis

Biological process	**GO ID**	**# reference genes in the category**	**# genes in the gene set and also in the category**	**expected # in the category**	**ratio of enrichment**	**p value from hypergeometric test**	**p value adjusted by the multiple test adjustment**
immune response	GO:0006955	C=680	O=13	E=0.97	R=13.42	rawP=3.91e-12	adjP=2.17e-09
immune system process	GO:0002376	C=1353	O=15	E=1.93	R=7.78	rawP=1.25e-10	adjP=3.47e-08
positive regulation of transport	GO:0051050	C=524	O=8	E=0.75	R=10.71	rawP=5.64e-07	adjP=0.0001
inflammatory response	GO:0006954	C=388	O=7	E=0.55	R=12.66	rawP=1.05e-06	adjP=0.0001
leukocyte migration	GO:0050900	C=165	O=5	E=0.24	R=21.26	rawP=3.52e-06	adjP=0.0002
defense response	GO:0006952	C=812	O=9	E=1.16	R=7.78	rawP=1.43e-06	adjP=0.0002
regulation of vascular endothelial growth factor production	GO:0010574	C=20	O=3	E=0.03	R=105.26	rawP=2.96e-06	adjP=0.0002
vascular endothelial growth factor production	GO:0010573	C=20	O=3	E=0.03	R=105.26	rawP=2.96e-06	adjP=0.0002
innate immune response	GO:0045087	C=299	O=6	E=0.43	R=14.08	rawP=3.69e-06	adjP=0.0002
positive regulation of multicellular organismal process	GO:0051240	C=520	O=7	E=0.74	R=9.45	rawP=7.24e-06	adjP=0.0003

Molecular function	**GO ID**	**# reference genes in the category**	**# genes in the gene set and also in the category**	**expected # in the category**	**ratio of enrichment**	**p value from hypergeometric test**	**p value adjusted by the multiple test adjustment**
cytokine activity	GO:0005125	C=186	O=7	E=0.25	R=27.49	rawP=5.14e-09	adjP=3.60e-07
chemokine activity	GO:0008009	C=35	O=4	E=0.05	R=83.47	rawP=1.47e-07	adjP=5.15e-06
cytokine receptor binding	GO:0005126	C=200	O=6	E=0.27	R=21.91	rawP=2.74e-07	adjP=6.39e-06
chemokine receptor binding	GO:0042379	C=45	O=4	E=0.06	R=64.92	rawP=4.14e-07	adjP=7.25e-06
receptor binding	GO:0005102	C=1173	O=10	E=1.61	R=6.23	rawP=2.29e-06	adjP=3.21e-05
G-protein coupled receptor binding	GO:0001664	C=185	O=5	E=0.25	R=19.74	rawP=4.98e-06	adjP=5.81e-05
copper ion binding	GO:0005507	C=51	O=3	E=0.07	R=42.96	rawP=4.64e-05	adjP=0.0005
growth factor activity	GO:0008083	C=144	O=3	E=0.20	R=15.22	rawP=0.0010	adjP=0.0088
protein binding	GO:0005515	C=6206	O=17	E=8.50	R=2.00	rawP=0.0012	adjP=0.0093
binding	GO:0005488	C=10881	O=23	E=14.90	R=1.54	rawP=0.0033	adjP=0.0231

Cellular component	GO ID	# reference genes in the category	# genes in the gene set and also in the category	expected # in the category	ratio of enrichment	p value from hypergeometric test	p value adjusted by the multiple test adjustment
extracellular space	GO:0005615	C=790	O=11	E=1.13	R=9.74	rawP=7.21e-09	adjP=3.24e-07
extracellular region part	GO:0044421	C=1046	O=11	E=1.50	R=7.36	rawP=1.27e-07	adjP=2.86e-06
extracellular region	GO:0005576	C=1831	O=12	E=2.62	R=4.58	rawP=4.62e-06	adjP=6.93e-05
cell surface	GO:0009986	C=563	O=7	E=0.80	R=8.70	rawP=1.24e-05	adjP=0.0001
plasma membrane	GO:0005886	C=3615	O=13	E=5.17	R=2.52	rawP=0.0009	adjP=0.0081
cell periphery	GO:0071944	C=3707	O=13	E=5.30	R=2.45	rawP=0.0011	adjP=0.0083
anchored to membrane	GO:0031225	C=136	O=2	E=0.19	R=10.29	rawP=0.0162	adjP=0.0885
plasma membrane part	GO:0044459	C=1440	O=6	E=2.06	R=2.91	rawP=0.0152	adjP=0.0885
perinuclear region of cytoplasm	GO:0048471	C=388	O=3	E=0.55	R=5.41	rawP=0.0177	adjP=0.0885
apical plasma membrane	GO:0016324	C=215	O=2	E=0.31	R=6.51	rawP=0.0378	adjP=0.1583

GO terms enriched for 35 genes highly expressed in lung3 and suppressed in MYBup cells (fold change>1.5, p<0.01). GO enrichment analysis was performed using WebGestalt (WEB-based GEne SeT AnaLysis Toolkit). Table shows significantly enriched GO categories under Biological Process, Molecular Function, and Cellular Component. GO categories in the top 10 that also have a p value < 0.05 are shown.
The table shows GO category, ID, the number of reference genes in the category (C), number of genes in the gene set and also in the category (O), expected number in the category (E), ratio of enrichment (R), p value from hypergeometric test (rawP), and p value adjusted by multiple test adjustment (Benjamini & Hochberg).

Supplementary methods

Plasmids and transfection

To generate expression plasmid pCMV.mCcl2 murine *Ccl2* was PCR amplified using primers listed in Table S1 from cDNA obtained by reverse transcription of mouse splenic mRNA. The amplicon (450 bp) was inserted into the p3xFLAG-CMV-14 vector backbone using *Eco*RI and *Xba*I. For generation of pcDNA3.Myc-tag-Myb plasmid, murine *Myb* was first re-amplified without stop codon from pcDNA3.mMyb (3) using primers in Table S1, inserted in pcDNA3.1(+) using *Hind*III/*Bam*HI and Myc-tag sequence with *Bam*HI/*Xho*I cohesive overhangs was subsequently inserted in frame at the 3'end of *Myb* cDNA. Stable clone expressing Myc-tagged Myb (C11) was generated by transfection with pcDNA3.Myc-tag-Myb using Lipofecamine® LTX (Life Technologies) and selection with G418 (400 μg/ml).

Cell cycle

Cells were washed with PBS and fixed in ice-cold 70% ethanol for 30 min at 4°C. Cells were centrifuged 200g/5min/4°C and stained with Vindel's solution (1 mM TrisCl, pH 8.0, 1 mM NaCl, 0.1% Triton X100, 10 mg/ml RNase, 50 mg/ml propidium iodide) and incubated for 30 min at 37°C. The DNA content was measured by the BD FACSVerse cytometer and data were analyzed with BD FACSuite software.

Proliferation

Cells were plated at the density $2x10^4$/ml and number of viable cells was determined by CASY cell counter (Roche) at the 24 hour interval for four days.

Cytotoxicity

Cells were treated with 2.5 and 5 μM cisplatin (cis-diamminedichloroplatinum(II), CDDP) (Sigma, St. Louis, MO) and 100 and 500 μg/ml 5-fluorouracil (2,4-Dihydroxy-5-fluoropyrimidine, 5-FU) (Sigma, St. Louis, MO) for 48 hours. Cell viability was estimated with the crystal violet staining. The adherent cells were washed with PBS, fixed with ice-cold methanol and stained with 0.4 % crystal violet for 20 min. Upon washing and air-drying, the dye was extracted with 10% acetic acid and absorbance at 570 nm was determined by the microplate reader (Synergy HT, BioTek, USA).

4T1-luc cells with overexpressed *Myb* gene

Puromycin-resistant luciferase expressing 4T1-12B cells (4) were transfected with pcDNA3.mMyb using Lipofectamine LTX (Life Technologies) and selected with 500 μg/ml of G418. Single clones were generated and *Myb* overexpression was verified by qPCR.

Intravital imaging

Mice injected with 4T1-luc cells were imaged using IVIS. Luciferin (1mg) was injected 10 min prior to imaging. Mice were under 2% Isofluran anesthesia, in a dorsal position placed in IVIS, and images taken.

Immunohistochemistry

The study group included 68 BC patients, 35 node-negative and 33 node-positive cases. According to the TNM staging 22 patients are of stage I (T1 N0 M0 = 22), 19 patients of stage II (T1 N1 M0 = 6; T2 N0 M0 = 10; T3 N0 M0 = 3), 25 patients of stage III (T2 N1 M0 = 10; T1 N2 M0 =3; T2 N2 M0 = 6; T2 N3 M0 = 2; T4 N1 M0 = 2; T3 N1 M0 = 1; T3 N2 M0 = 1), and 2 patients of stage IV (T1-2 N1 M1 = 2). Four μm thick tissue sections were

applied to positively charged slides, deparaffinized in xylene and rehydrated through a graded alcohol series. Endogenous peroxidase activity was quenched in 3% H_2O_2 for 5 minutes. The antigen retrieval was performed by heating the sections in citrate buffer (pH 6.0) in microwave (for 6 min at 120 W). UltraVision LP Large Volume Detection System HRP Polymer (Ready-To-Use) kit (Thermo Fisher Scientific, UK) was used according to the manufacturer´s instructions. 3,3´-diaminobenzidine (DAKO, Denmark) was used as chromogen. Slides were counterstained with Mayer´s hematoxylin. Negative controls were prepared by incubating samples in the absence of primary antibody. Evaluation of all IHC results was performed using a uniform microscope and camera setting (Zeiss microscope). The tumors were evaluated for percentage of immunostained positive cells in 10 random fields at magnification ×200.

References:

1. Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast Cancer Res Treat. 2010;123:725-31.
2. Nicolau M, Levine AJ, Carlsson G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. Proc Natl Acad Sci U S A. 2011;108:7265-70.
3. Knopfova L, Benes P, Pekarcikova L, Hermanova M, Masarik M, Pernicova Z, et al. c-Myb regulates matrix metalloproteinases 1/9, and cathepsin D: implications for matrix-dependent breast cancer cell invasion and metastasis. Mol Cancer. 2012;11:15.
4. Tao K, Fang M, Alroy J, Sahagian GG. Imagable 4T1 model for the study of late stage breast cancer. BMC Cancer. 2008;8:228.