Characterization of GECPAR, a noncoding RNA that regulates the transcriptional program of diffuse large B-cell lymphoma

Sara Napoli,[1] Luciano Cascione,[1,2] Andrea Rinaldi,[1] Filippo Spriano,[1] Francesca Guidetti,[1] Fangwen Zhang,[1] Maria Teresa Cacciapuoti,[3] Afua Adjeiwaa Mensah,[1] Giulio Sartori,[1] Nicolas Munz,[1] Mattia Forcato,[4] Silvio Bicciato[4], Annalisa Chiappella,[5] Paola Ghione,[6] Olivier Elemento,[7,8] Leandro Cerchietti, [6]Giorgio Inghirami[3] and Francesco Bertoni[1,9]

[1]Institute of Oncology Research, Faculty of Biomedical Sciences, USI, Bellinzona, Switzerland; [2]SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland; [3]Pathology and Laboratory Medicine Department, Weill Cornell Medicine, New York, NY, USA; [4]Center for Genome Research, Department of Life Sciences University of Modena and Reggio, Modena, Italy; [5]Ematologia, A.O.U. Città della Salute e della Scienza di Torino, Turin, Italy; [6]Department of Medicine, Division of Hematology and Medical Oncology, Weill Cornell Medicine, New York, NY, USA; [7]Institute for Computational Biomedicine, Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA; [8]Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY, USA and [9]Oncology Institute of Southern Switzerland, Bellinzona, Switzerland.

 doi:10.3324/haematol.2020.267096

Received: July 29, 2020.

Accepted: June 16, 2021.

Pre-published: June 24, 2021.

Correspondence: *FRANCESCO BERTONI* - francesco.bertoni@ior.usi.ch

SARA NAPOLI - sara.napoli@ior.usi.ch

Characterization of GECPAR, a noncoding RNA that regulates the transcriptional program of diffuse large B cell lymphoma

Sara Napoli [1], Luciano Cascione [1,2], Andrea Rinaldi [1], Filippo Spriano[1], Francesca Guidetti[1], Fangwen Zhang[1], Maria Teresa Cacciapuoti [3], Afua Adjeiwaa Mensah [1], Giulio Sartori [1], Nicolas Munz[1], Mattia Forcato [4], Silvio Bicciato [4], Annalisa Chiappella [5], Paola Ghione [6], Olivier Elemento [7,8], Leandro Cerchietti [6], Giorgio Inghirami [3] and Francesco Bertoni [1,9]

[1] Institute of Oncology Research, Faculty of Biomedical Sciences, USI, Bellinzona, Switzerland; [2] SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland; [3] Pathology and Laboratory Medicine Department, Weill Cornell Medicine, New York, NY, USA; [4] Center for Genome Research, Department of Life Sciences University of Modena and Reggio; [5] Ematologia, A.O.U. Città della Salute e della Scienza di Torino, Turin, Italy; [6]Department of Medicine, Division of Hematology and Medical Oncology, Weill Cornell Medicine, New York, NY, 10021, USA; [7]Institute for Computational Biomedicine, Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA; [8]Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY, USA; [9] Oncology Institute of Southern Switzerland, Bellinzona, Switzerland.

SUPPLEMENTARY INFORMATION

SUPPLEMENTARY METHODS

Cell lines, small interfering RNA transfection and drug treatment

A total of 22 established human DLBCL cell lines were used: six ABC DLBCL (RIVA, HBL-1, U2932, SUDHL-2, OCI-LY-3, OCI-LY-10) and 16 GCB DLBCL (Pfeiffer, OCI-LY-1, OCILY-2, OCI-LY-7, OCI-LY-8, OCI-LY-18, OCI-LY-19, KARPAS422, SU-DHL-4, SU-DHL-6, SU-DHL-16, SUDHL-8, SUDHL-10, FARAGE, VAL, TOLEDO, DOHH2). Cell lines were grown as previously described (1, 2). Cell lines identity was validated by STR DNA fingerprinting using the Promega GenePrint 10 System kit (B9510) (2). PDTX-RN, PDTX-SS, PDTX-KD and PDTX-RRR are Patient Derived Tumor Xenograft Cell lines (PDTX-CL) spontaneously derived from DLBCL patient derived tumor xenograft (PDTX) models (NY-PDTX-RN, NY-PDTX-SS, NY-PDTX-KD and NY-PDTX-RRR PDTX) cultured in vitro. Established PDTX-CL were maintained in RPMI 20% FBS 1% penicillin and streptomycin and 0.2 % Normocin (Invivogen)The siGL3 Negative Control siRNA (3) and siRNA-461 or 563 were purchased from Thermo Fisher, scramble control, LNA 461, LNA489, LNA 563 and LNA 856 from Qiagen. Sequences are reported in supplementary table S9. Cells (1 million per sample) were transfected with siRNAs (200 pmol) or LNA (1 nmol) using 4D Nucleofector (Amaxa-Lonza), according to the manufacturer's instructions and incubated for 24h. Cells were treated with OTX-015 (birabresib) (Selleckchem, Houston, TX, USA), or DMSO (Sigma) for 4h. Cells were treated with AZ6102 (Selleckchem) or DMSO for 48h.

Human subjects

All patients providing samples gave written informed consent. Molecular and clinical data acquisition and analysis and PDTX establishment were approved and carried out in accordance with Declaration of Helsinki and were approved by Institutional Review Boards of the New York Presbyterian Hospital, Weill Cornell Medicine (WCM), New York, NY, and Ospedale San Giovanni Battista delle Molinette, Turin, Italy.

IgM stimulation

Cells (3 million) per sample were washed and the pellet resuspend in 100 ul of PBS with 20 ug of anti-IgM or no antibody in 1.5 ml vials. After 30 minutes, IgM was washed out and RNA extracted 2.5h or 6h later.

Cell proliferation assay

Cells nucleofected with siRNAs or LNA oligonucleotides, or treated with AZ6102 were cultured for 72 h at 37°C 5% CO_2. Proliferation was assessed by MTT assay, as previously described (1). Proliferation of cells stably expressing GECPAR or of PDTX-RN after transient GECPAR knock down was followed in real time by Incucyte (Sartorius) live cells analysis for at least five days. Briefly, cells were counted and seeded in triplicates in 96-well plate coated with poly- L-ornithine (Sigma) to allow a monolayer growth. Different cell densities were tested to select the best cellular concentration for each model (OCI-Ly10, 10,000 cells/well, SUDHL2, 20,000 cells/well, PDTX-RN, 30,000 cells/well) Every 4h independent images (n=9) were acquired per each well. Analysis was performed by Incucyte Cell-by-Cell Analysis Software Module and cell proliferation was quantified by counting the number of phase objects over time. Cells expressing GFP were also counted by green object count module, based on fluorescence intensity. The count average of nine images was calculated for each replicate and normalized to the first acquired count (t_0). A specific green fluorescence threshold (GCU, green calibrated unit) was calculated for each cell line to distinguish cells with different fluorescence intensity. Statistical significance was determined using a two-tailed t-test with a threshold of $p< 0.05$.

RNA extraction

Total RNA was obtained from cell lines by phenol:chloroform extraction. RNA samples were treated with DNase I (Qiagen). To examine intracellular distribution of the transcripts cellular lysates were fractionated as previously described.(4)

Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)

Strand-specific quantitative RT-PCR (qRT-PCR) was performed using Quanti Fast SYBR Green RT-PCR Kit (Qiagen) on an ABI Step One Plus (Applied Biosystems). Only the forward primer was added to the reverse transcriptase reaction to selectively amplify the antisense strand and only the reverse primer to selectively amplify the sense strand. PolyA+ RNA was reverse transcribed with Superscript III and oligo dT while total RNA was reverse transcribed with random hexamers; mRNAs were measured from cDNA reverse transcribed with the SuperScript III First-Strand Synthesis SuperMix (ThermoFisher). Quantitative real time PCR (qPCR) was then performed using the SYBR Green FAST qPCR mix (KAPA Biosystem). qRT-PCR data were analyzed using ΔCt method after estimation of PCR efficiency with LinREG PCR software (5) and then normalized to GAPDH or β-actin as reference genes. Statistical significance was determined using a two-tailed t-test with a threshold of $p< 0.05$. Primer sequences are reported in Supplementary Table 10.

5' and 3' Rapid Amplification of cDNA Ends

5' RACE was performed with gene-specific primers for GECPAR (Supplementary Table 10) using the Invitrogen 5' RACE System and RNA from OCI-LY1 cells. cDNA was purified, tailed with dCTP and amplified consecutively with gene specific primers and either Abridged Anchor primer or Abridged Universal Amplification primer provided in the 5'RACE system kit. For 3' RACE, total RNA was polyadenylated with Poly(A) tailing kit (Applied Biosystem), or not. Artificially or naturally polyadenylated RNA was then reverse transcribed and amplified consecutively with gene-specific primers using theInvitrogen 3'RACE system kit. Final PCR products were cloned into the pGEM-T Easy vector (Promega) and sequenced.

GECPAR cloning and overexpression

The GECPAR sequence of 968 bp derived from RACE analysis was amplified from genomic DNA of OCI-LY1 cells using Expand™ High Fidelity PCR System (Roche), cloned into the pGEM T vector (Promega) and subcloned in pCDH-CMV-MCS-EF1-copGFP (System Biosciences, CD511B-1) using XbaI and BamHI restriction sites. Primers containing the restriction sites for PCR amplification are shown in Table S3. Plasmids were amplified in JM109 competent cells and purified by GenElute Plasmid Midiprep Kit (Sigma). DNA sequences of the construct was confirmed by DNA sequencing.

pCDH empty backbone or pCDH_GECPAR were transfected in HEK293 T together with pMD2.VSVG, envelope plasmid, and pCMV-R8.74, packaging plasmid. After 72h viral supernatant was collected and used to infect SUDHL2 or OCI-Ly10 cells (6 ml of viral supernatant, containing polybrene, 8μg/ml per 1 million lymphoma cells). After three consecutive infections, cells were washed and allowed to recover for 6 days before sorting by FACS to enrich for GFP+ cells. After 48h RNA was extracted to determine GECPAR overexpression Cells were then cultured and counted for 11 days to obtain proliferation curves, or seeded for Incucyte experiment.

PDTX-KD (2 million) were infected with 200 μl of viral particles concentrated 100-fold by Lenticoncentrator (Takara) according to manufacturing instructions. Virus was incubated with the cells in 4 ml of medium containing polybrene 8μg/ml, for 24h. Than cells were washed and seeded 30000 in 96-well plate for proliferation assay, or cultured at 1 million/ml to extract RNA and check GFP expression at the end of proliferation assay.

***In Silico* Genomic Analysis**

Public datasets of RNA-Seq from poly A+ and polyA- RNA of CD20+ cells and ChIP-Seq for H3K4 me1, H3K4 me3 and H3K27ac performed in K562 and GM12878, available in the Genome Browser at the UCSC Genome Bioinformatics Site (http://genome.ucsc.edu/index.html), were downloaded and reanalyzed to quantify the bidirectional transcription at *POU2AF1* super-enhancer locus.

The RNA-Seq datasets were pre-processed and analyzed following the ENCODE RNA-Seq pipeline. All details are available at https://www.encodeproject.org/pipelines/ENCPL002LPE/.

ChIP-Seq analysis

Public datasets of ChIP-Seq for BRD4, H3ac, H3K27me3 and RNA pol II after DMSO or JQ1 treatment of OCI-LY1 were downloaded and re-analyzed. Sequence reads obtained from ChIP fragments were aligned to human reference genome hg19 using Bowtie, allowing up to one mismatch per fragment length. Redundant reads were removed and only reads uniquely mapping to the reference genome were used for further analysis. The detection of peaks that are genomic regions enriched by ChIP, relative to the background reads, was carried out using HOMER (v2.6) (6), as previously described (7). All discovered putative peaks were ranked by their Normalized Tag Counts (number of tags found at the peak, normalized to 10 million total mapped tags) and annotated with annotatePeaks.pl subroutine.

RNA-Seq analysis

Total RNA-Seq reads from DLBCL patients (8) were kindly provided by G.I. and L.C.. The raw reads were quality assessed using fastqc (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). For each sample the distribution of unique, multi- and unmapped reads was checked for high proportions of unmapped or multi mapped reads. Reads obtained from RNA sequencing were mapped against the *human* hg38 genome build using the Genecode version 22 annotation. Alignment was done with STAR (v2.4.0h) (9), counting of reads overlapping gene features with HTSeq-Count. Differential gene expression analysis was performed using the voom/limma (10) R package. Transcripts that were expressed at > = 1 count per million mapped reads were considered for further analyses. Differentially expressed genes were defined as those with an empirical Bayes corrected (Benjamini- Hockberg procedure) p-value <0.05.

PolyA RNA-Seq was performed in U2932 transfected with GECPAR LNA 461, GECPAR LNA 563 or scramble control for 48h and in SUDHL2 stably overexpressing GECPAR and GFP or GFP alone. RNA was extracted and libraries prepared using NEBNext Ultra II Directional RNA Library Prep.

Public murine polyA RNA-Seq data (GSE72018) were interrogated to represent GECPAR expression by box plot graphs.

DNA Copy Number Alteration analysis

The cohort of patients analyzed for copy number alteration comprised 737 cases of mature lymphoid tumors and were previously described (11-15).

Microarray analysis

Gene expression profiles of untreated lymphoma cell lines were retrieved from our previously deposited NCBI GEO series GSE94669, and analyzed as previously described (1). Gene expression profiling of DLBCL patient samples was downloaded from GEO (GSE10846), the dataset includes 181 clinical samples from CHOP-treated patients and 233 clinical samples from Rituximab-CHOP-treated patients. The data were analyzed with Microarray Suite version 5.0 (MAS 5.0) using Affymetrix default analysis settings and global scaling as normalization method. The trimmed mean target intensity of each array was arbitrarily set to 500.

Kaplan-Meier analysis

Survival functions were defined according to the revised National Cancer Institute criteria and estimated using the Kaplan-Meier method. Patient groups were defined using the GECPAR gene expression profile: high expressor if GECPAR expression is higher than the 70^{th} percentile and low expressor if the GECPAR expression is lower than the 15^{th} percentile. The patients group were compared by the log-rank test. Cox proportional hazard models were used for univariate analysis and the estimation of hazard ratios (HRs).

CHARTseq

CHART Enrichment and RNAseH Mapping experiments were performed as previously described (16, 17). CHART extracts were prepared from 7 x 10^7 OCI-LY1 and U2932 per pulldown and hybridized with 750 pmol biotinylated oligonucleotides cocktail (IDT) (Supplementary Table S11) overnight with rotation at room temperature. Complexes were captured with 60 µl per sample of Streptavidin beads (Sigma), extensively washed and DNA eluted with RNAseH (Sigma) treatment. Cross-linking was reversed in the presence of Proteinase K (Roche), and DNA purified with a PCR purification kit (Qiagen). CHARTseq was performed in both cell lines with two independent samples of pulldown and matched negative control. An input DNA was also prepared and sequenced for each sample. The sequencing of the pre-pools was performed using the NextSeq500 sequencer with v2.0 chemistry from Illumina (San Diego, CA, USA) and 75 bp single reads. The NEBNext Ultra II DNA Library Prep Kit with Purification beads for Illumina (cat.n E7103S New England BioLabs Inc.) was employed with the NEBNext Multiplex Oligos for Illumina (cat.n. E7600S New England BioLabs Inc.) for libraries preparation. 75 bp single-end reads were mapped to hg19 using Bowtie aligner recording positions of uniquely mappable reads. The enrichment of CHART signal was determined relative to the oligo controls. Conservative enrichment profiles were determined using the SPP package (18) (lower bound of enrichment was determined based on a Poisson model, with a confidence interval of $p < 0.001$) and MACS (19) (-B –-bw 120 --broad), as described by Vance and colleagues. (20).

Data mining

For exploratory GECPAR function studies, differences in GEP of GCB DLBCL cell lines dichotomized for GECPAR expression based on median expression value were defined as statistically significant if log FC was > |0.59| with a $P < 0.05$ using the empirical Bayes moderated t-test as implemented in the LIMMA R-package by Carmaweb (https://carmaweb.genome.tugraz.at/carma) (17) Hierarchical clustering dendrograms and heatmaps for GCB DLBCL patients stratified by median GECPAR expression were created using the "heatplot" function of the bioconductor package made4 (21). Functional annotation was performed using Gene Set Enrichment Analysis (GSEA) (22) with all genes preranked by FC as determined by Limma test. Gene sets were considered significantly enriched if $p < 0.05$ and FDR<0.25. Gene ontology analysis was performed using the g-Profiler webtool. The p-value for pathway enrichment was computed using a Fisher's exact test and multiple-test correction was applied.

Characterization of GECPAR binding sites

Genes which were identified as GECPAR-bound from CHART analysis in OCI-LY1 and U2932, were functionally annotated by Panther (http://www.pantherdb.org/) (23) with Fisher's Exact with FDR multiple test correction. Peaks were considered concomitant in OCI-LY1 and U2932 if overlapping within a range of 10kb, as determined by BEDtool. Their FASTA sequences were interrogated by MEME software (24) for *de novo* motif discovery.

Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)

RT-PCR was performed using Verso 1 Step kit Thermostart (ThermoScientific with the indicated primers (Table S10). Samples were analyzed by agarose gel electrophoresis followed by staining with GelRed (Biotium) and imaging with AlphaImager (Innotech). To distinguish the strand direction of transcripts only the forward primer was added to the reverse transcriptase reaction to selectively amplify the antisense strand and only the reverse primer to selectively amplify sense strand**.**

Western blotting

U2932 nucleofected with LNAs against GECPAR were lysed 72h after treatment by hot SDS lysis buffer. SUDHL2 and OCI-Ly10 pCDH or pCDH GECPAR were lysed when they were in exponential growth. 10 µg of extracted proteins were separated on 4–20% precast polyacrylamide gel (Biorad). Immunoblotting was performed with the following antibodies: anti-TLE4 antibody (Abcam, ab140485), anti-CYLD antibody - N-terminal (Abcam, ab153698), anti-CREBBP antibody (Cell signaling, cat. 7389S).

References

1. Tarantelli C, Gaudio E, Arribas AJ, et al. PQR309 Is a Novel Dual PI3K/mTOR Inhibitor with Preclinical Antitumor Activity in Lymphomas as a Single Agent and in Combination Therapy. Clin Cancer Res. 2018;24(1):120-129.

2. Gaudio E, Tarantelli C, Spriano F, et al. Targeting CD205 with the antibody drug conjugate MEN1309/OBT076 is an active new therapeutic strategy in lymphoma models. Haematologica. 2020;105(11):2584-2591.

3. Napoli S, Pastori C, Magistri M, et al. Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells. EMBO J. 2009;28(12):1708-19.

4. Napoli S, Piccinelli V, Mapelli SN, et al. Natural antisense transcripts drive a regulatory cascade controlling c-MYC transcription. RNA Biol. 2017;14(12):1742-1755.

5. Ruijter JM, Ramakers C, Hoogaars WM, et al. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res. 2009;37(6):e45.

6. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell. 2010;38(4):576-89.

7. Mensah AA, Cascione L, Gaudio E, et al. Bromodomain and extra-terminal domain inhibition modulates the expression of pathologically relevant microRNAs in diffuse large B-cell lymphoma. Haematologica. 2018;103(12):2049-2058.

8. Teater M, Dominguez PM, Redmond D, et al. AICDA drives epigenetic heterogeneity and accelerates germinal center-derived lymphomagenesis. Nat Commun. 2018;9(1):222.

9. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29(1):15-21.

10. Law CW, Chen Y, Shi W, et al. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15(2):R29.

11. Chigrinova E, Rinaldi A, Kwee I, et al. Two main genetic pathways lead to the transformation of chronic lymphocytic leukemia to Richter syndrome. Blood. 2013;122(15):2673-82.

12. Boi M, Rinaldi A, Kwee I, et al. PRDM1/BLIMP1 is commonly inactivated in anaplastic large T-cell lymphoma. Blood. 2013;122(15):2683-93.

13. Martinez N, Almaraz C, Vaque JP, et al. Whole-exome sequencing in splenic marginal zone lymphoma reveals mutations in genes involved in marginal zone differentiation. Leukemia. 2014;28(6):1334-40.

14. Rinaldi A, Kwee I, Young KH, et al. Genome-wide high resolution DNA profiling of hairy cell leukaemia. Br J Haematol. 2013;162(4):566-9.

15. Rossi D, Trifonov V, Fangazio M, et al. The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. J Exp Med. 2012;209(9):1537-51.

16. Vance KW. Mapping Long Noncoding RNA Chromatin Occupancy Using Capture Hybridization Analysis of RNA Targets (CHART). Methods in molecular biology (Clifton, NJ). 2017;1468:39-50.

17. Sexton AN, Machyna M, Simon MD. Capture Hybridization Analysis of DNA Targets. Methods in molecular biology (Clifton, NJ). 2016;1480:87-97.

18. Kharchenko PV, Tolstorukov MY, Park PJ. Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nature biotechnology. 2008;26(12):1351-9.

19. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

20. Vance KW, Sansom SN, Lee S, et al. The long non-coding RNA Paupar regulates the expression of both local and distal genes. EMBO J. 2014;33(4):296-311.

21. Culhane AC, Thioulouse J, Perriere G, et al. MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics. 2005;21(11):2789-90.

22. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.

23. Mi H, Muruganujan A, Casagrande JT, et al. Large-scale gene function analysis with the PANTHER classification system. Nat Protoc. 2013;8(8):1551-66.

24. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994;2:28-36.

Table S3. Genes commonly enriched in GCB DLBCL cell lines and GCB-DLBCL patients according to high or low GECPAR expression

GENES CORRELATED TO GECPAR IN GCB-DLBCL CELL LINES AND DLBCL PATIENTS	
ESSENTIAL GENES	**CELL CYCLE**
TFDP1	CCNE2
PCNA	TFDP1
HNRNPC	PCNA
UBA52	MCM3
EIF2S2	MCM7
PSMC3	E2F2
DDB1	PRKDC
EIF3B	CDC7
RAN	CDC16
AFG3L2	MCM5
BUB1B	MCM4
NUTF2	BUB1B
PSMC2	ABL1
XPO1	MYC
PSMD2	ANAPC13
YBX1	RAD21
RPA2	MCM2
RUVBL1	STAG1
RRM1	MCM6
SMC4	HDAC1
SMU1	ATM
RPL7	SKP1
RAD21	
NCBP1	
NUP214	
EIF2B3	
U2AF2	
ZNF207	
CCT3	
COPS5	
DHX9	
CCT2	
SFPQ	
KIF11	
RNPS1	
HCFC1	
MED14	
POLA1	
TCERG1	
ABCE1	
DDX21	
E2F5	
SNRPB	
AP2M1	
POLR2B	
COPS6	
TPR	

Table S6. List of essential genes enriched in U2932 depleted of GECPAR (left) or in SUDHL2 overexpressing GECPAR (right)

Dataset	Limma_GECPARkdlogFCbase.rnk
Upregulated in class	na_pos
GeneSet	ESSENTIAL_ABC_DLBCL_ONCOGENIC_SIGNALING_PMID29925955
Enrichment Score (ES)	0.5168573
Normalized Enrichment Score (NES)	1.8684261
Nominal p-value	0
FDR q-value	0.006865541
	SYK
	PIK3CD
	CARD11
	REL
	BTK
	POU2AF1
	PLCG2
	PTPRC
	CD79B
	PIK3AP1
	PRKCB
	BATF
	CSK
	BCL2L1
	RNF31
	JAK1
	PAX5
	MYD88
	MTOR
	SPIB
	EBF1
	IKBKG
	IL10RA
	MALT1
	IRAK4
	IKBKB
	UNC93B1
	BLNK
	IRF4

Dataset	Limma_GECPARovlogFCbase.rnk
Upregulated in class	na_pos
GeneSet	ESSENTIAL_GCB_DLBCL_ONCOGENIC_SIGNALING_PMID29925955
Enrichment Score (ES)	0.6469427
Normalized Enrichment Score (NES)	1.5895188
Nominal p-value	0.003157895
FDR q-value	0.022455128
	EHD3
	FCRL1
	LCK
	GPR114
	CD27
	LHPP
	PACSIN1
	SYK
	SLC2A5
	KCNN3
	PTAFR
	SH2B2
	REL
	AIM2
	PIK3CG
	SEMA4A
	PPIL2
	TNFSF10
	GCNT1
	TPCN2
	TPCN1
	SYNE2
	POLD4
	BCAS4
	NCALD
	MAST3
	COTL1
	RECQL5
	MET
	PTPN18
	PITPNC1
	MYO1E
	SMARCA4
	ANK3
	FAM53B
	CD83
	DOK3
	FCRLB
	GPR160
	PRKCD
	ITPKB
	FANCA
	BPTF
	PIP4K2A
	S1PR2
	SH3KBP1
	CD86
	CCDC69
	STX7
	MEF2C
	EBF1

Table S9. siRNAs and LNAs

NAME	SENSE STRAND	ANTISENSE STRAND
GECPAR +461 siRNA	ACUGAUCUAAAGCCAAAGUTT	ACUUUGGCUUUAGAUCAGUTT
GECPAR +563 siRNA	GUGCUAUGAGGGAGUGAUUTT	AAUCACUCCCUCAUAGCACTT
GL3 siRNA	CUUACGCUGAGUACUUCGATT	UCGAAGUACUCAGCGUAAGTT
SCR LNA	----------	AA+CCATT+CTCC+GTCAA+ACC
GECPAR +461 LNA	----------	AC+TT+TGGCTT+TAGA+TCAGT
GECPAR +563 LNA	----------	AA+TCACT+CCCT+CATAG+CAC
GECPAR +489 LNA	----------	C+ATAGC+ACTG+TCTGAGGG+CT
GECPAR +856 LNA	----------	AGT+TCTGAC+TTGGCT+TCTG+T

+ LNA modified nucleotide

Table S10 Primers

NAME	SEQUENCE	APPLICATION
GECPAR +545 Fw	GTGGTCAGCCCTCAGACAGT	3'RACE, RT-PCR
GECPAR +625 Rev	CAGCATGAACTGCCCCTAAT	5'RACE, RT-PCR
GECPAR+804 Fw	ACCTAGGCGATGACCTTGTG	3'RACE, RT-PCR
GECPAR +900 Rev	GGCTGCACTTGCTTCTCTCT	5'RACE, RT-PCR
LOC100132078+3473 Rev	TTGAAAGCAGCAGCGAAAG	3'RACE
POU2AF1 ex2 Fw	AGGAGCCAGTGAAGGAACTG	qRT-PCR
POU2AF1 ex4 Rev	GGCAGCCTCCTCTGTCACT	qRT-PCR
CREBBP ex9 Fw	CATGTACGAGTCTGCCAACAG	qRT-PCR
CREBBP ex10 Fw	GCGACCTCCGTTTTTCTTCT	qRT-PCR
CREB5 ex6 Fw	AACCCTACAATGCCAGGATCT	qRT-PCR
CREB5 ex7 Rev	CACAGGGGTTGCTGAGATTT	qRT-PCR
TLE4 ex12 Fw	GGATTTGATCCACACCATCA	qRT-PCR
TLE4 ex13 Rev	TCTGACCATCTGCGCTAACA	qRT-PCR
CYLD Fw	CAGCCGGTTTCCAATCAG	qRT-PCR
CYLD Rev	ACCCTGGATGCCTTTCTTCT	qRT-PCR
GAPDH ex3 Fw	TCACCAGGGCTGCTTTTAAC	qRT-PCR
GAPDH ex4 Rev	GGGTGGAATCATATTGGAACA	qRT-PCR
GAPDH ctr neg Fw	CGTAGCTCAGGCCTCAAGAC	qPCR
GAPDH ctr neg Rev	GTCGAACAGGAGGAGCAGAG	qPCR
ALBUMIN ctr neg Fw	TTGCTAGATGGAGGGCAAAC	qPCR
ALBUMIN ctr neg Rev	TTTAAATCCGCACCCTTCTG	qPCR
BACH2_GECPAR_BS Fw	ATGTGGGGTCCTTTCCTTCT	qPCR
BACH2_GECPAR_BS Rev	TTGGAACCCAGTGAAAGATG	qPCR
11q23_GECPAR_BS Fw	AGCCACTCCTCGCAGTCTT	qPCR
11q23_GECPAR_BS Rev	GAGTCAGAATGTTGAAAGGCATAA	qPCR
TTK_GECPAR_BS FW	AATGGGACCATTTAAGTGAAAG	qPCR
TTK _GECPAR_BS REV	TCCTGAAGGAAATATCACAGAGTG	qPCR
ACTL6A_ GECPAR _BS FW	GACCCAGAAAACAAATCCAGAC	qPCR
ACTL6A_ GECPAR _BS REV	GGGGAACATGAAGGAAAAATC	qPCR
ATP11B_ GECPAR _BS FW	ACAGCTGATGCCTGGAGTTC	qPCR
ATP11B_ GECPAR _BS REV	GCATTAGCTGAGGTGGATTG	qPCR
XRCC4_ GECPAR _BS FW	ACAGATGTCTCTTCCACATTCTGA	qPCR
XRCC4_ GECPAR _BS REV	ATCCAGCAATCCCACTTCTG	qPCR
MCTP_ GECPAR _BS FW	TGGTAGTCATCCTCTGTCCAAATA	qPCR
MCTP_ GECPAR _BS REV	CAAATGCGTTCCTATGTGTCA	qPCR
BET1_ GECPAR _BS FW	AAGGGGTTGGCTATCTCTGA	qPCR
BET1_ GECPAR _BS REV	ATTGTCATGCATGGCTTCTG	qPCR
CREB5_ GECPAR _BS FW	TTAACCAAGGTTCCCCACAG	qPCR
CREB5_ GECPAR _BS REV	AGAGGTGGACAACCCAACTG	qPCR
ECT2_ GECPAR _BS FW	GGAATCTACACAGCCGTTACAA	qPCR
ECT2_ GECPAR _BS REV	GGTAATGAACATCTTTCCAGGTCTA	qPCR
XbaI GECPAR Fw	GCTCTAGAGCGCAGTGATTCAAGACACTTGG	GECPAR cloning
BamHI GECPAR Rev	CGGGATCCCGGTCATTCTTACTTTTAACAGCAC	GECPAR cloning

Table S11 CHART probes

NAME	SEQUENCE	APPLICATION
GECPAR_AS_oligo_1	CCTGGTTTCCAGTTTAGTTGTTC	RNAseH mapping
GECPAR _AS_oligo_2	TCCCTGGTTTCCAGTTTAGTTGT	RNAseH mapping
GECPAR _AS_oligo_3	GTTCCTGTTGTTATGCCTGAGGA	RNAseH mapping
GECPAR _AS_oligo_4	GTGTTCCTGTTGTTATGCCTGAG	RNAseH mapping
GECPAR _AS_oligo_5	CTGTGTTCCTGTTGTTATGCCTG	RNAseH mapping
GECPAR _AS_oligo_6	GCTTTGTGGAGAGTAAGACGTCG	RNAseH mapping
GECPAR _AS_oligo_7	TTGACCAAACTTGGCTTTGTGGA	RNAseH mapping
GECPAR _AS_oligo_8	GGAGCTTGACCAAACTTGGCTTT	RNAseH mapping
GECPAR _AS_oligo_9	CTTAGGGGATTTCCTCTCTGTGG	RNAseH mapping
GECPAR _AS_oligo_10	AACTTAGGGGATTTCCTCTCTGT	RNAseH mapping
GECPAR _AS_oligo_11	GTTTTCATGTTCTTGGGGCATGG	RNAseH mapping
GECPAR _AS_oligo_12	GGACTGTTTTCATGTTCTTGGGG	RNAseH mapping
GECPAR _AS_oligo_13	GCATCTGGACTGTTTTCATGTTC	RNAseH mapping
GECPAR _AS_oligo_14	TGCATTGCAGGTTCATGCATCTG	RNAseH mapping
GECPAR _AS_oligo_15	TAGCACTGTCTGAGGGCTGACCA	RNAseH mapping
GECPAR _AS_oligo_16	TCCCTCATAGCACTGTCTGAGGG	RNAseH mapping
GECPAR _AS_oligo_17	CAATCACTCCCTCATAGCACTGT	RNAseH mapping
Biotin_AS_oligo_2	TCCCTGGTTTCCAGTTTAGTTGT	CHART
Biotin _AS_oligo_4	GTGTTCCTGTTGTTATGCCTGAG	CHART
Biotin _AS_oligo_6	GCTTTGTGGAGAGTAAGACGTCG	CHART
Biotin _AS_oligo_16	TCCCTCATAGCACTGTCTGAGGG	CHART
Biotin_scr-oligo1	ctCCactgatCAtgcTgtcgGaG	CHART
Biotin_scr-oligo2	cttccGtgTTgcacTTatGgttT	CHART

Table Captions

Table S1 (separate file)

Limma results comparing gene expression profiles of GCB-DLBCL cell lines dichotomized by median GECPAR expression. MeanM represents modulation (fold change) for each gene in high GECPAR vs low GECPAR expression level.

Table S2 (separate file)

Limma test performed on gene expression profiles of GCB-DLBCL patients dichotomized for median GECPAR expression. LogFC represents modulation for each gene in high GECPAR vs low GECPAR expression level.

Table S4 (separate file)

Limma test performed on gene expression profile of U2932 after GECPAR knockdown versus control. LogFC represents modulation for each gene in GECPAR knockdown vs control.

Table S5 (separate file)

Limma test performed on gene expression profile of SUDHL2 overexpressing GECPAR versus control. LogFC represents modulation for each gene in GECPAR overexpressing cells vs control.

Table S7 (separate file)

GECPAR binding sites detected by CHARTseq in OCI-LY1. Fold change represents enrichment of GECPAR binding relative to negative control.

Table S8 (separate file)

GECPAR binding sites detected by CHARTseq in U2932. Fold change represents enrichment of GECPAR binding relative to negative control.

Supplementary figures legends

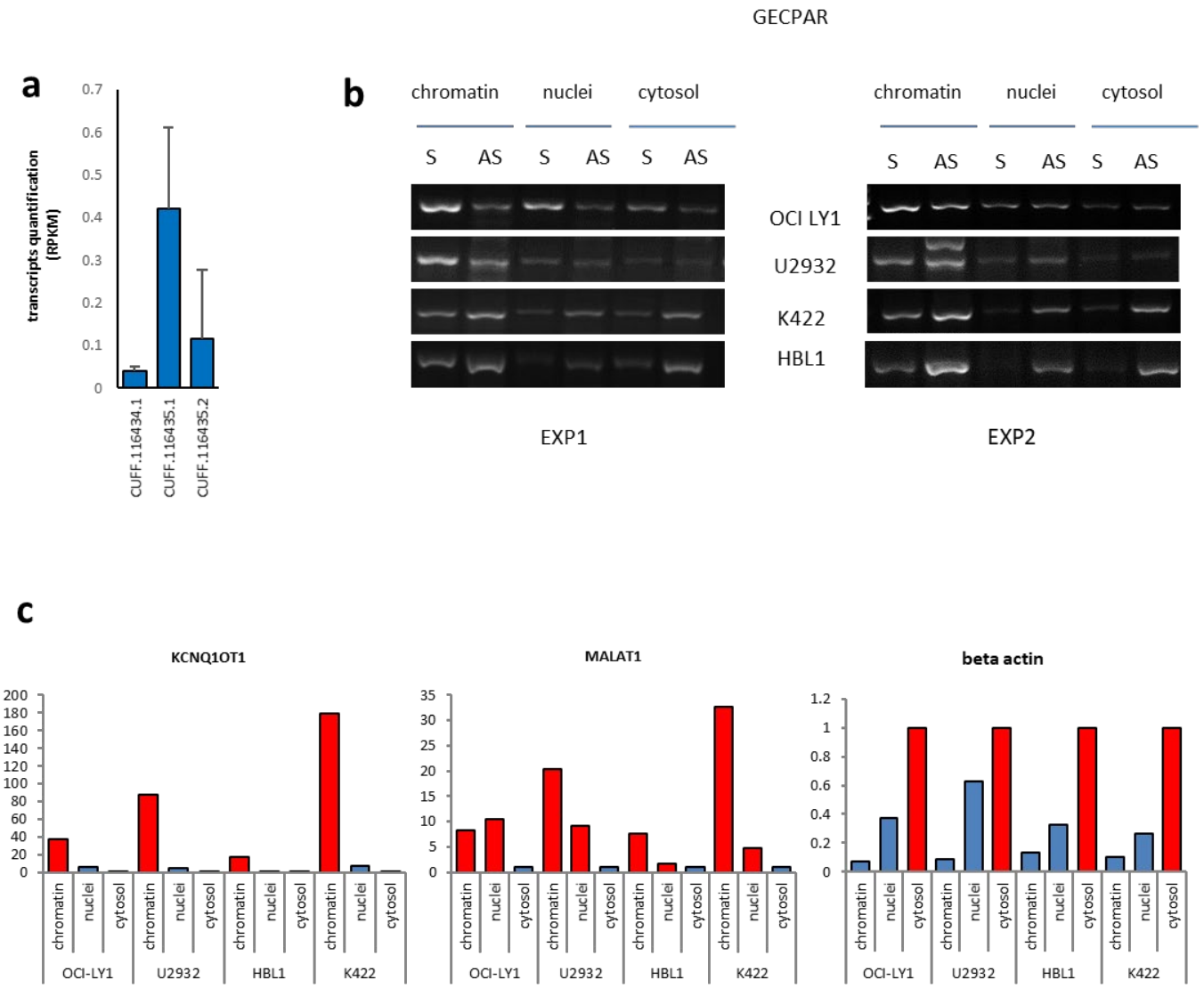
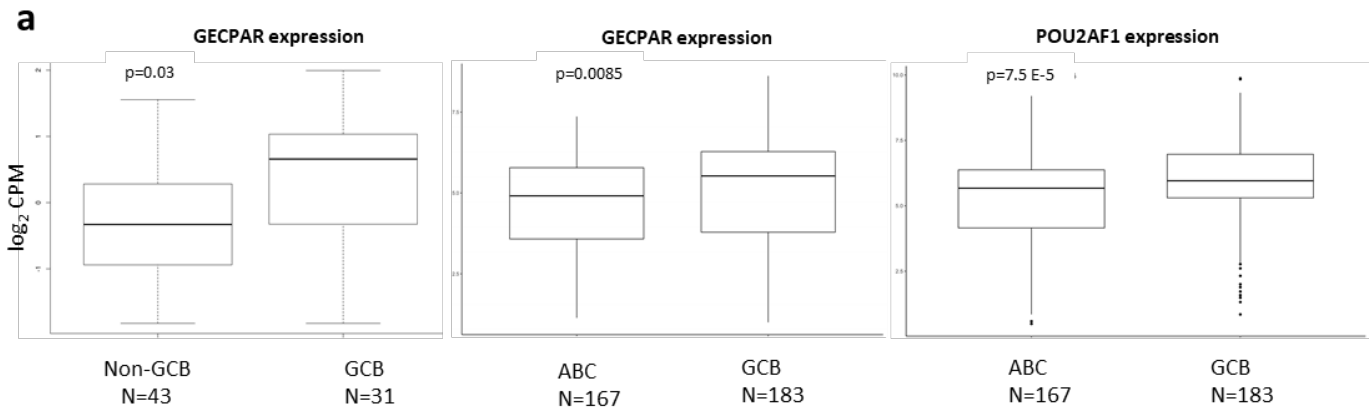
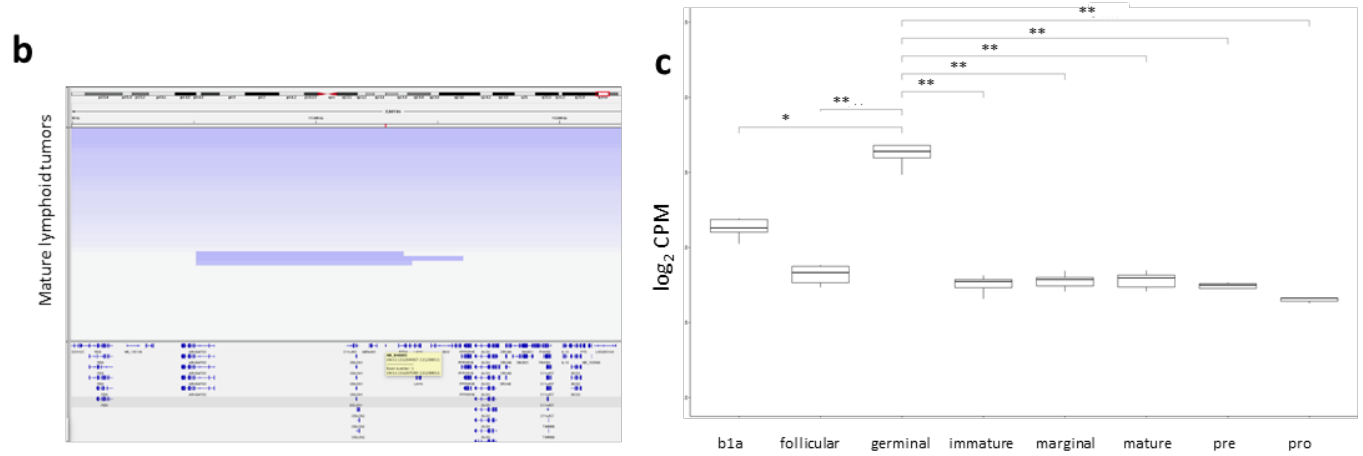


Fig. S1 a, Quantification of De Novo reconstructed transcripts in CD20+ RNAseq in correspondence of LOC100132078 transcript **b**. Directional semiquantitative RT-PCR of two independent experiments of subcellular fractionation of GECPAR and its antisense transcript. **c**, qRT-PCR of KCNQ1OT1 as a positive control for chromatin associated RNA, MALAT1 as a nuclear soluble RNA and mature beta-actin mRNA as a cytosolic RNA.


a
GECPAR expression
p=0.03
log2 CPM
Non-GCB
N=43
GCB
N=31
GECPAR expression
p=0.0085
ABC
N=167
GCB
N=183
POU2AF1 expression
p=7.5 E-5
ABC
N=167
GCB
N=183


b
Mature lymphoid tumors
c
log2 CPM
b1a
follicular
germinal
immature
marginal
mature
pre
pro

116

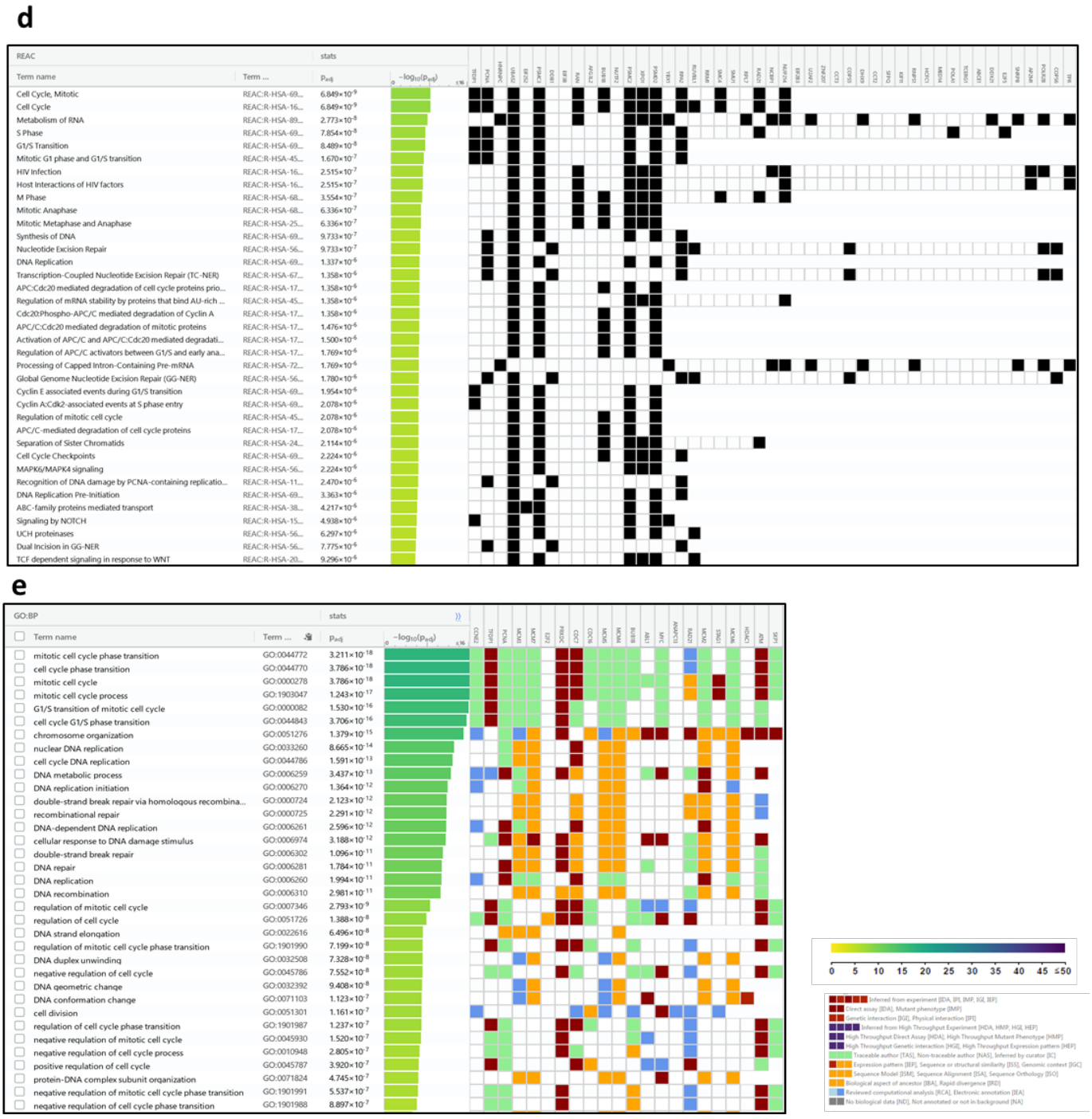
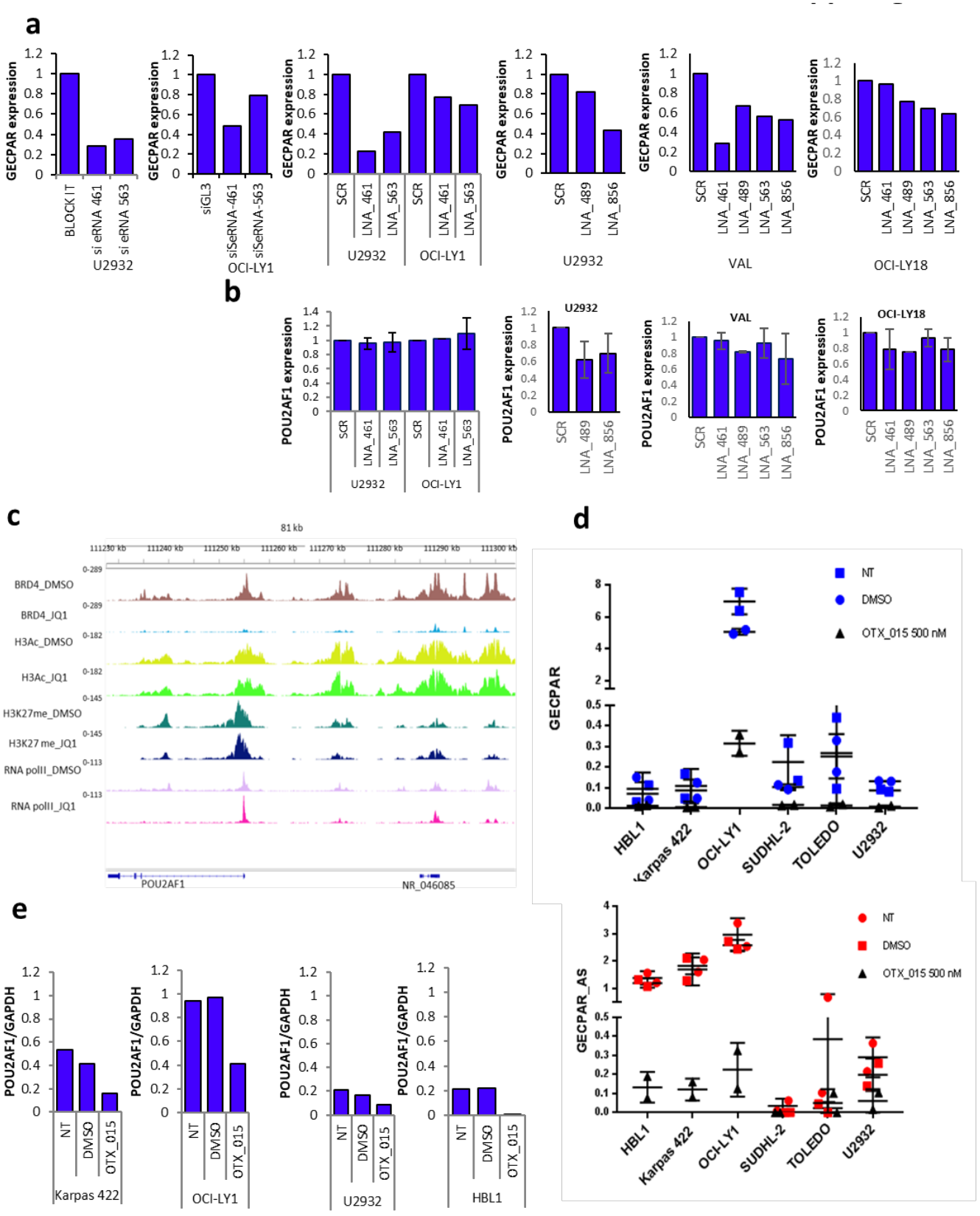


Fig. S2 a, Box plots of GECPAR expression quantified by total RNA seq in GCB or ABC DLBCL patients in a validation cohort (left), box plots of GECPAR (middle) and POU2AF1 (right) expression quantified by microarray in a large validation cohort of GCB or ABC DLBCL patients. **b**, Copy number alterations of 11q23 in 737 mature lymphoid tumors. The red interval indicates the genomic locus of GECPAR and its RefSeq ID and relative coordinates are indicated in the yellow box.**c**, Boxplots of murine GECPAR orthologue expression stratified for cell of origin, * p<0.05, **<0.005 **d**, Gene ontology classification by gProfiler of the essential genes commonly enriched in patients and cell lines with high GECPAR expression. **e**, Gene ontology classification by gProfiler of cell cycle gene set elements enriched in cell lines and patients with high GECPAR expression.


a
GECPAR expression
BLOCK IT
si eRNA 461
si eRNA 563
U2932
siGL3
siSeRNA-461
siSeRNA-563
OCI-LY1
SCR
LNA_461
LNA_563
U2932
OCI-LY1
SCR
LNA_489
LNA_856
U2932
SCR
LNA_461
LNA_489
LNA_563
LNA_856
VAL
SCR
LNA_461
LNA_489
LNA_563
LNA_856
OCI-LY18
b
POU2AF1 expression
SCR
LNA_461
LNA_563
U2932
OCI-LY1
U2932
SCR
LNA_489
LNA_856
VAL
SCR
LNA_461
LNA_489
LNA_563
LNA_856
OCI-LY18
SCR
LNA_461
LNA_489
LNA_563
LNA_856
c
81 kb
BRD4_DMSO
BRD4_JQ1
H3Ac_DMSO
H3Ac_JQ1
H3K27me_DMSO
H3K27 me_JQ1
RNA polII_DMSO
RNA polII_JQ1
POU2AF1
NR_046085
d
GECPAR
NT
DMSO
OTX_015 500 nM
HBL1
Karpas 422
OCI-LY1
SUDHL-2
TOLEDO
U2932
e
POU2AF1/GAPDH
NT
DMSO
OTX_015
Karpas 422
OCI-LY1
U2932
HBL1
GECPAR_AS
NT
DMSO
OTX_015 500 nM
HBL1
Karpas 422
OCI-LY1
SUDHL-2
TOLEDO
U2932

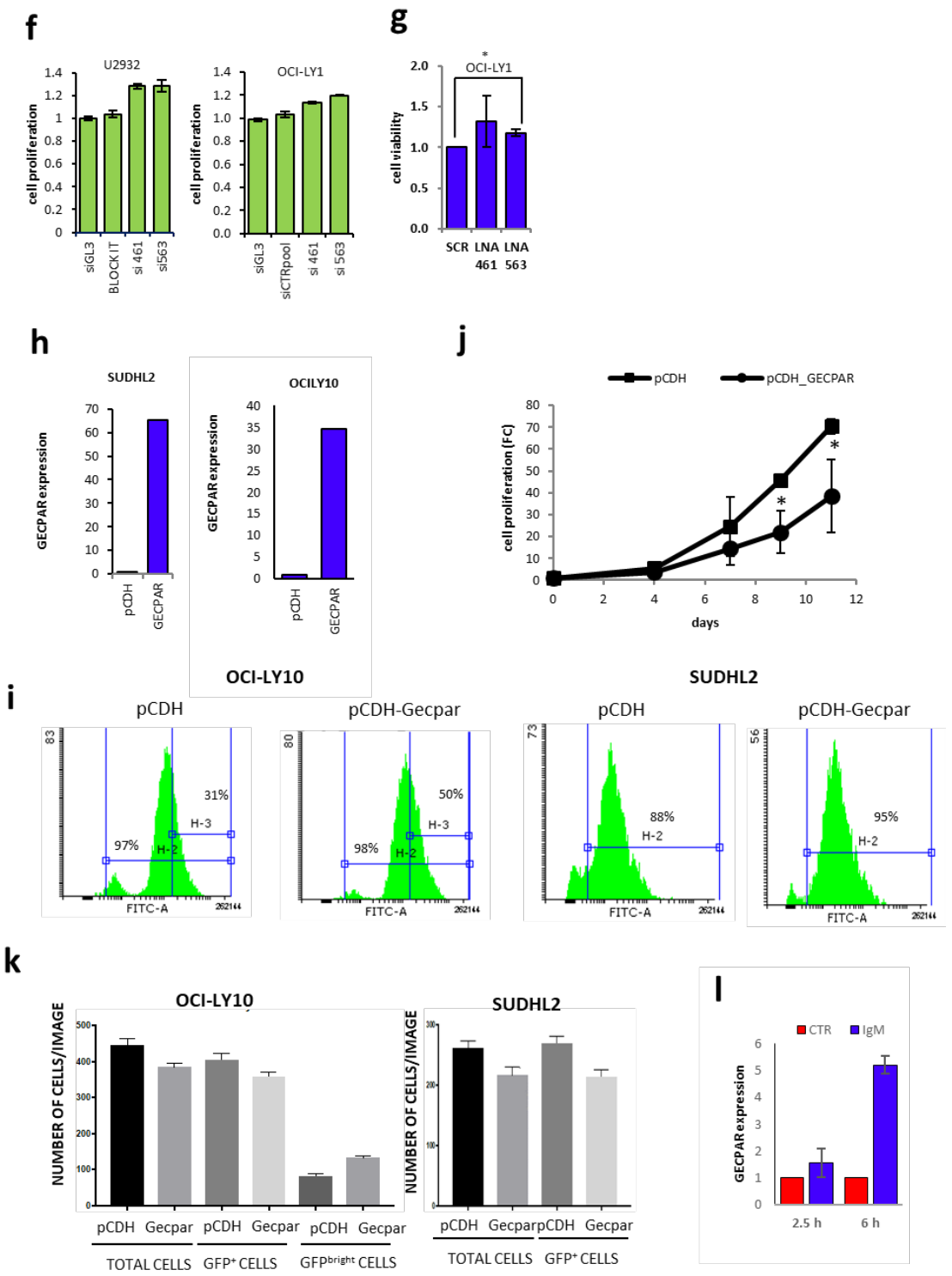


Fig. S3 a,GECPAR expression 24h after interference with two different siRNA in U2932 and OCI-Ly1 and with four different LNA antisense oligonucleotides in U2932, OCI-Ly1, VAL and OCI-Ly18. GECPAR expression is normalized to samples transfected with negative controls. Numerical codes associated to siRNA and LNAs are referred to the first nucleotide recognized in GECPAR transcript relative to its transcription start site **b,** POU2AF1 gene expression after interference with GECPAR by four different LNA antisense oligonucleotides in U2932, OCI-LY1, VAL and OCI-Ly18. **c**, Occupancy of BRD4, H3Ac and RNA pol II at *POU2AF1* and LOC100132078 loci determined by ChIP-Seq after treatment of OCI-LY1 with DMSO or JQ1. **d**, **top**, GECPAR expression in six DLBCL cell lines treated with DMSO or OTX-015 for 4 h. Pool of two independent experiments**; bottom,** GECPAR antisense transcript expression in 6 DLBCL cell lines treated with DMSO or OTX-015 for 4 h. Pool of two independent experiments. **e**, POU2AF1 downregulation 4h after OTX-015 treatment in 4 DLBCL cell lines . **f**, MTT proliferation assay 72 h after transfection with negative controls or

siRNAs 461 or 563. Representative experiment. **g**, MTT proliferation assay 72 h after transfection with negative controls or LNA 461 or 563 in OCI-Ly1. Average of three independent experiments. **h,** GECPAR levels in SUDHL2 and OCI-Ly10 transduced with empty vector or overexpression vector. Representative experiment. **i**, GFP expression measured by FACS at t_0 of Incucyte experiment in OCI-Ly10 and SUDHL2 stably transduced with pCDH empty vector or pCDH-Gecpar vector. H2, percentage of total GFP positive cells, H3, percentage of GFP ^{bright} cells. **j,** Growth curve of SUDHL2 parental and SUDHL2_overexpressing GECPAR, performed after sorting of GFP positive cells. Average of three independent experiments **k,** Number of total cells**,** GFP positive cells and GFP ^{bright} cells counted by Incucyte instrument at t_0 of proliferation assay in OCI-Ly10 and SUDHL2 stably transduced with pCDH empty vector or pCDH-Gecpar vector. **l,** GECPAR levels in U2932 stimulated for 2.5 or 6h with 20 µg of anti-IgM. Average of three independent experiments.

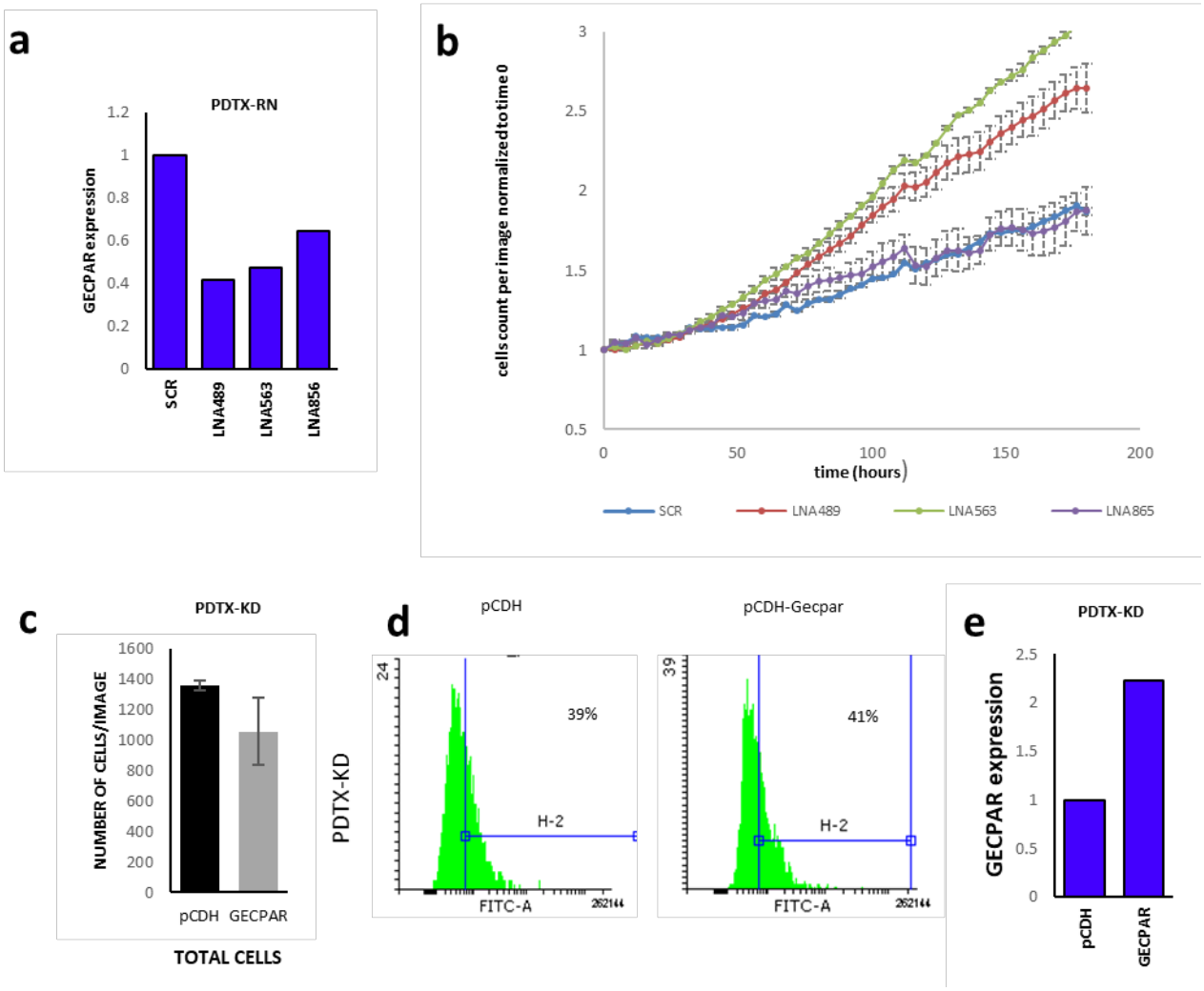


Fig. S4 a, GECPAR expression 48h after interference with four different LNA antisense oligonucleotides in PDTX-RN. **b,** Proliferation assay performed with Incucyte instrument in PDTX-RN nucleofected with negative control (SCR) and three different GECPAR specific LNA antisense oligonucleotides and followed for 8 days. Representative experiment. **c,** Number of total cells counted by Incucyte instruments at t_0 in PDTX-KD transduced with pCDH or pCDH-Gecpar vector. **d**, Percentage of GFP positive PDTX-KD, 9 days after transduction with pCDH or pCDH- Gecpar vectors. **e,** Gecpar expression quantified by qRT-PCR in PDTX-KD, 9 days after transduction with pCDH or pCDH- Gecpar vectors.

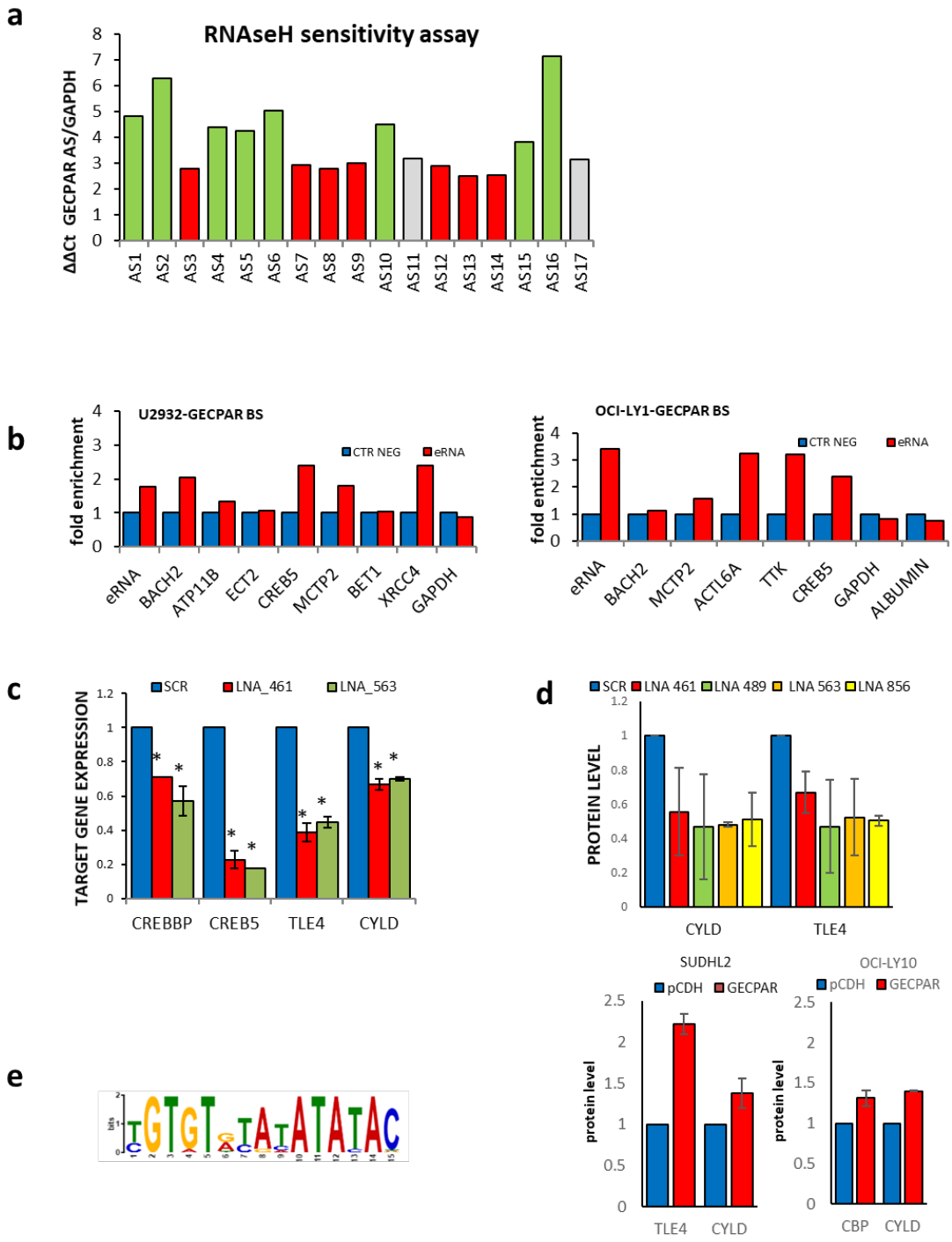
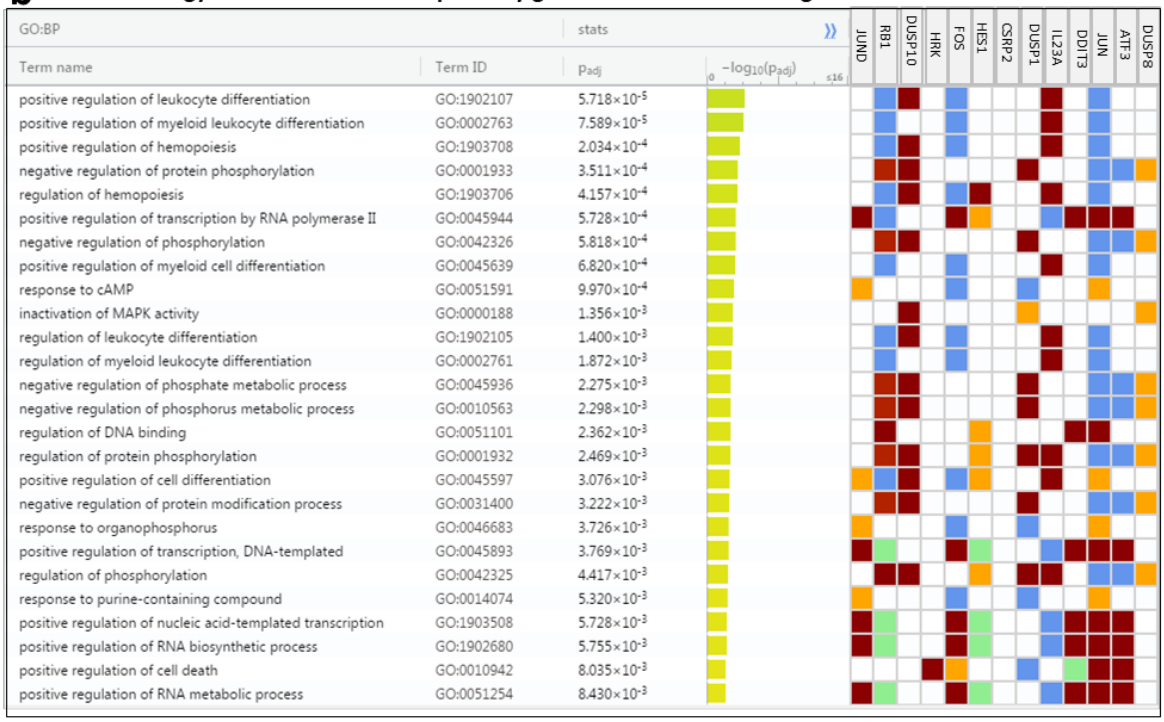


Fig. S5 a, GECPAR level in OCI-LY1 RNA extracted from chromatin after incubation with 17 different antisense oligonucleotides designed to bind GECPAR and treatment with RNAse H. **b**, DNA enrichment after GECPAR pulldown in U2932 (left) or OCI-LY1 (right), concordant with representative peaks from CHARTseq. **c**. Downregulation of direct targets of GECPAR after GECPAR inhibition by two different LNA oligonucleotides in U2932. Average of three independent experiments. * P <0.05 **d**. **Top**, Downregulation at protein level of direct GECPAR targets after GECPAR inhibition by four different LNA oligonucleotides in U2932. **Bottom**, Upregulation at protein level of direct GECPAR targets of in SUDHL2 and OCI-Ly10 stably overexpressing GECPAR Average of three independent experiments. **e**. GECPAR binding motif predicted by MEME

a **Gene ontology classification of TGF β pathway geneset elements downregulated in GECPAR knock down**

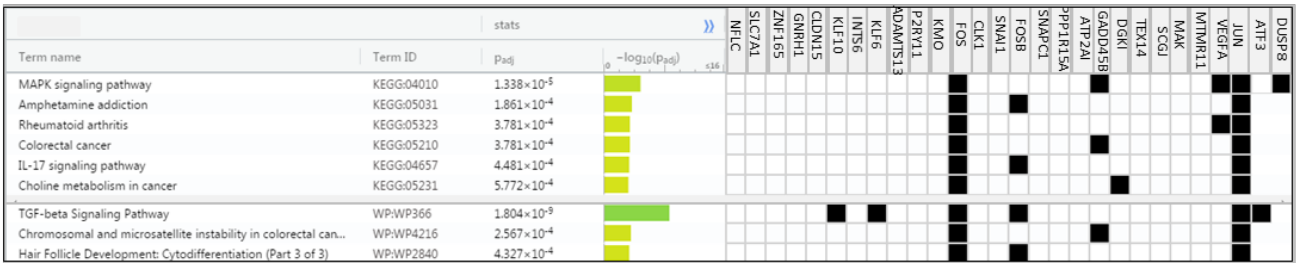
GO:BP		stats	TGFBR1	SMURF1	SMURF2	SMAD7	SKI	KLF10	FNTA	ARID4B	SKIL	LTBP2	SLC20A1	PPP1R15A	TJP1
Term name	Term ID	p_{adj}													
transforming growth factor beta receptor signaling pathway	GO:0007179	1.876×10^{-12}													
cellular response to transforming growth factor beta stimulus	GO:0071560	1.535×10^{-11}													
response to transforming growth factor beta	GO:0071559	1.988×10^{-11}													
transmembrane receptor protein serine/threonine kinase si...	GO:0007178	3.685×10^{-10}													
negative regulation of transforming growth factor beta rece...	GO:0030512	1.522×10^{-8}													
negative regulation of cellular response to transforming gro...	GO:1903845	1.779×10^{-8}													
regulation of transforming growth factor beta receptor sign...	GO:0017015	1.736×10^{-7}													
regulation of cellular response to transforming growth facto...	GO:1903844	1.926×10^{-7}													
cellular response to growth factor stimulus	GO:0071363	2.158×10^{-7}													
negative regulation of transmembrane receptor protein seri...	GO:0090101	2.360×10^{-7}													
negative regulation of BMP signaling pathway	GO:0030514	3.103×10^{-7}													
response to growth factor	GO:0070848	3.132×10^{-7}													
negative regulation of cellular response to growth factor sti...	GO:0090288	1.268×10^{-6}													
regulation of BMP signaling pathway	GO:0030510	5.720×10^{-6}													
enzyme linked receptor protein signaling pathway	GO:0007167	7.081×10^{-6}													
regulation of transmembrane receptor protein serine/threo...	GO:0090092	1.264×10^{-5}													
regulation of cellular response to growth factor stimulus	GO:0090287	4.148×10^{-5}													
BMP signaling pathway	GO:0030509	8.312×10^{-5}													
cellular response to endogenous stimulus	GO:0071495	9.693×10^{-5}													
response to BMP	GO:0071772	1.250×10^{-4}													
cellular response to BMP stimulus	GO:0071773	1.250×10^{-4}													
response to endogenous stimulus	GO:0009719	4.267×10^{-4}													
activin receptor signaling pathway	GO:0032924	5.286×10^{-3}													

b **Gene ontology classification of ATF2 pathway geneset elements downregulated in GECPAR knock down**



GO:BP		stats	JUND	RB1	DUSP10	HRK	FOS	HES1	CSRP2	DUSP1	IL23A	DDIT3	JUN	ATF3	DUSP8
Term name	Term ID	p_{adj}													
positive regulation of leukocyte differentiation	GO:1902107	5.718×10^{-5}													
positive regulation of myeloid leukocyte differentiation	GO:0002763	7.589×10^{-5}													
positive regulation of hemopoiesis	GO:1903708	2.034×10^{-4}													
negative regulation of protein phosphorylation	GO:0001933	3.511×10^{-4}													
regulation of hemopoiesis	GO:1903706	4.157×10^{-4}													
positive regulation of transcription by RNA polymerase II	GO:0045944	5.728×10^{-4}													
negative regulation of phosphorylation	GO:0042326	5.818×10^{-4}													
positive regulation of myeloid cell differentiation	GO:0045639	6.820×10^{-4}													
response to cAMP	GO:0051591	9.970×10^{-4}													
inactivation of MAPK activity	GO:0000188	1.356×10^{-3}													
regulation of leukocyte differentiation	GO:1902105	1.400×10^{-3}													
regulation of myeloid leukocyte differentiation	GO:0002761	1.872×10^{-3}													
negative regulation of phosphate metabolic process	GO:0045936	2.275×10^{-3}													
negative regulation of phosphorus metabolic process	GO:0010563	2.298×10^{-3}													
regulation of DNA binding	GO:0051101	2.362×10^{-3}													
regulation of protein phosphorylation	GO:0001932	2.469×10^{-3}													
positive regulation of cell differentiation	GO:0045597	3.076×10^{-3}													
negative regulation of protein modification process	GO:0031400	3.222×10^{-3}													
response to organophosphorus	GO:0046683	3.726×10^{-3}													
positive regulation of transcription, DNA-templated	GO:0045893	3.769×10^{-3}													
regulation of phosphorylation	GO:0042325	4.417×10^{-3}													
response to purine-containing compound	GO:0014074	5.320×10^{-3}													
positive regulation of nucleic acid-templated transcription	GO:1903508	5.728×10^{-3}													
positive regulation of RNA biosynthetic process	GO:1902680	5.755×10^{-3}													
positive regulation of cell death	GO:0010942	8.035×10^{-3}													
positive regulation of RNA metabolic process	GO:0051254	8.430×10^{-3}													

c **Kegg and Wikipathway classification of RELA DN V1 UP geneset elements downregulated in GECPAR knock down**



		stats	NFLC	SLC7A1	ZNF165	GNRH1	CLDN15	KLF10	INTS6	KLF6	ADAMTS13	P2RY11	KMO	FOS	CLK1	SNA1	FOSB	SNAPC1	PPP1R15A	ATP2A1	GADD45B	DGKI	TEX14	SCGJ	MAK	MTMR11	VEGFA	JUN	ATF3	DUSP8
Term name	Term ID	p_{adj}																												
MAPK signaling pathway	KEGG:04010	1.338×10^{-5}																												
Amphetamine addiction	KEGG:05031	1.861×10^{-4}																												
Rheumatoid arthritis	KEGG:05323	3.781×10^{-4}																												
Colorectal cancer	KEGG:05210	3.781×10^{-4}																												
IL-17 signaling pathway	KEGG:04657	4.481×10^{-4}																												
Choline metabolism in cancer	KEGG:05231	5.772×10^{-4}																												
TGF-beta Signaling Pathway	WP:WP366	1.804×10^{-9}																												
Chromosomal and microsatellite instability in colorectal can...	WP:WP4216	2.567×10^{-4}																												
Hair Follicle Development: Cytodifferentiation (Part 3 of 3)	WP:WP2840	4.327×10^{-4}																												

Fig. S6 Gene ontology classification by gProfiler of TGF-β **(a)** and ATF2 **(b)** pathway gene set elements and genes upregulated after RELA knock down **(c),** downregulated after GECPAR knock down in U2932.

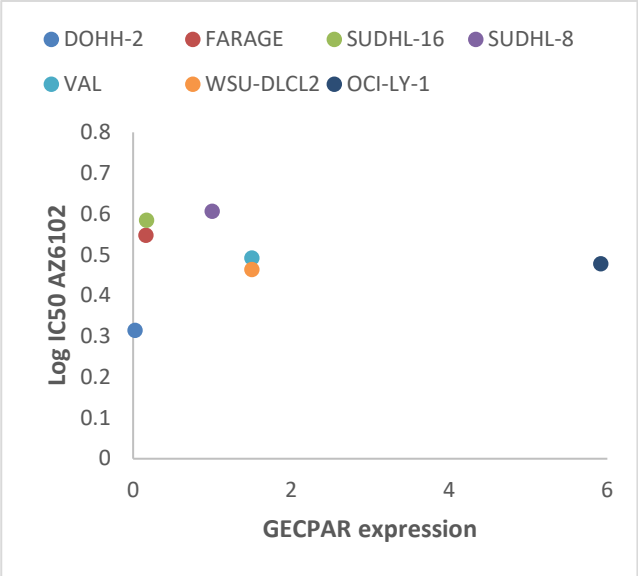


Fig. S7. GECPAR expression and Log IC50 of AZ6102 in 7 GCB-DLBCL cell lines tested for tankyrase inhibitor sensitivity.