

SUPPLEMENTAL METHODS

Generation of GFP+ A20 cells through GFP-expressing lentiviral vector

Lentiviral vector pRRLCMVGFPSin-18 (a kind gift from Dr. G. Ferrari from San Raffaele Hospital, Milan, Italy) has been used to transduce A20 cells. A third-generation packaging system was used to produce viral particles. Lentiviral stocks were produced in 293T cells by Ca3PO4 co-transfection of the plasmids as previously described (1). A20 cells were infected with lentiviral particles added at a multiplicity of infection (MOI) of 20. After 24 hours, cells were washed, and expanded. Three days later cells were checked for efficiency of transduction. To obtain a pure population of GFP-expressing cells, we sorted GFP+ A20 cells using a FACSAria sorter (BD).

Immunophenotypical analyses

For immunophenotypical analyses, human and mouse tissue sections were deparaffinized, rehydrated and unmasked using Novocastra Epitope Retrieval Solutions pH 6 and pH 9 in PT Link Dako at 98°C for 30 minutes. Subsequently, the sections were brought to room temperature and washed in PBS. After neutralization of the endogenous peroxidase with 3% H₂O₂ and Fc blocking by a specific protein block (Novocastra), the samples were incubated with antibodies. The following primary antibodies were adopted for IHC and IF: Ki67 (Abcam, 1:1000 pH6), c-Myc (clone Y69 Abcam, 1:500 pH6), p-γH2ax (Abcam, 1:1000 pH6), p-Chk1 (Abcam, 1:500 pH9), CD3 (Abcam, 1:100 pH9), CD8 (clone D4W2Z Cell Signaling, 1:400 pH9), αSMA (clone 1A4(asm-1 Biocare, 1:500 pH9), NGFR (clone 7F10 Leica Microsystems, 1:50 pH6), VCAM1 (clone EPR5047 Abcam, 1:500 pH9), CD146 (clone N1238 Abcam, 1:25 pH9), CD146 (clone EPR3208 Abcam, 1:25 pH9), PDGFR-β (clone Y92 Abcam, 1:250 pH6), Nestin (clone rat-40 1 Millipore, 1:100 pH9), DDX39 (clone EPR13507 B Abcam, 1:50 pH9), Lig1/Dna ligase1 (Invitrogen, 1:100 pH9), eIF6 (Cell Signaling, 1:50 pH9), TIAM-2 (clone EPR16838 Abcam, 1:500 pH9), PIK3R2/P85 Beta (LS BIO, 1:50 pH6), SIK2 (LS BIO, 1:100 pH6). Anti-mouse and anti-rabbit (Alexa Fluor 488 and 568 Conjugate) secondary antibodies were used for fluorescence detection.

IHC staining was revealed using Novolink Polymer Detection Systems (Novocastra) or IgG (H&L) specific secondary antibodies (Life Technologies, 1:500) or MACH 2 Double Stain 1 kit and DAB (3,3'-Diaminobenzidine, Novocastra) or Vulcan Fast Red or Ferangi Blue as substrate-chromogens.

Multi-resolution analysis

For multiresolution analysis of double-marker immunostained tissue sections, an efficient convolution of the redundant "à trous" algorithm with bilinear spline interpolation kernel has been realized; this has guaranteed an isotropic filter that therefore does not prefer particular directions (i.e. shapes). For the 20x and 40x magnifications, planes 5 and 6 of the dyadic analysis were chosen respectively, since they experimentally showed the best automatic segmentation with simple binarization on threshold equal to the mean value of the wavelet coefficients, increased by 1.5 times their standard deviation. This technique is described in literature for a rapid and correct segmentation, while attenuating eventual noise. Among the measures we considered (e.g. Pearson statistical index, Structural Similarity index, Peak Signal-to-Noise Ratio), the Euclidean distance between the centroids of the two classes (SMA/MYC and NGFR/MYC) was effective in underlining the three conditions of co-location, non-localization and substantially random distribution.

Flow cytometry

For flow cytometry analyses, the following antibodies were used: I H-2K^d (PE, clone SF1-1.1.1, eBioscience), I-A/I-E (APCeFluor780, clone M5/114.15.2, eBioscience), CD134 (OX-40) (PE, clone OX-

86, Biolegend), CD278 (ICOS) (BV510, clone C398.4A, Biolegend), CD8 (BV605, clone 53-6.7, BD), CD25 (BV650, clone PC61, BD), CD279 (PD-1) (BV786, clone J43, BD), Foxp3 (PerCpCy5.5, clone FJK-16.5, eBioscience), Ki-67 (APC, clone SolA15, eBioscience), CD4 (APC-Cy7, clone RM4-5, Biolegend), CD366 (TIM3) (BV421, clone 5D12/TIM-3, BD), CD45 (FITC, clone 30-F11, Tonbo Biosciences), CD31 (FITC, clone 390, eBioscience), TER119 (FITC, clone Ter119, Invitrogen), CD34 (FITC, clone RAM34, BD), cKIT (FITC, clone 2B8, eBioscience), SCA-1 (APC, clone D7, BD), CD44 (APC-Cy7, clone IM7, BD).

RNA sequencing and data analysis

The sample B-BM1 was removed from the analyses due to the low number of mapped reads (~6.7 million reads) and following the evaluation of the Euclidean distance among the samples. Differentially expressed genes were called using DESeq2 R package (2) starting from the count tables generated by HTSeq (3).

Quantitative gene set analysis was performed with the QuSAGE R package (4) using the log₂ normalized expression values and either the c2 curated gene sets of the Broad Molecular Signatures Database (MSigDB; <http://software.broadinstitute.org/gsea/msigdb/collections.jsp>) or a custom list of gene sets manually selected for being implicated in DNA damage, MYC targets or B-cell biology (Supplemental Table S11).

Heatmaps representing the unsupervised hierarchical clustering of the samples according to the log₂ normalized expression values of genes belonging to the gene set of interest were generated with the function `pheatmap` of the `pheatmap` (v 1.012) R package.

Circular layout visualization of the log₂FC of the activity of selected pathways were generated with a custom R script that made use of the `GOplot` (5) R library.

Gene fusion events were detected in RNA-seq data using the tool FusionCatcher (<https://github.com/ndaniel/fusioncatcher>), and *ad-hoc* R scripts were used to process the data in order to remove possible false positives according to FusionCatcher “annotation” field that takes into account several factors like known false positives, fusion genes seen in healthy sample or paralogy of genes involved in the fusion event (see FusionCatcher manual) (Supplemental Table S4). Circular layout visualization of the genomic location of the predicted gene fusions in WT and *Sparc*^{-/-} samples were generated with a custom R script based on the `circlize` R package.

Preranked gene set enrichment analysis (GSEA) for MSigDB Hallmark gene sets was carried out using the Bioconductor package `fgsea` package (<https://github.com/ctlab/fgsea>)

REFERENCES

1. De Palma M, Naldini L. Transduction of a gene expression cassette using advanced generation lentiviral vectors. *Methods Enzymol.* 2002;346:514-29. PubMed PMID: 11883088.
2. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. PubMed PMID: 25516281; PubMed Central PMCID: PMC4302049.
3. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015 Jan 15;31(2):166-9. doi: 10.1093/bioinformatics/btu638. Epub 2014 Sep 25. PubMed PMID: 25260700; PubMed Central PMCID: PMC4287950.
4. Yaari G, Bolen CR, Thakar J, Kleinstein SH. Quantitative set analysis for gene expression: a method to quantify gene set differential expression including gene-gene correlations. *Nucleic Acids Res.* 2013 Oct;41(18):e170. doi: 10.1093/nar/gkt660. Epub 2013 Aug 5. PubMed PMID: 23921631; PubMed Central PMCID: PMC3794608.

5. Walter W, Sánchez-Cabo F, Ricote M. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics*. 2015 Sep 1;31(17):2912-4. doi: 10.1093/bioinformatics/btv300. Epub 2015 May 11. PubMed PMID: 25964631.

SUPPLEMENTAL FIGURE LEGENDS

Supplemental Figure S1 Flow cytometry-based sorting of A20 cells from tissues.

(a) Representative flow cytometry sorting of A20 cells from BM, Spleen (highly and lowly infiltrated) and Liver of mice injected with A20 cells according to the experiment's graphical scheme shown in Figure 1A. Within the SSC-high CD45⁺ gate, A20 cells were characterized for high B220 and CD19 expression.

Supplemental Figure S2 Morphological and phenotypical heterogeneity in OPL239 and OPL241 DLBCL-like tumors.

(a) Representative microphotographs of H&E-stained sections showing the increasing degree of pleiomorphism and cytological atypia of DLBCL-like tumors from BM to LI and SPL tissues. Original magnification x400 and x630. (b-c) IHC for Ki-67 (b) and quantitative analysis (c) showing comparable proliferative fractions in BM, LI, and SPL tissue infiltrates. (d-e) IHC for Myc (d) and quantitative analysis (e) highlighting significant differences in Myc nuclear expression among BM, LI, and SPL infiltrates. (f-i) IHC for p- γ H2aX and p-Chk1 DNA damage response markers (f, h) and quantitative analysis (g, i) showing an increase fraction of immunoreactive nuclei in SPL infiltrates as compared with LI and, even more significantly, BM. Original magnifications x400.

Supplemental Figure S3 Intralesional heterogeneity of the mesenchymal components in DLBCL infiltrates.

(a) IHC for SMA, NGFR, CD146, and PDGFRb mesenchymal markers on three different regions (coloured insets) from two representative human DLBCL (see Supplemental Table 1) lymph node biopsies (panoramic insets) highlighting the disconnected staining pattern of the four markers in the three regions, which indicates the existence of different stromal adaptations to lymphoma infiltration within the same lesion. Regions were selected from whole slide digital scans. Original magnifications, x2 (panoramic insets) and x200 (coloured insets). (b) double marker IF for NGFR (green signal) and either SMA, PDGFRb, or VCAM-1 (all in red), on lymphoid follicles from a human lymph node with reactive follicular hyperplasia (upper panels) or infiltrates from a human DLBCL lymph node biopsy (lower panels). Mesenchymal markers that identify the FDC meshwork (NGFR and VCAM-1) or the inter-follicular fibroblastic reticular and sinusoidal network (SMA and PDGFRb) in the reactive condition (upper panels) are aberrantly focally associated or dissociated in DLBCL infiltrates (lower panels, arrows). Original magnification, x200.

Supplemental Figure S4 Mesenchymal changes in DLBCL-like tumors from OPL239/241 lines in nude mice at different tissue sites.

(a-i) IF for SMA (cyan signal), CD146 (violet signal), Ngfr (yellow signal), Pdgfrb (green signal), and Nestin (red signal) mesenchymal markers (a, d, g), quantitative analysis (histograms, b, e, h) and average relative fractions (pie charts, c, f, i) of uninvolved (CTRL), OPL239- or OPL241-infiltrated BM (a-c), LI (d-f), and SPL (g-i) of nude mice, detailing that different combinations of mesenchymal markers are variably induced across the three tissue sites. Original magnifications x200.

Supplemental Figure S5 Principal component analysis and clustering of the A20 transcriptomes from BM, LI, and SPL, and quantitative set analysis for gene expression.

(a) Principal component analysis of the samples. (b) Unsupervised hierarchical clustering of the samples according to their Euclidean distance calculated on the rlog-transformed data. (c-e) Bar plots displaying the 100 most modulated pathways between LI and BM (c), SP and BM (d) and SP and LI (e), as assessed by QuSAGE. Each bar shows the mean and the confidence interval of the gene set.

Supplemental Figure S6 GSEA for MSigDB Hallmark gene sets in A20 cells sorted from SPL and BM

(a) Preranked gene set enrichment analysis (GSEA) for MSigDB Hallmark gene sets on the RNA-seq transcriptomes of A20 cells sorted from SPL and BM showing the differential enrichment in MYC targets, DNA repair programs (positively enriched in the SPL), HYPOXIA, KRAS signaling and IFN γ response programs (positively enriched in the BM).

Supplemental Figure S7 In situ IHC validation of selected up- and down-modulated targets according to A20 SPL versus BM transcriptomes.

(a-f) IHC on BM and SPL A20 and OPL(239/241) infiltrates for selected markers listed among the top-50 significantly up- or down-modulated transcripts in the comparison between RNA-seq profiles of A20 cells sorted from the BM and SPL of WT mice (Supplemental Table S5). The targets selected according to their up-modulation in the SPL (a-c) included (a) Eif6, a key regulator of cell metabolism and Myc signaling enforcer, (b) the modulator of Rho family gtpases Tiam2, and (c) the DNA ligase I (Lig1) involved in DNA replication, recombination and repair. The targets selected according to their up-modulation in the BM (d-f) included (d) the salt-inducible kinase Sik2, (e) the Pi3k regulatory subunit Pik3r2, and (f) the ATP-dependent RNA helicase Ddx39. The in situ expression trend of the selected markers proved consistent with their transcriptional modulation in lymphomatous infiltrates in the two environments. Original magnifications x400.

Supplemental Figure S8 Transcriptional modulation of *Hif1 α* in A20 cells upon co-culture with different mesenchymal elements under hypoxic condition.

(a) *Hif1 α* gene expression by Real-time PCR on total RNA isolated from A20 cells at basal conditions and following 4 days co-culture with BM-MSCs, Spl-MSCs or BM-derived An2+ pericytes, under hypoxia (5%O₂). Under hypoxic conditions, *Hif1 α* was induced in A20 cells cultured alone and, less prominently, in the co-culture with Spl-MSC, while it was not induced by BM-MSC or BM-derived pericyte co-cultures, suggesting that A20 lymphoma cells could display a different hypoxic response according to the source of mesenchymal interactors.

Supplemental Figure S9 Multiresolution analysis on human DLBCL tumors

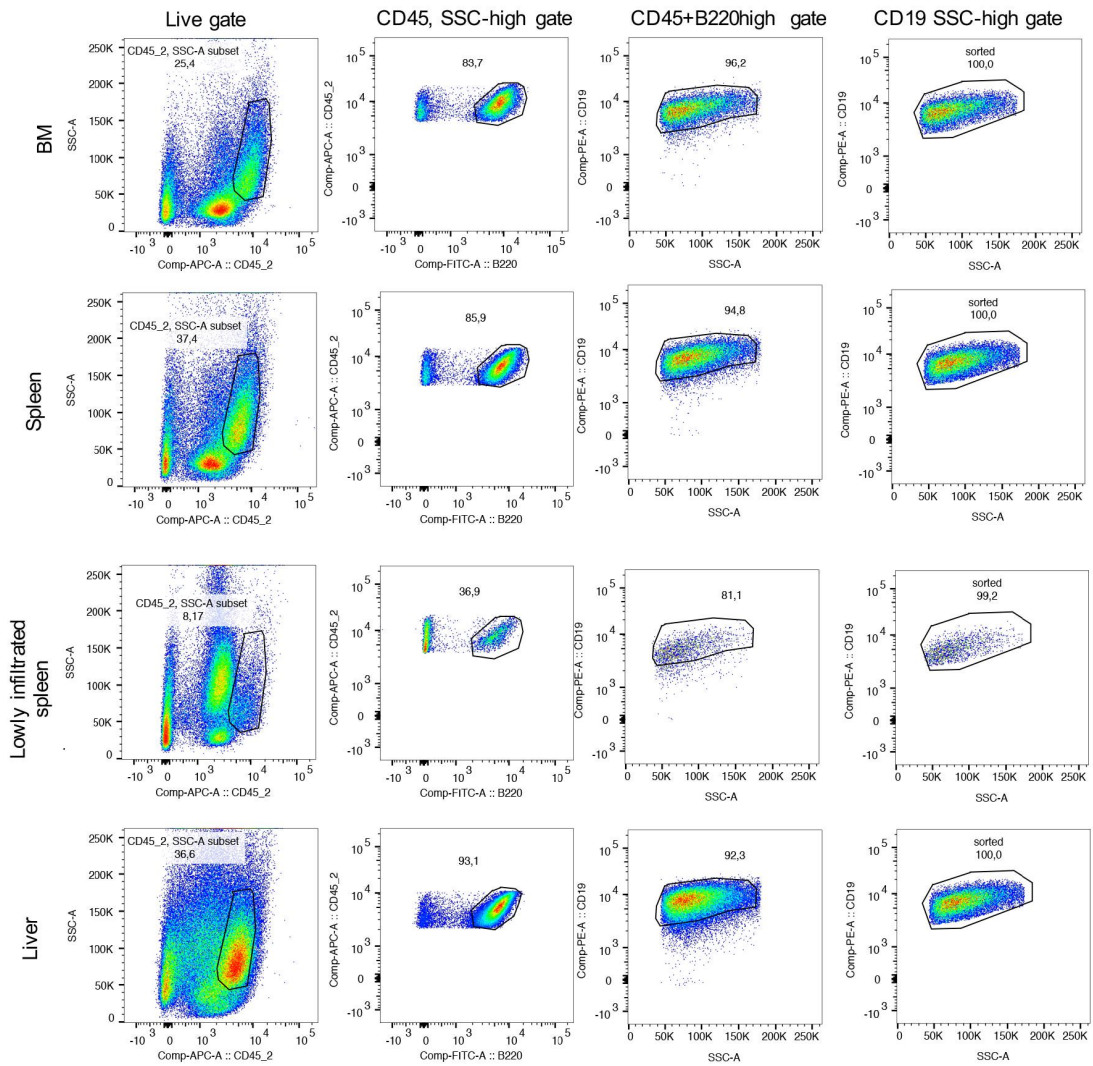
(a) Representative input (left panels) and output (right panels) images of Multiresolution Analysis (see Supplemental Methods Section) applied to double-marker IF stainings for SMA (green signal, upper panels) and MYC (red signal) or NGFR (green signal, lower panels) and MYC (red signal) on human DLBCL samples (Supplemental Table S1), showing the spatial segregation of NGFR- and MYC-stained areas (nearly complete absence of yellow signal in the output panel) and the intermingling SMA and MYC stainings (appearance of yellow signal in the output panel). Original magnification x200. (b) Graphic output of the Association index calculated according to the Multiresolution Analysis applied to SMA/MYC and NGFR/MYC immunostainings on human DLBCL samples (Supplemental Table S1) and expressed as the inverse of the average distance between the signals corresponding to the protein markers. The output shows a significantly lower degree of association between NGFR and MYC.

Supplemental Figure S10 Quantitative set analysis for gene expression in WT versus *Sparc*^{-/-} comparison and lymphoma cell frequency in the SPL and BM upon transplantation of A20-GFP+ cells.

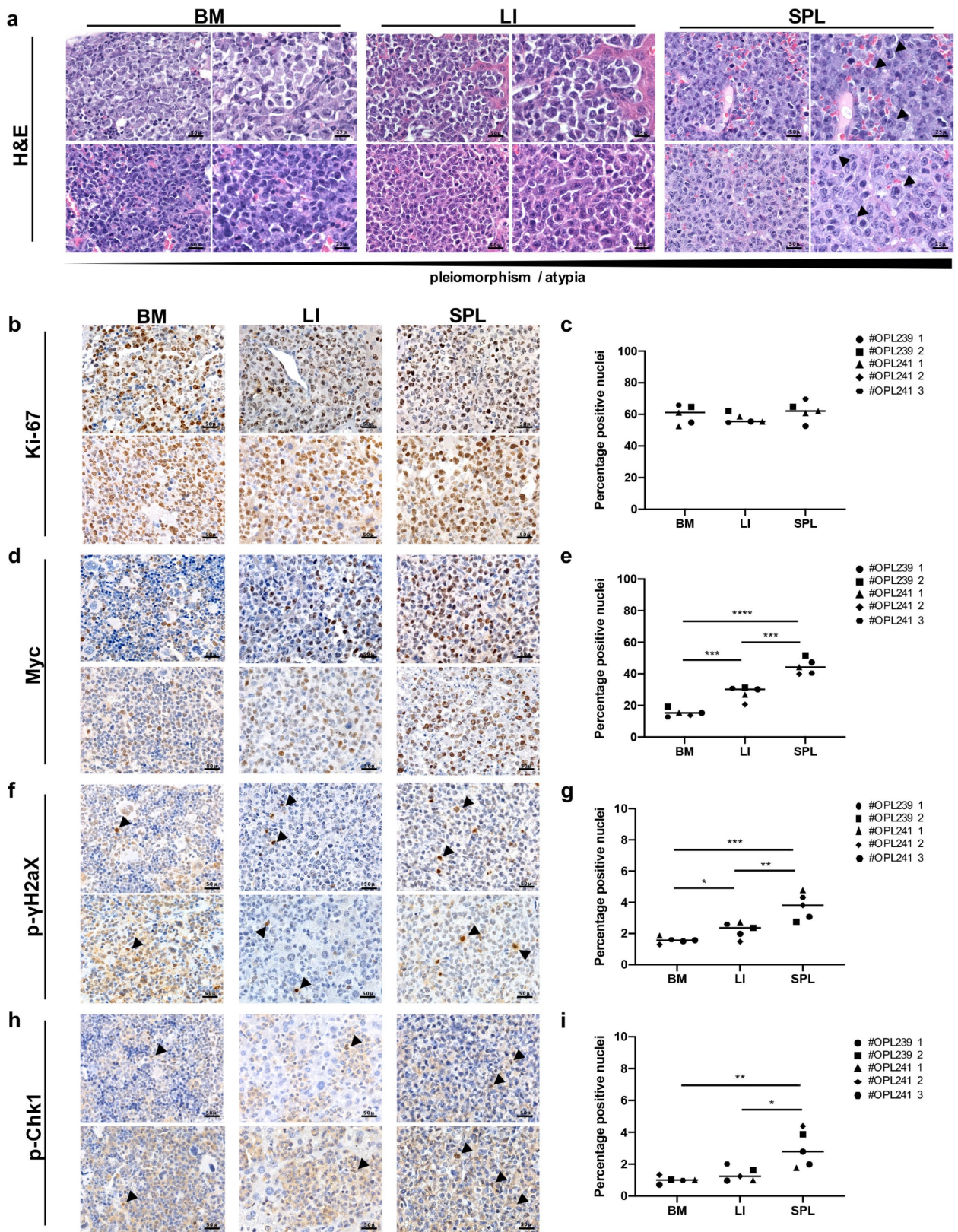
(a) Bar plots displaying the 100 most modulated pathways between WT and *Sparc*^{-/-} samples (including BM, LI and SPL tissue sites), as assessed by QuSAGE. Each bar shows the mean and the confidence interval of the gene set. (b) The frequency of A20-GFP+ cells infiltrating the SPL of WT and *Sparc*^{-/-} mice has been evaluated after 5 weeks of the injection of 5×10^5 A20-GFP+ cells from the BM, passaged three times in WT or *Sparc*^{-/-} hosts as detailed in the experiment's graphical scheme in Figure 6l. The SPL frequency of A20-GFP+ cells is extremely low and does not significantly vary according to the *Sparc* genotype of the priming or recipient hosts. (c) The frequency of A20-GFP+ cells infiltrating the BM has been evaluated after 5 weeks of the injection of 5×10^5 A20-GFP+ cells from the BM, passaged three times in WT hosts. The frequency of A20-GFP+ cells retrieved from the BM increases along passages suggesting a positive selection in BM homing capacity.

Supplemental Figure S1

a

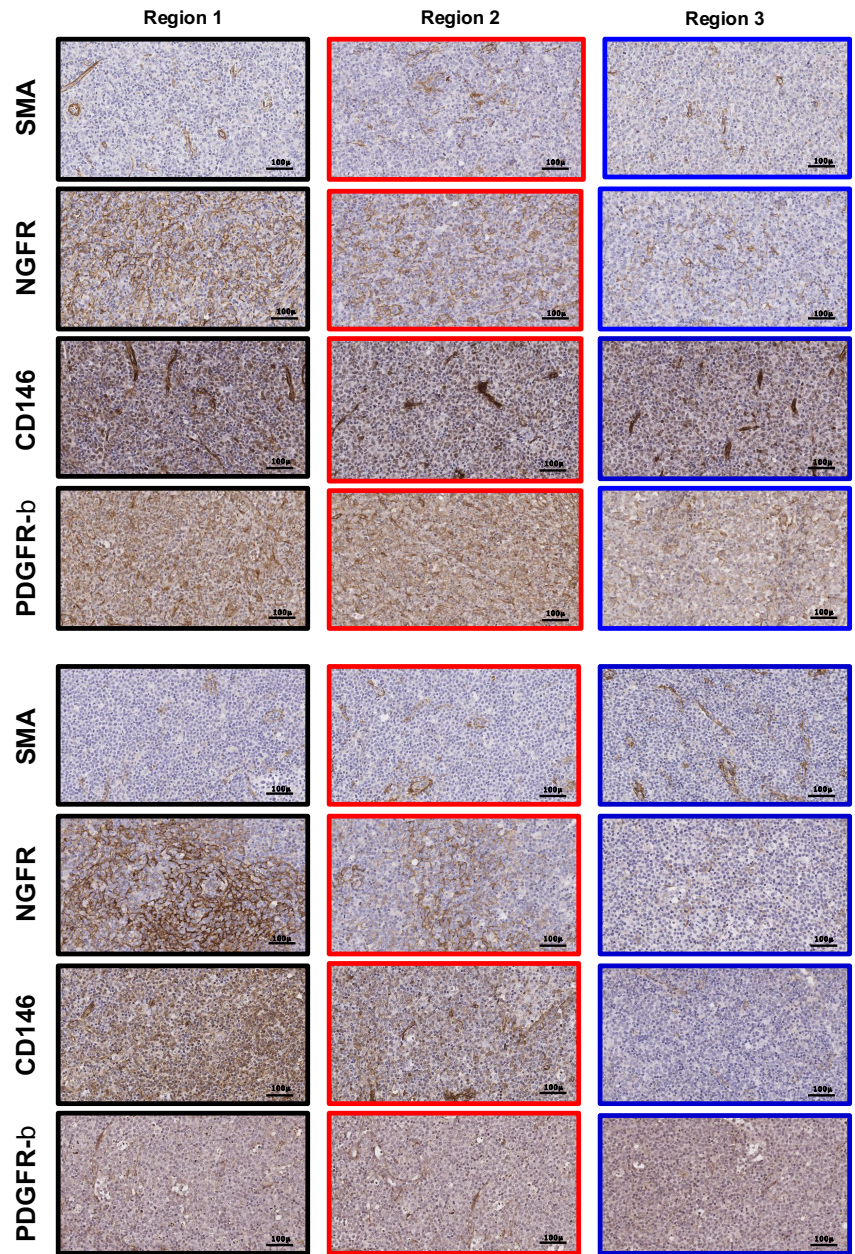
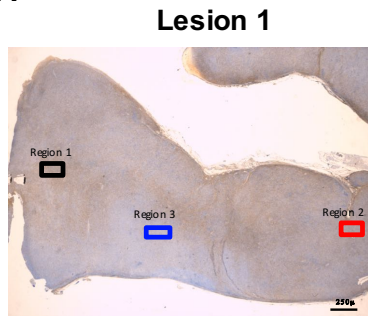


Supplemental Figure S2

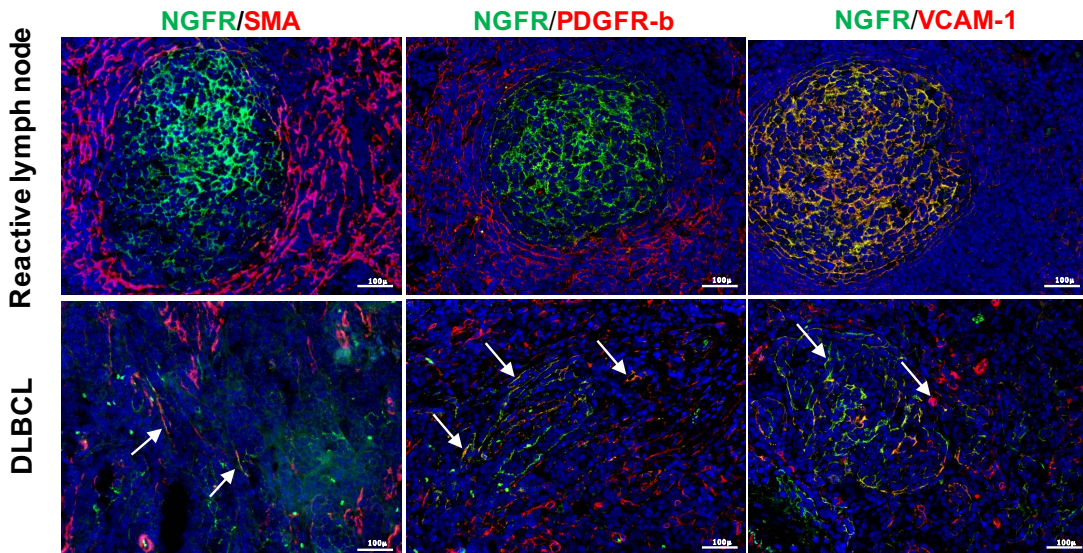


Supplemental Figure S3

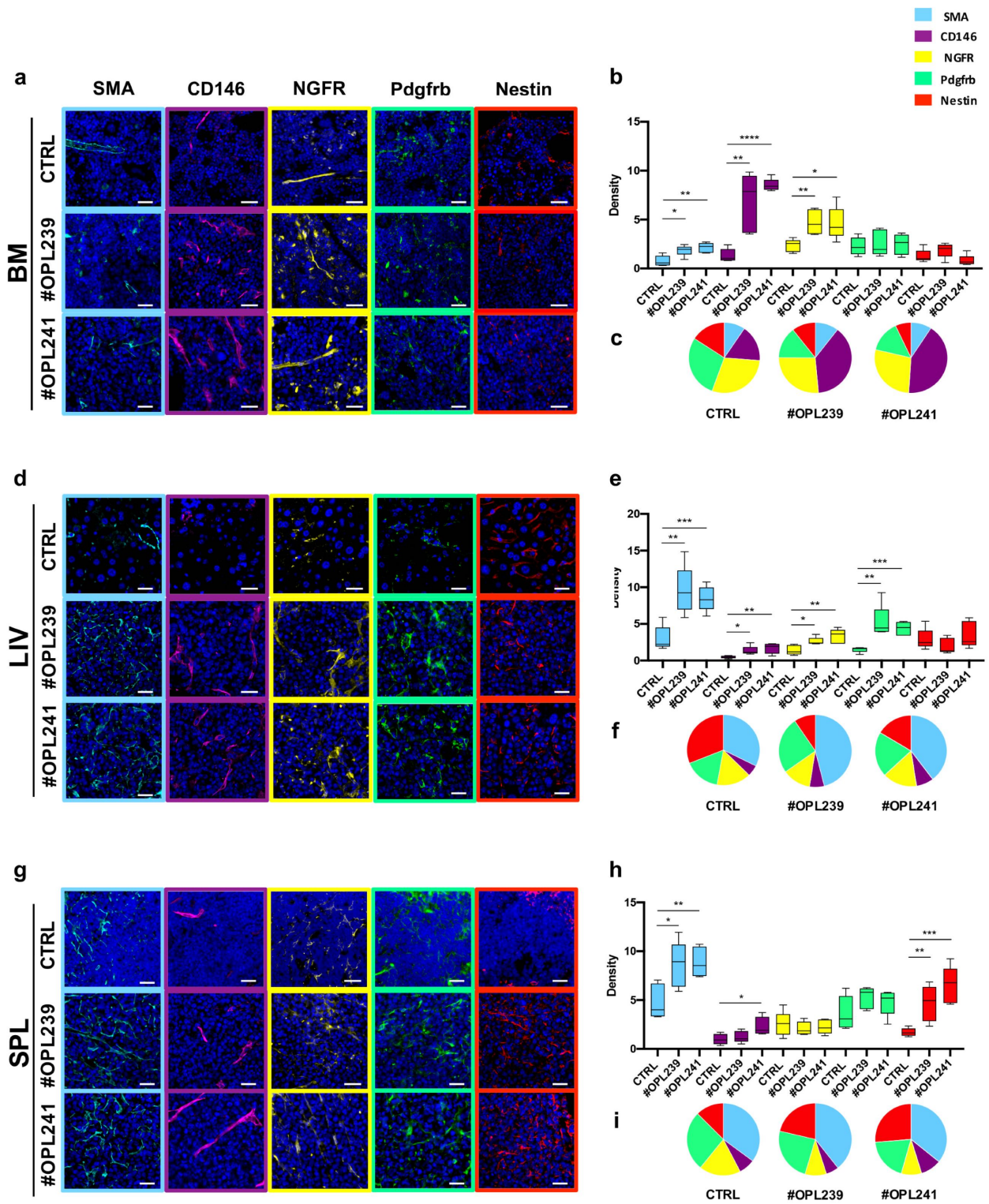
A



B

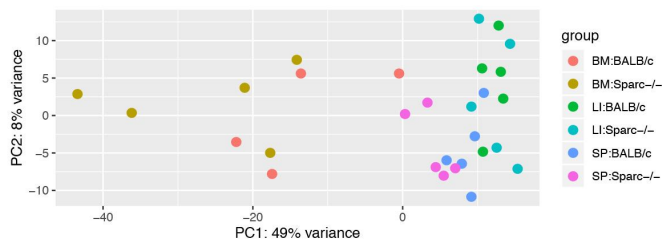


Supplemental Figure S4

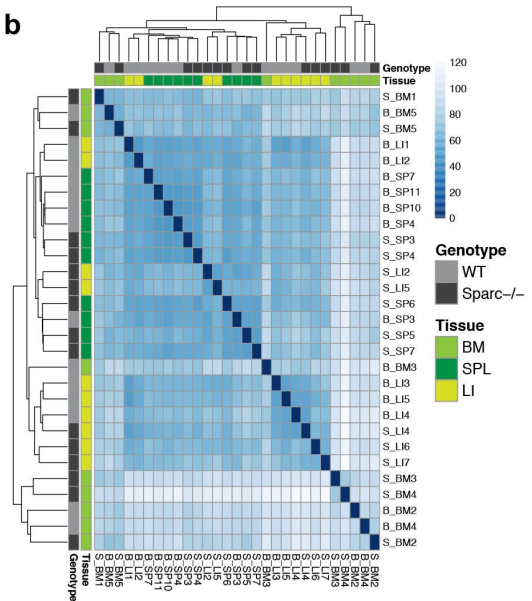


Supplemental Figure S5

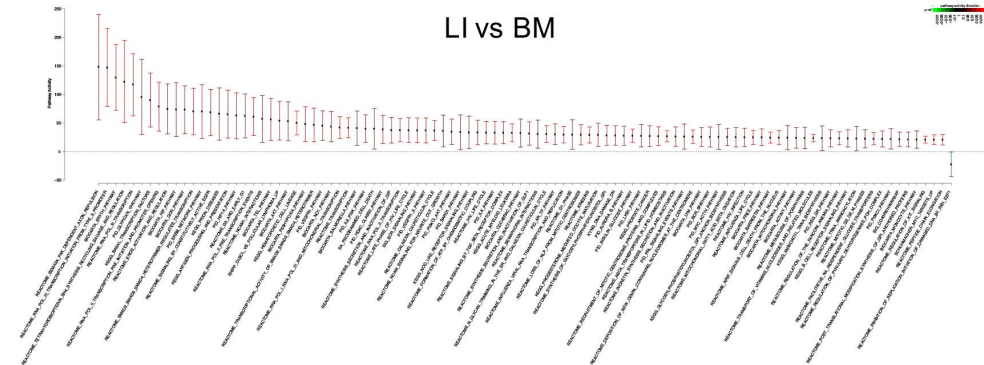
a



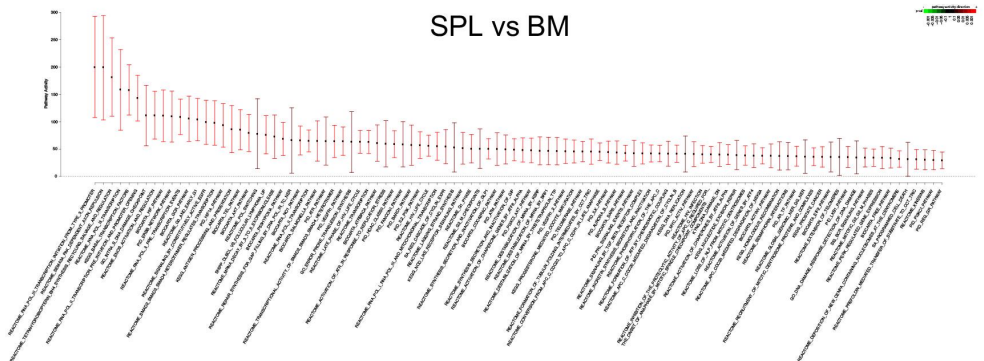
b



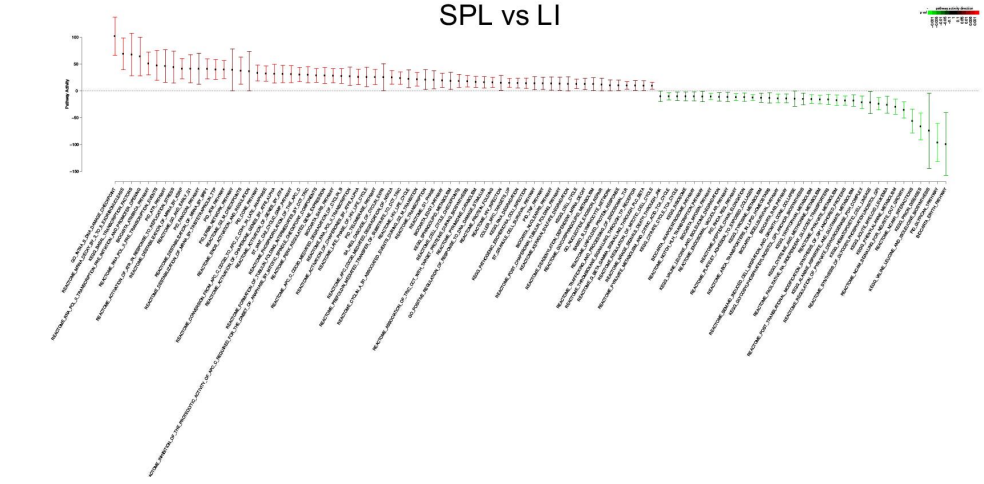
c



d

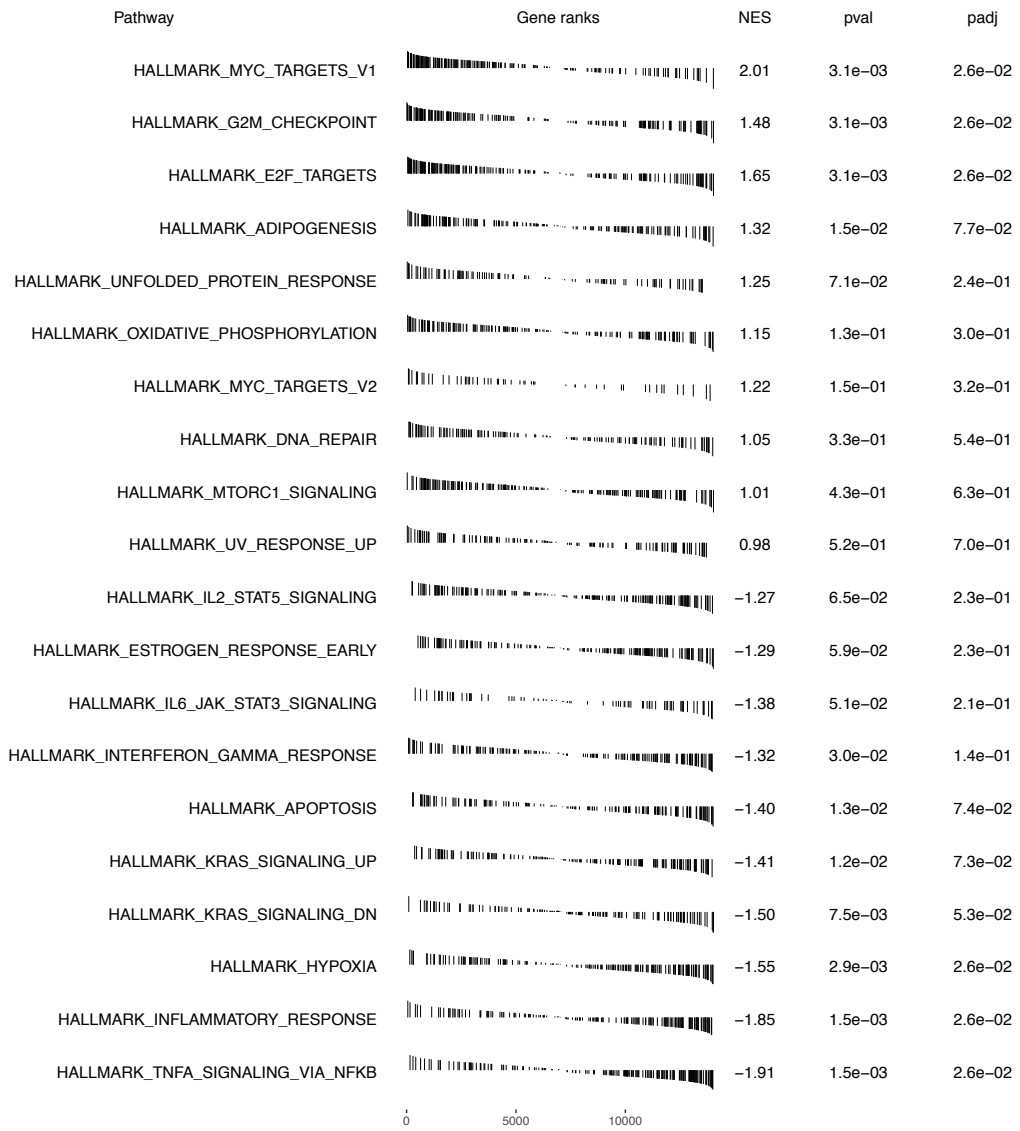


e

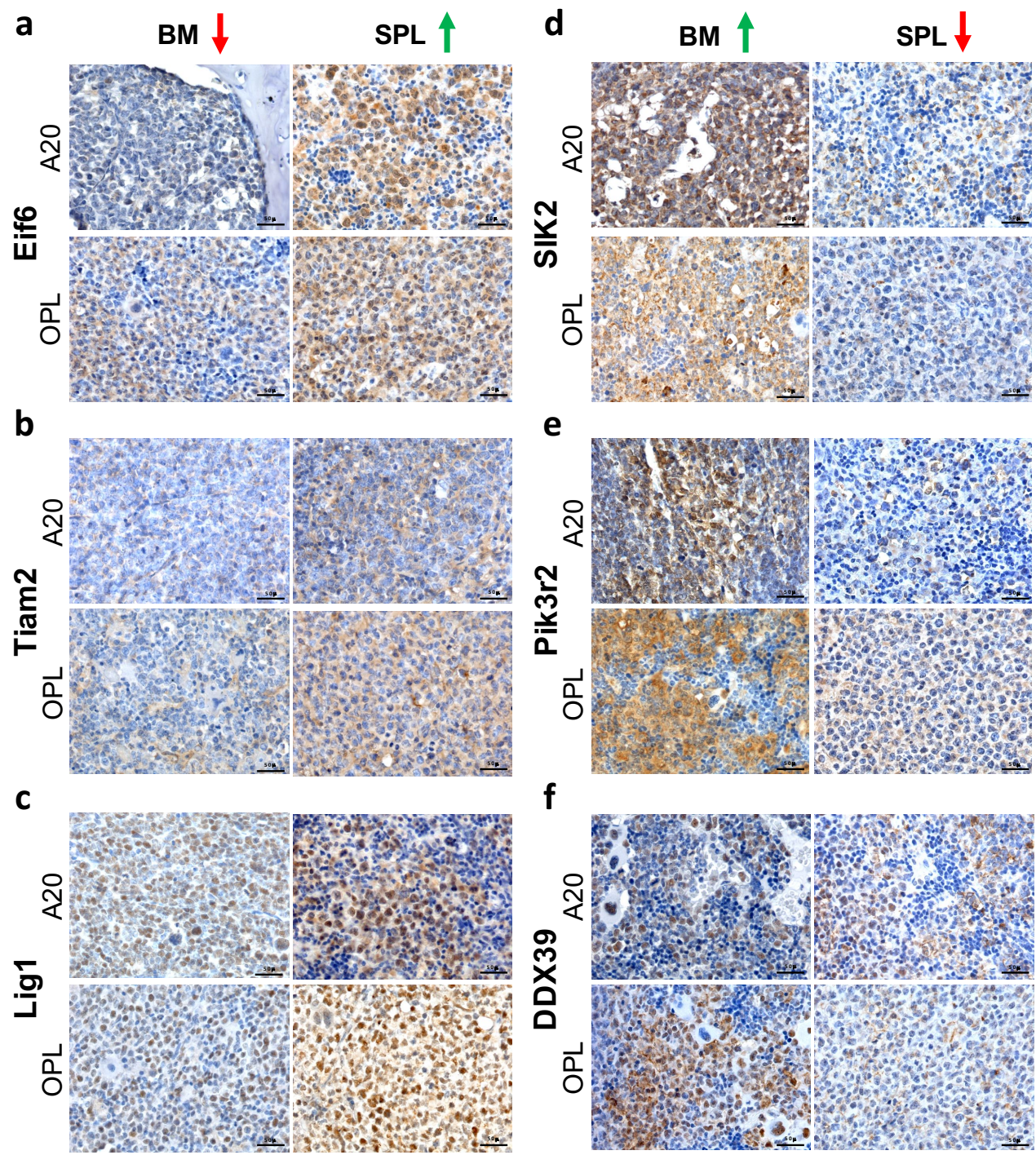


Supplemental Figure S6

a

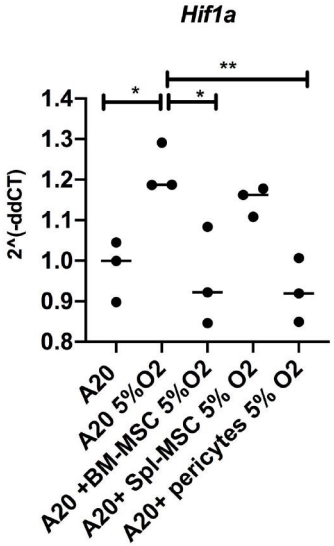


Supplemental Figure S7

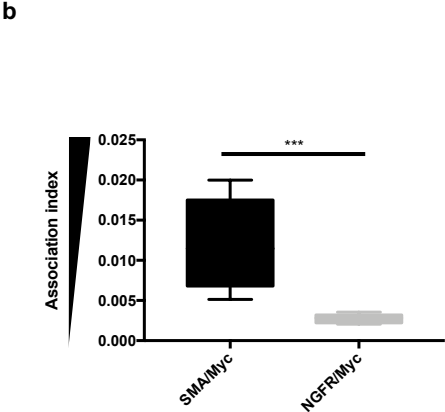
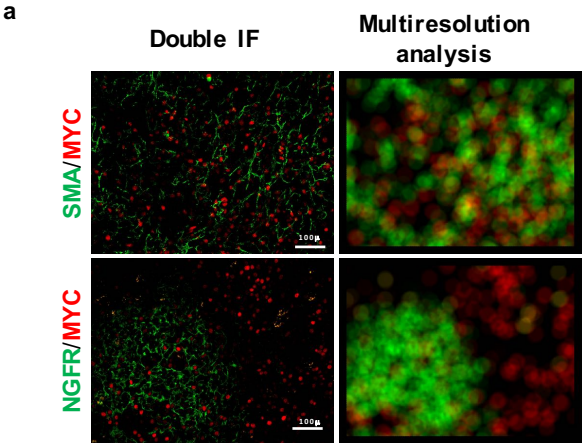


Supplemental Figure S8

a

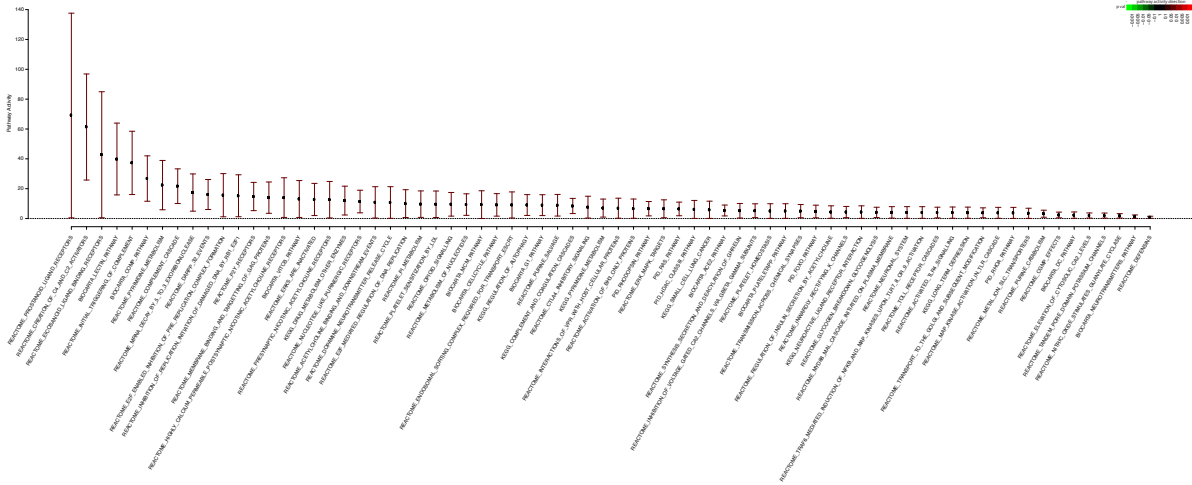


Supplemental Figure S9

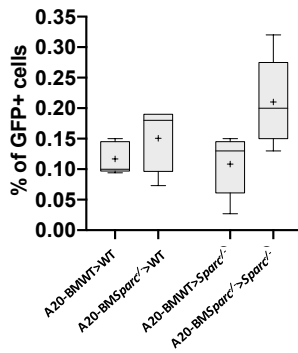


Supplemental Figure S10

a



b



c

A20 passage in the BM

