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

Sara Napoli,1 Luciano Cascione,1,2 Andrea Rinaldi,2 Filippo Spriano,1 Francesca Guidetti,2 Fangwen Zhang,2 Maria Teresa Cacciapuoti,3 Afua Adjeiwaa Mensah,2 Giulio Sartori,1 Nicolas Munz,1 Mattia Forcato,4 Silvio Bicciato4, Annalisa Chiappella,5 Paola Ghione,6 Olivier Elemento,7,8 Leandro Cerchietti,6 Giorgio Inghirami3 and Francesco Bertoni1,9

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

©2022 Ferrata Storti Foundation. This is an open-access paper. 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; 6Department of Medicine, Division of Hematology and Medical Oncology, Weill Cornell Medicine, New York, NY, 10021, USA; 7Institute for Computational Biomedicine, Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA; 8Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY, USA; 9 Oncology Institute of Southern Switzerland, Bellinzona, Switzerland.

SUPPLEMENTARY INFORMATION

1. Supplementary methods
2. Table S3. Genes commonly enriched in GCB cell lines and GCB-DLBCL patients high vs low GECPAR
3. Table S6 List of essential genes enriched in U2932 depleted of GECPAR or in SUDHL2 overexpressing GECPAR
4. Table S9. siRNAs and LNAs
5. Table S10 Primers
6. Table S11. CHART probes
7. List of external supplementary files
8. Supplementary figure legends
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, SUDDL-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, SU-DHL-8, SU-DHL-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 48h. 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% CO2. 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 (t0). 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 ∆Ct 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 reanalyzed. 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 \( \geq 1 \) 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 70th percentile and low expressor if the GECPAR expression is lower than the 15th 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 “heatmap” 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 Alphalmager (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 | CELL CYCLE |
|-------------------------------------------------|------------|
| TFD1                                           | CCNE2      |
| PCNA                                           | TFD1       |
| HNRNPC                                         | PCNA       |
| UBA52                                          | MCM3       |
| EIF2S2                                         | EIF2S2     |
| 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                                           |            |
| COP55                                          |            |
| DHX9                                           |            |
| CCT2                                           |            |
| SFPQ                                           |            |
| KIF11                                          |            |
| RNPS1                                          |            |
| HCPC1                                          |            |
| MED14                                          |            |
| POLA1                                          |            |
| TCGER1                                         |            |
| ABCE1                                          |            |
| DDX21                                          |            |
| E2F5                                           |            |
| SNRPB                                          |            |
| AP2M1                                          |            |
| POLR2B                                         |            |
| COP56                                          |            |
| 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_DLBCI_ONCOGENIC_SIGNALING_PMID29925955 |
| Enrichment Score (ES) | 0.5168573 |
| Normalized Enrichment Score (NES) | 1.8684261 |
| Nominal p-value | 0 |
| FDR q-value | 0.006865541 |

| Dataset | Limma_GECPARovlogFCbase.rnk |
|---------|-----------------------------|
| Upregulated in class | na_pos |
| GeneSet | ESSENTIAL_GCB_DLBCI_ONCOGENIC_SIGNALING_PMID29925955 |
| Enrichment Score (ES) | 0.8469427 |
| Normalized Enrichment Score (NES) | 1.5805188 |
| Nominal p-value | 0.003157895 |
| FDR q-value | 0.022455128 |
### Table S9. siRNAs and LNAs

| NAME            | SENSE STRAND                      | ANTISENSE STRAND                        |
|-----------------|-----------------------------------|----------------------------------------|
| GECPAR +461 siRNA | ACUGAUCAUGGACCCCAAGUTT            | ACUUUGGUUUGAUCAGUTT                     |
| GECPAR +563 siRNA | GUCCUAUGAGGGAGUACUUTT             | AAUCACUCUCCUCAUAGCCTT                   |
| GL3 siRNA       | CUUACCUGAGACUUCGATT               | UCGAAGUACUCCAGGUAAGTT                   |
| SCR LNA         | ----------------------------------| ---------------------------------------|
| GECPAR +461 LNA | ----------------------------------| ---------------------------------------|
| GECPAR +563 LNA | ----------------------------------| ---------------------------------------|
| GECPAR +489 LNA | ----------------------------------| ---------------------------------------|
| GECPAR +856 LNA | ----------------------------------| ---------------------------------------|

+ LNA modified nucleotide
| NAME                          | SEQUENCE                                      | APPLICATION          |
|-------------------------------|-----------------------------------------------|----------------------|
| GECPAR +545 Fw                | GTGGTCAGCCCTAGACAGTT                        | 3'RACE, RT-PCR       |
| GECPAR +625 Rev               | CAGCATGAACTGCCCCTAAAT                      | 5'RACE, RT-PCR       |
| GECPAR +804 Fw                | ACCTGAGCTGACCTTTATTG                       | 3'RACE, RT-PCR       |
| GECPAR +900 Rev               | GGCTGACCTGGCTTTTCTTTCT                     | 5'RACE, RT-PCR       |
| LOC100132078+3473 Rev         | TGGAAACGACAGCGGAAAG                       | 3'RACE               |
| POU2AF1 ex2 Fw                | ACGGACCCATAGGAGGACTG                      | qRT-PCR              |
| POU2AF1 ex4 Rev               | GCCAGGCTCCTCGCTTCTGTACT                 | qRT-PCR              |
| CREBBP ex9 Fw                 | CATGATCGAGTCTGCTAACA                      | qRT-PCR              |
| CREBBP ex10 Fw                | GGACCTCCGTCTTTTCTCT                     | qRT-PCR              |
| CREBS ex6 Fw                  | AACCCTCAATGCCCAGGATC                    | qRT-PCR              |
| CREB5 ex7 Rev                 | CACAGGGTGCTGAGGATTT                    | qRT-PCR              |
| TLE4 ex12 Fw                  | GGATTTGATCCACACCATA                    | qRT-PCR              |
| TLE4 ex13 Rev                 | TCTGAACATCCTGCGTAACA                   | qRT-PCR              |
| CYLD Fw                       | CAGCAGGTGCCACACTCA                    | qRT-PCR              |
| CYLD Rev                      | ACCCTGAGTCCCTTCTCT                     | qRT-PCR              |
| GAPDH ex3 Fw                  | TCACCGAGGCTGCTTTTTAC                   | qRT-PCR              |
| GAPDH ex4 Rev                 | GGTGGAATCATATGGAAACA                    | qRT-PCR              |
| GAPDH ctr neg Fw              | CGTAGCTCAGGCCATCAGAC                 | qPCR                 |
| GAPDH ctr neg Rev             | GCAGAAGGGAGGAGGAGGAG                     | qPCR                 |
| ALBUMIN ctr neg Fw            | TTGCTAGATGAGGGAGGAC                      | qPCR                 |
| ALBUMIN ctr neg Rev           | TTTAAATCCGCACCCCTCTG                 | qPCR                 |
| BACH2 GECPAR BS Fw            | ATGTGGGGTCCTTTTCTCT                  | qPCR                 |
| BACH2 GECPAR BS Rev           | TGGAAACCGATGGAAAGAT                     | qPCR                 |
| 11q23 GECPAR BS Fw            | AGCCACTCCTGCGAGTT                    | qPCR                 |
| 11q23 GECPAR BS Rev           | GAGTCAGAATGTTGAAGGCCATAA               | qPCR                 |
| TTK GECPAR BS FW              | AATGGGACCATTGAAAGATG                 | qPCR                 |
| TTK GECPAR BS REV             | TCTGAAAGAAATATCACAGAGTG               | qPCR                 |
| ACTL6A GECPAR BS FW           | GACCCAGAAAACAAATCAGAGC                | qPCR                 |
| ACTL6A GECPAR BS REV          | GGGGAACATGAGGAAAAATC                   | qPCR                 |
| ATP11B GECPAR BS FW           | ACAGCTGATGCCGTGAGTCC                   | qPCR                 |
| ATP11B GECPAR BS REV          | GCATTAGCTGAGTGGATTG                  | qPCR                 |
| XRCC4 GECPAR BS FW            | ACAGATGTCCCTTCCACATCTCTGA             | qPCR                 |
| XRCC4 GECPAR BS REV           | ATCCAGCAATCCCCACTCTCG                | qPCR                 |
| MCTP GECPAR BS FW             | TGGTAGATCGTTCTGAGCTCTGCAAATA            | qPCR                 |
| MCTP GECPAR BS REV            | CAAATGGGTTCTCTATGTTGCA               | qPCR                 |
| BET1 GECPAR BS FW             | AAGGGTGGGTGCTATCTCTGA                 | qPCR                 |
| BET1 GECPAR BS REV            | ATGTGCTGATGGCCATTCTTG                | qPCR                 |
| CREB5 GECPAR BS FW            | TTAACCAAGGGCTCTCCACAG                 | qPCR                 |
| CREB5 GECPAR BS REV           | AGAGGTGGACAACCCCAACTG                | qPCR                 |
| ECT2 GECPAR BS FW             | GGAATCTACACAGCCGTTACAA                | qPCR                 |
| ECT2 GECPAR BS REV            | GTAATGAACATCTTCTCTCTTGCTA            | qPCR                 |
| XbaI GECPAR Fw                | GTCTCAGAGGGCAATGATTCAAGACACTTG         | GECPAR cloning      |
| BamHI GECPAR Rev              | CGGGATCCCTAGCTCTTACCAACAGC              | GECPAR cloning      |
| NAME                  | SEQUENCE                                                                 | APPLICATION     |
|-----------------------|---------------------------------------------------------------------------|-----------------|
| GECPAR AS oligo 1     | CCTGGTTTCCAGTTTAGTTGTC                                                  | RNAseH mapping  |
| GECPAR AS oligo 2     | TCCCTGGTTTCCAGTTTAGTTGTC                                                  | RNAseH mapping  |
| GECPAR AS oligo 3     | GTTCTGGTTTCCAGTTTAGTTGTC                                                  | RNAseH mapping  |
| GECPAR AS oligo 4     | GTGTTCTGGTTTCCAGTTTAGTTGTC                                                  | RNAseH mapping  |
| GECPAR AS oligo 5     | CTTGTTCTGGTTTCCAGTTTAGTTGTC                                                  | RNAseH mapping  |
| GECPAR AS oligo 6     | GTTTCTGGAGAGTAAGACGTCG                                                   | RNAseH mapping  |
| GECPAR AS oligo 7     | TTGACCAAACCTGGCTTGTGGA                                                   | RNAseH mapping  |
| GECPAR AS oligo 8     | GGAGCTTGACCAAACCTTGGCTT                                                  | RNAseH mapping  |
| GECPAR AS oligo 9     | CTAGGGGGATTTTCTCTCCTG                                                   | RNAseH mapping  |
| GECPAR AS oligo 10    | AACTTAGGGGATTTTCTCTCCTG                                                  | RNAseH mapping  |
| GECPAR AS oligo 11    | GCTTTTCTGTGGGCTGAGTGG                                                   | RNAseH mapping  |
| GECPAR AS oligo 12    | GGACTGTTTCTGTGGGCTGAG                                                   | RNAseH mapping  |
| GECPAR AS oligo 13    | GCTTTGAGAGTAGAAGAGCGTC                                                   | RNAseH mapping  |
| GECPAR AS oligo 14    | GCTTGACTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATC
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.
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

- Non-GCB: N=43
- GCB: N=31

- p=0.03
- p=0.0085
- p=7.38E-5

(b) Mature lymphoid tumors

(c) Log2 CPM

- bia, follicular, germinal, immature, marginal, mature, pre, pro
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.
**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 t0 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 t0 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 t0 in PDTX-KD transduced with pCDH or pCDH-Gecpar vector. d, Percentage of GFP positive PDTX-KD 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 gene set elements downregulated in GECPAR knock down

| Term name                                                  | stats | GOBP |
|------------------------------------------------------------|-------|------|
| transforming growth factor beta receptor signaling pathway | 1.876e-12 |     |
| cellular response to transforming growth factor beta stimulus | 1.935e-12 |     |
| response to transforming growth factor beta                | 1.988e-10 |     |
| transmembrane receptor protein serine/threonine kinase...  | 3.688e-10 |     |
| negative regulation of transforming growth factor beta rece... | 9.322e-08  |     |
| negative regulation of cellular response to transforming growth factor beta stimulus | 1.779e-08  |     |
| regulation of transforming growth factor beta receptor signaling... | 1.738e-10  |     |
| regulation of cellular response to transforming growth factor stimulus | 1.626e-10  |     |
| cellular response to growth factor stimulus                | 2.158e-07  |     |
| negative regulation of transmembrane receptor protein serine/threonine kinase... | 2.800e-07  |     |
| negative regulation of BMP signaling pathway               | 3.103e-07  |     |
| response to growth factor                                  | 3.132e-07  |     |
| negative regulation of cellular response to growth factor stimulus | 1.258e-07  |     |
| regulation of BMP signaling pathway                        | 5.720e-08  |     |
| enzyme linked receptor protein signaling pathway            | 7.081e-09  |     |
| regulation of transmembrane receptor protein serine/threonine kinase... | 1.254e-09  |     |
| regulation of cellular response to growth factor stimulus | 4.148e-10  |     |
| BMP signaling pathway                                      | 8.312e-10  |     |
| cellular response to endogenous stimulus                   | 9.693e-04  |     |
| response to BMP                                            | 1.250e-04  |     |
| cellular response to BMP stimulus                          | 1.250e-04  |     |
| response to endogenous stimulus                            | 4.267e-10  |     |
| activin receptor signaling pathway                         | 5.286e-02  |     |

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

| Term name                                                  | stats | GOBP |
|------------------------------------------------------------|-------|------|
| positive regulation of leukocyte differentiation           | 5.718e-12 |     |
| positive regulation of myeloid leukocyte differentiation | 7.589e-12 |     |
| positive regulation of hemopoiesis                        | 2.034e-09 |     |
| negative regulation of protein phosphorylation            | 3.511e-10 |     |
| regulation of hemopoiesis                                  | 4.157e-09 |     |
| positive regulation of transcription by RNA polymerase II | 5.728e-04  |     |
| negative regulation of phosphorylation                    | 5.818e-04  |     |
| positive regulation of myeloid cell differentiation        | 6.820e-10  |     |
| response to CAMP                                           | 9.870e-10  |     |
| inactivation of MAPK activity                              | 1.135e-09  |     |
| regulation of leukocyte differentiation                   | 1.400e-10  |     |
| regulation of myeloid leukocyte differentiation            | 1.872e-10  |     |
| negative regulation of phosphate metabolic process        | 2.275e-09  |     |
| negative regulation of phosphorus metabolic process       | 2.298e-09  |     |
| regulation of DNA binding                                  | 2.362e-09  |     |
| regulation of protein phosphorylation                      | 2.469e-10  |     |
| positive regulation of cell differentiation                | 3.076e-10  |     |
| negative regulation of protein modification process       | 3.222e-10  |     |
| response to organophosphorus                               | 3.726e-10  |     |
| positive regulation of transcription, DNA-templated        | 3.750e-10  |     |
| regulation of phosphorylation                             | 4.117e-10  |     |
| response to purine-containing compound                     | 5.320e-10  |     |
| positive regulation of nucleic acid-templated transcription | 5.328e-10  |     |
| positive regulation of RNA biosynthetic process            | 5.755e-10  |     |
| positive regulation of RNA metabolic process              | 8.035e-10  |     |
| positive regulation of RNA metabolic process              | 8.430e-10  |     |

**C** Kegg and Wikipathway classification of RELA DN V1 UP gene set elements downregulated in GECPAR knock down

| Term name         | stats | Kegg and Wikipathway classification |
|-------------------|-------|-----------------------------------|
| MAPK signaling pathway | 1.388e-10 |     |
| Amyotrophic lateral sclerosis | 1.861e-10 |     |
| Rheumatoid arthritis | 3.781e-10 |     |
| Colorectal cancer | 3.781e-10 |     |
| U-LT signaling pathway | 4.402e-07 |     |
| Oxidative metabolism in cancer | 5.772e-10 |     |
| TGF-β Signaling Pathway | 1.804e-10 |     |
| Chromosomal and microsatellite instability in colorectal cancer | 2.587e-10 |     |

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.