haematologica
Journal of The Ferrata Storti Foundation

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

by Sara Napoli, Luciano Cascione, Andrea Rinaldi, Filippo Spriano, Francesca Guidetti, Fangwen Zhang, Maria Teresa Cacciapuoti, Afua Mensah, Giulio Sartori, Nicolas Munz, Mattia Forcato, Silvio Bicciato, Annalisa Chiappella, Paola Ghione, Olivier Elemento, Leandro Cerchietti, Giorgio Inghirami, and Francesco Bertoni

Haematologica 2021 [Epub ahead of print]

Citation: Sara Napoli, Luciano Cascione, Andrea Rinaldi, Filippo Spriano, Francesca Guidetti, Fangwen Zhang, Maria Teresa Cacciapuoti, Afua Mensah, Giulio Sartori, Nicolas Munz, Mattia Forcato, Silvio Bicciato, Annalisa Chiappella, Paola Ghione, Olivier Elemento, Leandro Cerchietti, Giorgio Inghirami, and Francesco Bertoni. Characterization of GECPAR, a noncoding RNA that regulates the transcriptional program of diffuse large B cell lymphoma. Haematologica. 2021; 106:xxx
doi:10.3324/haematol.2020.267096

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.

Short title for the running head

GECPAR, eRNA master regulator of germinal center

Corresponding authors:

-Prof Dr Francesco Bertoni, Institute of Oncology Research, via Vincenzo Vela 6, 6500 Bellinzona, Switzerland. Phone: +41 91 8200 322; e-mail: francesco.bertoni@ior.usi.ch

-Dr Sara Napoli, Institute of Oncology Research, via Vincenzo Vela 6, 6500 Bellinzona, Switzerland. Phone: +41 91 8200 322; e-mail: sara.napoli@ior.usi.ch

CONTRIBUTIONS

S.N. performed the experiments, analyzed and interpreted the data, wrote the manuscript. S.N. and F.B. conceived and supervised the study, and extensively reviewed the manuscript. L.C. performed data mining. A.R. performed next generation sequencing. A.A.M reviewed the manuscript. F.S., F.G., F.Z., G.S. performed experiments. M.F., S.B. provided bioinformatic support. M.T.C., A.C., P.G. and G.I. provided clinical samples data. O.E. and G.I. provided transcriptome data. All authors read and edited the manuscript.

ACKNOWLEDGMENTS

Partially supported by funds from Gelu Foundation (to F. Bertoni).

DISCLOSURE OF CONFLICTS OF INTEREST

Luciano Cascione: travel grant from HTG. Francesco Bertoni: institutional research funds from Acerta, ADC Therapeutics, Bayer AG, Cellestia, CTI Life Sciences, EMD Serono, Helsinn, ImmunoGen, Menarini Ricerche, NEOMED Therapeutics 1, Nordic Nanovector ASA, Oncology Therapeutic Development, PIQUR Therapeutics AG; consultancy fee from Helsinn, Menarini; expert statements provided to HTG; travel grants from Amgen, Astra Zeneca, Jazz Pharmaceuticals, PIQUR Therapeutics AG. The other Authors have nothing to disclose.

Key points

- GECPAR is an enhancer RNA transcribed from the super-enhancer of the *POU2AF1* gene in normal and neoplastic germinal center B cells
- GECPAR has a tumor suppressor activity in DLBCL and is involved in cell proliferation and differentiation
- GECPAR expression has favorable prognostic impact in GCB DLBCL patients

Abstract

Enhancers are regulatory regions of DNA, which play a key role in cell-type specific differentiation and development. Most active enhancers are transcribed into enhancer RNAs (eRNAs) that can regulate transcription of target genes by means of *in cis* as well as *in trans* action. eRNAs stabilize contacts between distal genomic regions and mediate the interaction of DNA with master transcription factors. Here, we characterised an enhancer RNA, GECPAR (GErminal Center Proliferative Adapter RNA), that is specifically transcribed in normal and neoplastic germinal center B-cells from the super-enhancer of *POU2AF1*, a key regulatory gene of the germinal center reaction. Using diffuse large B cell lymphoma cell line models, we demonstrated the tumor suppressor activity of GECPAR, which is mediated via its transcriptional regulation of proliferation and differentiation genes, particularly MYC and the Wnt pathway

INTRODUCTION

Enhancers are regulatory DNA regions that positively drive gene transcription across neighbouring genomic regions spanning many megabases and are characterised by distinct epigenetic features (1, 2): a high ratio of H3K4me1 to H3K4me3; enrichment of H3K27ac, which is deposited by the CREBBP/p300 complex (3); high accessibility to chromatin readers such as bromodomain and extraterminal domain (BET) proteins and transcription factors (TFs). Some enhancers are actively transcribed giving rise to noncoding RNAs called enhancer RNAs (eRNAs) (4). Transcribed enhancers are more acetylated, more enriched of TFs and coactivators, and are also more active in the transactivation of promoters, with which they interact inside 3D structures called enhancer-promoter loops (5). Clusters of enhancers, called super-enhancers (SEs), are strongly transcribed and produce several eRNAs controlling key genes, which regulate cellular development and differentiation (6, 7). eRNAs are crucial components of the regulatory chromatin machinery that controls the expression of key context-specific, protein-coding genes. They usually stabilize multiprotein complexes and constitute a scaffold for DNA loops by enforcing interactions between distant DNA regions, including those located on different chromosomes (8-11). As they lack a poly A tail, their activity is restrained to the site of transcription and they undergo rapid decay. However, polyadenylated long intergenic noncoding RNAs (lincRNAs) also comprise enhancer-derived noncoding transcripts (e-lncRNA) (12), and the stabilization of these eRNAs confers them the capability to act in trans, regulating several distant targets (13).

Individual eRNAs are expressed in a tissue-specific manner. In normal B cells at various stages of differentiation, the expression of noncoding RNAs can more precisely define cellular subsets than protein-coding transcripts (14, 15). In particular, eRNAs are differentially expressed during B cell development and they are associated with protein-coding genes that play an essential role in B cell differentiation.

Diffuse large B cell lymphoma (DLBCL) derives from germinal center (GC) B cells. DLBCL is typically divided into two main subtypes: GC B cell-like (GCB DLBCL), whose transcriptional profile resembles that of light zone GC B cells, and activated B cell-like (ABC DLBCL), whose transcriptome resembles that of plasmablasts (16). However, DLBCLs within each of these subgroups exhibit biological, genetic and transcriptional heterogeneity (17-19). Lineage-specific and growth-dependent transcription factors like BCL6, Myc, NF-κB, p53, and E2F1 can activate specific genetic signatures, depending on the activation of unique subsets of enhancers (20, 21), and contribute to disease heterogeneity. Here, we studied a unique eRNA associated with the *POU2AF1* gene, that we termed GECPAR, for GErminal Center Proliferative Adapter RNA. *POU2AF1* encodes the protein OCA-B, coactivator of OCT2, a B cell specific transcription factor which plays a pivotal role in the regulation of normal and neoplastic GC B cells (22, 23). The SE proximal to *POU2AF1* is the most activated SE in GCB DLBCL(23). Loss of GECPAR correlated with reduced transcription of *TLE4*, which is a negative regulator of LEF1, a Wnt pathway effector protein that in turn regulates also NF-κB. GECPAR loss also increased MYC expression and proliferation of DLBCL cell lines. Conversely, its overexpression impaired cell proliferation. Collectively, our data provides evidence of the nodal role of GECPAR in the regulatory network modulating B cell differentiation and proliferation.

METHODS

Detailed descriptions of experimental methods are included in the Supplementary data.

Human samples, cell lines, small interfering RNA transfection

Established human DLBCL cell lines and Patient Derived Tumor Xenograft Cell lines (PDTX-CL) were grown as previously described (24). All patients providing samples gave written informed consent. Molecular and clinical data acquisition 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. Cell lines were checked for their identity (24). Cells were transfected with siRNAs or LNA using the 4D Nucleofector.

GECPAR cloning and infection into lymphoma cells, RNA-Seq

Cellular lysates were fractionated as previously described (25). For strand-specific qRT-PCR, only the forward primer was used to amplify the antisense strand and only the reverse primer to amplify the sense strand. 5' and 3' Rapid Amplification of cDNA Ends (RACE) was done using Invitrogen RACE System kits. GECPAR was cloned into the pGEM-T vector and subcloned in pCDH-CMV-MCS-EF1-copGFP. pCDH empty backbone or pCDH_GECPAR were transfected in HEK293T, and viral supernatant was then used to infect lymphoma cells. RNA-Seq in cell lines was performed using the NEBNext Ultra II Directional RNA Library Prep.

CHART-seq

CHART enrichment and RNAseH mapping experiments were performed following previously reported protocols (26, 27). The enrichment of CHART signals was determined relative to the oligo controls. Conservative enrichment profiles were determined using the SPP package (28) and MACS (29), as described by Vance and colleagues (30).

RESULTS

The super-enhancer associated with the *POU2AF1* gene locus is transcribed in normal B cells and DLBCL cell lines

Analysis of publicly available RNA-Seq data on RNA polyA+ or poly A- (31) showed that CD20+ cells express a non-polyadenylated portion of the LOC100132078 transcript and also two isoforms of a more abundant antisense transcript (Figure 1a, Figure S1a) . Due to its proximity to the *POU2AF1* gene and its localization in a genomic region with characteristic SE features (highly acetylated, enriched in H3K4me1 but not H3K4me3, based on ENCODE ChIP-Seq data), we hypothesized that it could be an eRNA with particular relevance for GC B cells

To confirm the eRNA length reconstructed in CD20+ cells, we performed 5' and 3' RACE in the DLBCL cell line OCI-LY1. For the 3'-end detection we ran two reactions, with or without the addition of an artificial polyA tail. We identified a transcript lacking a polyA tail and another that was 400 bases longer and naturally polyadenylated. Similarly to the forementioned poly A- transcript reported in CD20+ normal B cells, neither of the transcripts identified in DLBCL cells extended beyond the annotated first exon. The 5' RACE reaction reverse transcribed from exon 4 did not identify a specific 5'-end for exon 1, indicating that the long annotated transcript, LOC100132078, was likely not stable in our model. Conversely, reverse transcribing from exon 1, we identified a 5'-end located at nucleotide +366, mirroring our *in silico* observations for CD20+ normal B cells (Figure 1a-b). We renamed the stabilized portion of LOC100132078 we had sequenced in the OCI-LY1 model as "GECPAR".

GECPAR is mainly chromatin associated and partially polyadenylated

To further characterize the physical characteristics of GECPAR, RNA was extracted from cytoplasm, nucleoplasm and chromatin fractions. In GCB-DLBCL (OCI-LY1 and Karpas422) and ABC-DLBCL (HBL1, U2932) cell lines, GECPAR was transcribed but mostly retained on chromatin, in accordance with reported features of eRNA (6, 7). It was also clearly detected in the nucleoplasm and cytoplasm of OCI-LY1, a cell line with five-fold higher levels of chromatin-associated GECPAR than the other cell lines (Figure 1c).Semiquantitative directional RT-PCR showed that chromatin association was particular to GECPAR since its antisense transcript, when expressed, was more ubiquitously distributed (Figure S1b).

Quantification of *KCNQ1OT1, MALAT1* and beta-actin mRNAs served as a control for chromatin-associated, nuclear and cytosolic RNA, respectively (Figure S1c).

Strong association of a transcript to chromatin usually correlates with its lack of polyadenylation consequent rapid degradation by the RNA exosome (32). To determine if these features were applicable to GECPAR, we assessed its polyadenylation status. The latter was abundant in total transcripts reverse-transcribed using random hexamers, especially in the two GCB-DLBCL cell lines. Conversely, when oligo-dT was used for reverse transcription, GECPAR was clearly detectable in only OCI-LY1, in agreement with the higher abundance of GECPAR in this cell line. (Figure 1d).

GECPAR is predominantly transcribed in GCB DLBCL cell lines and patients

We measured GECPAR transcription by directional qRT-PCR in 22 DLBCL cell lines (GCB, n.= 16; ABC, n.=8). The overlapping antisense transcript was evaluated in parallel as a control. GECPAR was more frequently expressed in GCB- than ABC-DLBCL cell lines (11/16 vs 0/8; P 0.001). In particular, it was expressed at high levels in five (OCI-LY1, OCI-LY1b, OCI-LY8, OCI-LY18, VAL), and at lower levels in six (SU-DHL-4, SU-DHL-6, SU-DHL-16, SU-DHL-8, SU-DHL-10, TOLEDO) GCB DLBCL cell lines. The transcript was barely detectable in the remaining five GCB and in all the eight ABC DLBCL cell lines, while the antisense transcript was more broadly expressed in all cell lines (Figure 2a).

We also evaluated GECPAR level in a total RNA-Seq dataset (33) obtained from specimens derived from normal tonsil (n.=31) and DLBCL patients (GCB, n.=16; ABC, n.18). The transcript was significantly more expressed in normal cells compared to tumor cells, and, in accordance with our cell lines data, it was generally more abundant in GCB than in ABC DLBCL (Figure 2b). The higher GECPAR expression in GCB-DLBCL was confirmed in a validation cohort of 74 patients (GCB, n.=31; non-GCB, n.=43) (GSE145043) (Figure S2a) and in a second one of 350 patients (GCB, n.=183; ABC, n.167) (GSE10846). Variation of GECPAR expression in DLBCL cell lines and patients might be partially explained by its unstable genomic locus (34-36). A focal deletion of the chromosomal region containing the eRNA was observed in 3/737 mature lymphoid tumors (37-41) (Figure S2b).

The normal tonsil derived cells were then subdivided according to B cell maturation stage (42). GECPAR was most highly expressed by centroblasts while naïve B cells expressed the lowest levels. This observation further underlined the specific transcription of GECPAR in GC derived cells. We also analysed a catalogue of murine lncRNA expressed in different developmental stages of B cell maturation (43). Similar to our observations in humans, the murine GECPAR orthologue was mainly expressed in GC B cells, confirming the specific and conserved association of GECPAR with the GC B cell transcriptional program (Figure S2c).

GECPAR expression correlates with cell cycle genes and the GCB DLBCL oncogenic signature

To identify a gene expression signature associated with GECPAR, we focused on the 16 GCB DLBCL cell lines with available expression profiling data (44) and split them in two groups based on the median GECPAR expression. We identified 122 significantly upregulated and 73 downregulated genes (abs log2 fold change ≥ 0.59 and P ≤ 0.05), that could divide GCB DLBCL cell lines into high and low GECPAR expressers (Figure 2c, Table S1). Transcripts that were more expressed in GECPAR high than in GEPCAR low expressers showed a significant enrichment of cell cycle genes and essential cell survival genes, while genes involved in MAPK and PI3K pathways, as well as LEF1 targets were comparatively less enriched (Figure 2d).

When we divided the 16 GCB-DLBCL patient specimens according to GECPAR expression GECPAR-high specimens showed an enrichment of cell cycle genes, particularly the G2M checkpoint as well as genes essential for cell survival (Figure 2e, Table S2). Conversely, LEF1 targets and genes downstream of TGF-β and ATF2 were downregulated in DLBCLs with high GECPAR expression (Figure 2f). Comparison of the genes associated with differential GECPAR expression in cell lines and clinical specimens (Table S3, Figure S2d-e) revealed that common genes were mainly involved in negative regulation of the cell cycle Due to these observations, we hypothesized that GECPAR had an antiproliferative function.

GECPAR exhibits anti-proliferative activity in DLBCL cells

To investigate the putative antiproliferative role of GECPAR we induced degradation of GECPAR using LNA oligonucleotides in VAL, OCI-LY18 and OCI-LY1, three GCB DLBCL cell lines with high level of GECPAR and U2932, an ABC DLBCL with moderate GECPAR expression (Figure S3a). After 24 hours we measured *POU2AF1* mRNA and observed a negligible effect on its expression (Figure S3b).Therefore, despite GECPAR transcription being dependent on activation of the same super-enhancer (Figure S3c-d) needed for

POU2AF1 transcription (Figure S3e), GECPAR itself was not essential for *POU2AF1* transcription. Degradation of GECPAR led to an increase in cell proliferation in all the tested cell lines, suggesting a tumor suppressor function of GECPAR (Figures 3a, S3f-g). To further confirm the antiproliferative activity of GECPAR, we then overexpressed GECPAR in SUDHL2 and OCI-Ly10, two ABC cell lines with low GECPAR levels. The growth of stable GFP positive GECPAR-expressing cells (Fig. S3 h, i) was followed by imaging in real time for 5 days. In both cell models, we measured a significant reduction in proliferation of cells overexpressing GECPAR compared to control infected cells (Figure 3b,S3j). In particular, OCI-Ly10 expressed very intense GFP fluorescence (Fig. S3i) and could grow as monolayer on L-poly-ornithin coated surface allowing monitoring the growth of cells with specific green fluorescence intensity. On the contrary, SUDHL2 tended to form clusters, despite of the L-poly-ornithin coating, and the instrument could hardly discriminate fluorescence from single cells over time. In that case, we could measure the cell growth by phase contrast image analysis, more accurately. The number of total cells and of GFP expressing cells counted at time 0 are reported in Fig. S3k.

As further confirmation, we analyzed GECPAR function also in two ABC (PDTX-KD and PDTX-RRR) and two GCB (PDTX-SS and PDTX-RN) DLBCL PDX models. We confirmed that GECPAR was higher in the two GCB than ABC cases (Figure 3e). Furthermore, we selected the PDX cells with the highest GECPAR expression (PDTX-RN) and we silenced GECPAR by LNA antisense oligonucleotides (Figure S4a). GECPAR silencing increased the proliferation rate also in this model (Figure 3f, S4b). In addition, we overexpressed GECPAR in PDTX-KD cells, which had a very low amount of the transcript. We seeded the cells 24h alter transduction and we followed them (Figure S4c). As for SUDHL2, although we could not monitor their growth along the whole experiment due to their tendency to form clusters, we measured GFP expression by FACS (Figure S4d), GECPAR expression by qRT-PCR (Figure S4e) and cell viability by MTT assay (Figure 3f) after nine days. As observed with ABC DLBCL cell lines, also PDX cells, derived from an ABC DBCL with low GECPAR expression, reduced their proliferation rate after GECPAR overexpression.

GECPAR polarizes cells towards GCB transcriptional program

We performed transcriptional analysis after GECPAR knockdown and overexpression in U2932 and SUDHL2 cells, respectively. Knockdown of GECPAR resulted in 1,099 significantly downregulated and 528 upregulated genes (Table S4), while overexpression of GECPAR led to significant upregulation of 3,152 genes and downregulation of 787 genes (Table S5). Genes upregulated after GECPAR silencing comprised proliferation genes, which were conversely downregulated in GECPAR-overexpressing cells. Further, while U2932, an ABC DLBCL with moderate basal GECPAR expression still presented an enrichment of oncogenic genes typical of ABC DLBCL after GECPAR knockdown (Figure 3d, Table S6, left), the other ABC DLBCL SUDHL2, showed an enrichment of GCB DLBCL genes (Figure 3e, Table S6, right), after GECPAR overexpression. Finally, GECPAR transcription was strongly induced by anti-IgM stimulation of the BCR (Figure S3l). Together, these observations provided further support of GECPAR's role in maintaining the GC transcriptional program.

GECPAR expression has favorable prognostic impact in GCB DLBCL patients

We assessed the expression of GECPAR in 91 DLBCL patients treated with R-CHOP and its potential impact on the clinical outcome. We classified patients in three subgroups: low expressor (below the 15th percentile of the whole population), high expressor (over the 70th percentile), and neutral (in between). High expressor patients had a higher survival probability than low expressor (p value = 0.01) (Fig. 4a). Then, we looked at the high and low expressors based on their cell of origin. As expected by previous analysis (Fig S2a), the high expressors were mainly GCB patients. However, among GCB DLBCL patients, cases with low GECPAR expression had the same risk of death as ABC patients, while the high expressors showed a better outcome (p value=0.03). All together, these observations further sustain the tumor suppressor role we attributed to GECPAR based on our *in vitro* experiments.

GECPAR acts *in trans* regulating cell growth and differentiation by means of Wnt pathway

To identify the genes directly regulated by GECPAR, we performed CHART-Seq in OCI-LY1 and U2932. We identified 4,172 peaks in OCI-LY1 and 692 peaks in U2932 (Figure 5a, Table S7-8). The most prominent peaks were validated in an independent CHART experiment by qRT, confirming the robustness of both the enrichment experiment and downstream analysis (Figure S5b). As an additional control, we measured the levels of transcripts associated with GECPAR binding including *CREBBP, CREB5, TLE4* and *CYLD*. After 24h of GECPAR silencing with LNA oligonucleotides in U2932, the levels of these transcripts were reduced by 50-80% (Figure S5c) and after 72h we noticed a reduction of 50% also in the level of CYLD and TLE4

proteins (Fig.S5d, top). We also measured the increase in protein levels in SUDHL2 and OCI-Ly10 stably overexpressing GECPAR, for TLE4 and CYLD, or CREBBP and CYLD, respectively (Fig.S5d, bottom).

GECPAR capture was done with a set of probes, selected after RNAseH sensitivity assay (Figure S5a). Only peaks called by two different algorithms (MACS and SPP) were taken in account: 4,172 in OCI-LY1 and 692 in U2932 (Figure 5a, Table S7-8). We identified a putative GECPAR binding motif. Among 78 CHARTseq peaks that fell within an interval of 10 kb in both cell lines there was a significant putative GECPAR binding motif (13 matches, *p*-values between 2.15 E-07 and 1.9 E-09) (Figure S5e),

To identify biological processes directly influenced by GECPAR independently of the cell of origin, we analysed 325 genes bound by the eRNA in both OCI-LY1 and U2932. The most significantly enriched classes of genes belonged to the Wnt signaling pathway, cell growth and differentiation (Figure 5b). RNA-Seq data after GECPAR knockdown showed modulation of three pathways associated with development, differentiation and proliferation and known to cross-talk with the Wnt pathway, such as TGF β, NF-κB and MAPK (Figure 5c) (45-47) . Negative regulators of TGF-β pathways including SMAD7, SMURF1 and SMURF2 (Figure S6a) and negative regulators of MAPK signaling, DUSP1, DUSP8 and DUSP10 (Figure S6b), were downregulated, after GECPAR silencing. Some of the downregulated genes belonging to the aforementioned pathways are also negatively regulated by NF-κB (Figure S6c). Notably, WNT and MAPK pathways were also affected in SUDHL2 cells overexpressing GECPAR (Figure 5d).

Intersection of CHARTseq and RNA-Seq data for U2932 cells with GECPAR knockdown identified *MYC* and *PRDM1* among seven genes negatively regulated by GECPAR, indicating that the eRNA influenced both the proliferative capability, reducing MYC, and the terminal differentiation to plasma cells, reducing PRDM1, the genes coding for BLIMP1. Interestingly, 21 direct GECPAR upregulated targets were positively correlated with GECPAR expression also in GCB DLBCL specimens (Figure 54e). Among them there were *KLF6, NOTCH2,* components of BMP, cAMP and TNF-alpha pathways. Strikingly, we also identified *TLE4* (Groucho), which forms a corepressor complex with TCF/LEF1 and recruits HDACs to inhibit transactivation of TCF/LEF1 target genes (48).

Our identification of GECPAR involvement in Wnt signaling prompted us to evaluate the activity of the tankyrase 1/2 (TNKS1/2) inhibitor, AZ6102, that prevents nuclear translocation of beta-catenin. (49). For the four ABC DLBCL cell lines we tested, GECPAR expression and sensitivity to AZ6102 were significantly anti-correlated (Figure 6a), suggesting that expression of GECPAR sensitized cells to Wnt pathway inhibition. All seven GCB DLBCL cell lines tested where equally sensitive to Wnt pathway inhibition (Figure S7). The differential sensitivity to AZ6102 in ABC DLBCLs was not related to tankyrase expression, since protein levels were similar for the four cell lines (Figure 6a). Further, GECPAR overexpressing SUDHL2 cells were more sensitive to Wnt inhibition than the parental control, in terms of cell cycle perturbation. AZ6102 treatment more readily caused G2/M arrest, subG1 accumulation and decreased re-entry in G1 in GECPAR overexpressing cells (Figure 6b).

DISCUSSION

eRNAs have recently started to be recognized as potent modulators of coding gene transcription (50, 51). Here, we provide the first evidence of a lncRNA, transcribed in a SE specifically active during maturation of GC B cells, which plays an antiproliferative role in DLBCL models and is associated with favorable clinical outcome in GCB DLBCL patients.

The lncRNA LOC100132078 was previously annotated as an unknown ncRNA, mainly expressed in lymph nodes and testis (52), and reported among p53-induced eRNAs in breast cancer (53). Since it mapped inside a SE relevant for GC formation (3, 23, 43) and in a site of recurrent genomic instability in lymphoid tumors (34-36), we elucidated its role in DLBCL, the neoplastic counterpart derived from GC B cells. We defined this lncRNA as eRNA according to the main features of this class of ncRNA: it was encoded within a SE; it was a non-polyA chromatin-associated transcript: its expression, highly cell type specific, was dependent on enhancer activation. We also identified a stabilized 970 nucleotide-long transcript, which, based on its expression pattern, we named GECPAR. It was less expressed in DLBCL samples than in normal tonsil B cells and *in vitro* experiments showed an inverse correlation with cell proliferation, suggesting an antitumoral function. The latter was further supported by the association between high GECPAR expression and favorable outcome in GCB DLBCL patients. GECPAR did not seem to act by *in cis* transactivation of the juxtaposed *POU2AF1* gene, which is strongly expressed in GC-derived malignancies (22). Indeed, although GECPAR and *POU2AF1* transcript levels were correlated in cell lines and in clinical specimens, silencing of the eRNA did not strongly impair expression of the coding gene. This is not uncommon and might be due to redundant functions of multiple enhancers that target a given promoter (54).

On the contrary, GECPAR showed *in trans* activity and directly regulated the expression of several transcripts, mainly involved in cell growth and differentiation. These regulated genes were identified as common GECPAR targets in a GCB- and an ABC DLBCL cell line, both of which had constitutively high GECPAR expression.

GECPAR expression was increased after BCR activation, an event that causes transcriptional reprograming of B cells. The exogenous overexpression of GECPAR in an ABC DLBCL cell line confirmed its ability to switch the lymphoma cell towards the GCB DLBCL transcriptional signature.

Nuclear enriched lncRNA regulating transcription *in trans* have been described and they often modulate cell development (43, 55). We propose that GECPAR is used by normal GC B cells to fine-tune the balance between proliferation and differentiation by directly repressing *MYC* and *PRDM1* expression. MYC has a stage-specific role in the GC, particularly in light zone B cells, namely centrocytes, from which GCB DLBCL tumor cells derive. After antigen-driven selection, B cells that still need to improve their antigen affinity can re-enter in the dark zone where they undergo additional cycles of somatic hypermutation. This so-called "cyclic re-entry" is critical for maintaining the GC and is induced by the re-expression of MYC via BCR activation through NF-κB and FOXO1 (56, 57). We propose GECPAR as a key surveillant of this process, as it directly represses MYC in that phase. Termination of the GC reaction is modulated by NF-κB activation downstream of the BCR. It induces IRF4, master regulator of terminal B cell differentiation which in turn activates the plasma cell master regulator BLIMP1, encoded by *PRDM1* (58). GECPAR itself directly represses *PRDM1*, impeding terminal differentiation into plasma blast. In conclusion, GECPAR, which is induced by BCR activation, would retain B cells in the GC light zone, reducing the tendency to re-enter in the dark zone or to exit and differentiate to plasma cells.

GECPAR also reduces B cell proliferation rate and the tendency to differentiate, possibly by directly inducing TLE4, a negative repressor of TCF/LEF1. LEF1 is the key mediator of nuclear Wnt signaling and is important in lymphopoiesis. LEF1 is overexpressed in the nucleus of approximately 40% of DLBCL (59). MYC and Wnt pathway are connected in a positive feedback-loop involving LEF1 (60). GECPAR, which directly inhibited *MYC* expression, indirectly enhanced its anti-proliferative activity via TLE4 that contributed to the arrest of terminal differentiation induced by NF-κB. Indeed, GECPAR expression was inversely correlated with many LEF1 targets, in both DLBCL cell lines and specimens, and some of them were related to NF-κB regulation. Moreover, GECPAR silencing induced upregulation of important NF-κB genes, such as *CARD11, REL and IKBKB*, supporting the link between GECPAR and Wnt/NF-κB crosstalk. Several bidirectional connections between Wnt and NF-κB pathways (45) have been reported in cancer and in particular, in DLBCL (61). We propose GECPAR as an additional layer of control of NF-κB activation in GC B cells, pausing terminal differentiation to plasma blasts.

The greater sensitivity of ABC DLBCL with high GECPAR expression to pharmacological inhibition of Wnt further supports the relationship between GECPAR and Wnt pathway regulation and uncovers alternative therapeutic options for ABC DLBCL patients.

In conclusion, our work describes a novel mechanism of regulation of GC differentiation, which might contribute to DLBCL pathogenesis, and could help in understanding the heterogeneity of this disease.

REFERENCES

1. Loven J, Hoke HA, Lin CY, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell. 2013;153(2):320-334.

2. Whyte WA, Orlando DA, Hnisz D, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013;153(2):307-319.

3. Zhang J, Vlasevska S, Wells VA, et al. The CREBBP Acetyltransferase Is a Haploinsufficient Tumor Suppressor in B-cell Lymphoma. Cancer Discov. 2017;7(3):322-337.

4. Kim TK, Hemberg M, Gray JM, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010;465(7295):182-187.

5. Arnold PR, Wells AD, Li XC. Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. Front Cell Dev Biol. 2019;7:377.

6. Li W, Notani D, Rosenfeld MG. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat Rev Genet. 2016;17(4):207-223.

7. Liu F. Enhancer-derived RNA: A Primer. Genomics Proteomics Bioinformatics. 2017;15(3):196-200.

8. Jiao W, Chen Y, Song H, et al. HPSE enhancer RNA promotes cancer progression through driving chromatin looping and regulating hnRNPU/p300/EGR1/HPSE axis. Oncogene. 2018;37(20):2728-2745.

9. Li W, Notani D, Ma Q, et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature. 2013;498(7455):516-520.

10. Fanucchi S, Shibayama Y, Burd S, et al. Chromosomal contact permits transcription between coregulated genes. Cell. 2013;155(3):606-620.

11. Tan SH, Leong WZ, Ngoc PCT, et al. The enhancer RNA ARIEL activates the oncogenic transcriptional program in T-cell acute lymphoblastic leukemia. Blood. 2019;134(3):239-251.

12. Marques AC, Hughes J, Graham B, et al. Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. Genome Biol. 2013;14(11):R131.

13. Alvarez-Dominguez JR, Knoll M, Gromatzky AA, et al. The Super-Enhancer-Derived alncRNA-EC7/Bloodlinc Potentiates Red Blood Cell Development in trans. Cell Rep. 2017;19(12):2503-2514.

14. Agirre X, Meydan C, Jiang Y, et al. Long non-coding RNAs discriminate the stages and gene regulatory states of human humoral immune response. Nat Commun. 2019;10(1):821.

15. Verma A, Jiang Y, Du W, et al. Transcriptome sequencing reveals thousands of novel long non-coding RNAs in B cell lymphoma. Genome Med. 2015;7:110.

16. Mlynarczyk C, Fontan L, Melnick A. Germinal center-derived lymphomas: The darkest side of humoral immunity. Immunol Rev. 2019;288(1):214-239.

17. Pasqualucci L, Dalla-Favera R. Genetics of diffuse large B-cell lymphoma. Blood. 2018;131(21):2307-2319.

18. Wright GW, Huang DW, Phelan JD, et al. A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications. Cancer Cell. 2020;37(4):551-568.

19. Chapuy B, Stewart C, Dunford AJ, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nat Med. 2018;24(5):679-690.

20. Ceribelli M, Kelly PN, Shaffer AL, et al. Blockade of oncogenic IkappaB kinase activity in diffuse large B-cell lymphoma by bromodomain and extraterminal domain protein inhibitors. Proc Natl Acad Sci U S A. 2014;111(31):11365-11370.

21. Reddy A, Zhang J, Davis NS, et al. Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. Cell. 2017;171(2):481-494.

22. Teitell MA. OCA-B regulation of B-cell development and function. Trends Immunol. 2003;24(10):546-553.

23. Chapuy B, McKeown MR, Lin CY, et al. Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. Cancer Cell. 2013;24(6):777-790.

24. 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.

25. 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.

26. Vance KW. Mapping Long Noncoding RNA Chromatin Occupancy Using Capture Hybridization Analysis of RNA Targets (CHART). Methods Mol Biol. 2017;1468:39-50.

27. Sexton AN, Machyna M, Simon MD. Capture Hybridization Analysis of DNA Targets. Methods Mol Biol. 2016;1480:87-97.

28. Kharchenko PV, Tolstorukov MY, Park PJ. Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat Biotechnol. 2008;26(12):1351-1359.

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

30. 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.

31. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.

32. Pefanis E, Wang J, Rothschild G, et al. RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. Cell. 2015;161(4):774-789.

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

34. Poretti G, Kwee I, Bernasconi B, et al. Chromosome 11q23.1 is an unstable region in B-cell tumor cell lines [Research Support, Non-U.S. Gov't]. Leuk Res. 2011;35(6):808-813.

35. Auer RL, Starczynski J, McElwaine S, et al. Identification of a potential role for POU2AF1 and BTG4 in the deletion of 11q23 in chronic lymphocytic leukemia. Genes Chromosomes Cancer. 2005;43(1):1-10.

36. Zhao C, Inoue J, Imoto I, et al. POU2AF1, an amplification target at 11q23, promotes growth of multiple myeloma cells by directly regulating expression of a B-cell maturation factor, TNFRSF17. Oncogene. 2008;27(1):63-75.

37. 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):2673-2682.

38. Boi M, Rinaldi A, Kwee I, et al. PRDM1/BLIMP1 is commonly inactivated in anaplastic large T-cell lymphoma. Blood. 2013;122(15):2683-2693.

39. 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):1334-1340.

40. Rinaldi A, Kwee I, Young KH, et al. Genome-wide high resolution DNA profiling of hairy cell leukaemia. Br J Haematol. 2013;162(4):566-569.

41. 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):1537-1551.

42. De Silva NS, Klein U. Dynamics of B cells in germinal centres. Nat Rev Immunol. 2015;15(3):137-148.

43. Brazao TF, Johnson JS, Muller J, et al. Long noncoding RNAs in B-cell development and activation. Blood. 2016;128(7):e10-19.

44. 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.

45. Ma B, Hottiger MO. Crosstalk between Wnt/beta-Catenin and NF-kappaB Signaling Pathway during Inflammation. Front Immunol. 2016;7:378.

46. Guo X, Wang XF. Signaling cross-talk between TGF-beta/BMP and other pathways. Cell Res. 2009;19(1):71-88.

47. Grumolato L, Liu G, Haremaki T, et al. beta-Catenin-independent activation of TCF1/LEF1 in human hematopoietic tumor cells through interaction with ATF2 transcription factors. PLoS Genet. 2013;9(8):e1003603.

48. Daniels DL, Weis WI. Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. Nat Struct Mol Biol. 2005;12(4):364-371.

49. Johannes JW, Almeida L, Barlaam B, et al. Pyrimidinone nicotinamide mimetics as selective tankyrase and wnt pathway inhibitors suitable for in vivo pharmacology. ACS Med Chem Lett. 2015;6(3):254-259.

50. Franco HL, Nagari A, Malladi VS, et al. Enhancer transcription reveals subtype-specific gene expression programs controlling breast cancer pathogenesis. Genome Res. 2018;28(2):159-170.

51. Zhang Z, Lee JH, Ruan H, et al. Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer. Nat Commun. 2019;10(1):4562.

52. Fagerberg L, Hallstrom BM, Oksvold P, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics. 2014;13(2):397-406.

53. Leveille N, Melo CA, Rooijers K, et al. Genome-wide profiling of p53-regulated enhancer RNAs uncovers a subset of enhancers controlled by a lncRNA. Nat Commun. 2015;6:6520.

54. Li X, Fu XD. Chromatin-associated RNAs as facilitators of functional genomic interactions. Nat Rev Genet. 2019;20(9):503-519.

55. Guttman M, Donaghey J, Carey BW, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature. 2011;477(7364):295-300.

56. Pasqualucci L. Molecular pathogenesis of germinal center-derived B cell lymphomas. Immunol Rev. 2019;288(1):240-261.

57. Luo W, Weisel F, Shlomchik MJ. B Cell Receptor and CD40 Signaling Are Rewired for Synergistic Induction of the c-Myc Transcription Factor in Germinal Center B Cells. Immunity. 2018;48(2):313-326.

58. Boi M, Zucca E, Inghirami G, et al. PRDM1/BLIMP1: a tumor suppressor gene in B and T cell lymphomas. Leuk Lymphoma. 2015;56(5):1223-1228.

59. Tandon B, Peterson L, Gao J, et al. Nuclear overexpression of lymphoid-enhancer-binding factor 1 identifies chronic lymphocytic leukemia/small lymphocytic lymphoma in small B-cell lymphomas. Mod Pathol. 2011;24(11):1433-1443.

60. Hao YH, Lafita-Navarro MC, Zacharias L, et al. Induction of LEF1 by MYC activates the WNT pathway and maintains cell proliferation. Cell Commun Signal. 2019;17(1):129.

61. Bognar MK, Vincendeau M, Erdmann T, et al. Oncogenic CARMA1 couples NF-kappaB and beta-catenin signaling in diffuse large B-cell lymphomas. Oncogene. 2016;35(32):4269-4281.

Figure legends

Figure 1. POU2AF1 super-enhancer derived transcript in normal B cells and DLBCL cell lines. a, top. Schematic representation of transcripts annotated in chromosome 11q23, between POU2AF1 and BTG genes, according UCSC Genome Browser. Bottom, close-up of LOC100132078 annotated transcript, aligned with CAGE signals on strand plus and minus, transcripts sequenced and reconstructed in RNA poly A + or polyA- from CD20+ cells, and histone marks from ENCODE project. Red lines show positions of exact 5' and 3' ends of GECPAR determined by RACE in OCI-LY1. Arrows indicate position of primers used for 5' and 3' RACE, in particular red arrows primers used for the retrotranscription step. **b**. 5' (left) and 3' (right) RACE performed in OCI-LY1. Numbers on the right of the bands indicate the exact nucleotides corresponding to 5' and 3'ends of GECPAR respect to nucleotide +1, the TSS of annotated LOC100132078. **c**. GECPAR level measured by qRT in subcellular compartments in four DLBCL cell lines, two GCB and two ABC. **d**. GECPAR level measured by qRT in total RNA transcripts or polyadenylated only, in four DLBCL cell lines. Data are mean ± SD of independent determinations. * P <0.05.

Figure 2. GECPAR specific expression in GC B cells and correlation with essential genes. a, top, GECPAR expression in a panel of 22 DLBCL cell lines, 16 GCB and 8 ABC, bottom, expression level of GECPAR antisense transcript, measured as control. **b**, top, Box plots of GECPAR expression quantified by total RNA seq in normal individuals or GCB or ABC DLBCL patients. Bottom, box plots of GECPAR expression in normal individuals stratified for cell of origin. **c**, Heat map of differential gene expression, in GCB DLBCL cell lines dichotomized for GECPAR expression. **d**, Preranked gene set enrichment analysis, in GCB DLBCL cell lines classified for GECPAR expression. **e**, Heat map of differential gene expression, in 16 GCB-DLBCL patients, classified for GECPAR expression. **f**, Preranked gene set enrichment analysis, in DLBCL patients classified for GECPAR expression

Fig 3. GECPAR anti proliferative activity and activation of GCB transcriptional program. a, Proliferation assay after interference with GECPAR by four different LNA antisense oligonucleotides in U2932, VAL and OCI-LY18. Average of three independent experiments * P <0.05, ** P <0.01. **b,** Growth curve of SUDHL2 GFP+ and SUDHL2 Gecpar- GFP+, left, or OCI-Ly10 GFP ^{bright} and SUDHL2 Gecpar- GFP ^{bright}, right, measured by Incucyte. Average of three independent experiments * P <0.05, ** P <0.01**c**, Preranked GSEA of RNAseq data after GECPAR knock down in U2932. **d**, Preranked GSEA of RNAseq data in GECPAR overexpressing SUDHL2 respect to control. **e**, GECPAR expression in four PDX models derived from two ABC and two GCB patients. **f**, **Left,** Proliferation assay in PDX-RN five days after GECPAR knock down. **Right**, Proliferation assay in PDX-KD nine days after GECPAR infection.

Figure 4. GECPAR has a favorable impact on the outcome of GCB DLBCL patients. a, Kaplan-Meier curves of DLBCL patients treated with R-CHOP and stratified for GECPAR expression. **b**, Kaplan-Meier curves of GCB and ABC DLBCL patients treated with R-CHOP and stratified for GECPAR expression.

Fig.5 GECPAR *in trans* transcriptional regulatory function. a. Pipeline of CHART experiment and analysis. **b**, Panther gene ontology classification of 325 GECPAR target genes identified both in OCI-LY1 and U2932 by CHART. **c**, Preranked GSEA of RNAseq data after GECPAR knock down in U2932 **d**, Preranked GSEA of RNAseq data after GECPAR overexpression in SUDHL2 **e,** Top, Venn diagram crossing genes with GECPAR binding detected by CHARTseq and significant expression modulation after GECPAR knocked down, in U2932. Direct downregulated (left) and upregulated (right) GECPAR targets are listed. Bottom, preranked GSEA of direct GECPAR positively regulated targets, in GCB DLBCL patients dichotomized for GECPAR expression

Fig.6 Wnt inhibitor sensitivity in ABC-DLBCL cell lines in dependence of GECPAR expression. a, Anticorrelation between AZ6102 Log IC_{50} and GECPAR expression, left, or tankirase protein level, right, in ABC DLBCL cell lines. **b** Cell cycle analysis in two different GECPAR overexpressing SUDHL2 clones and relative controls, exposed to DMSO or 5 µM of AZ6102 for 48h.

Fig. 1

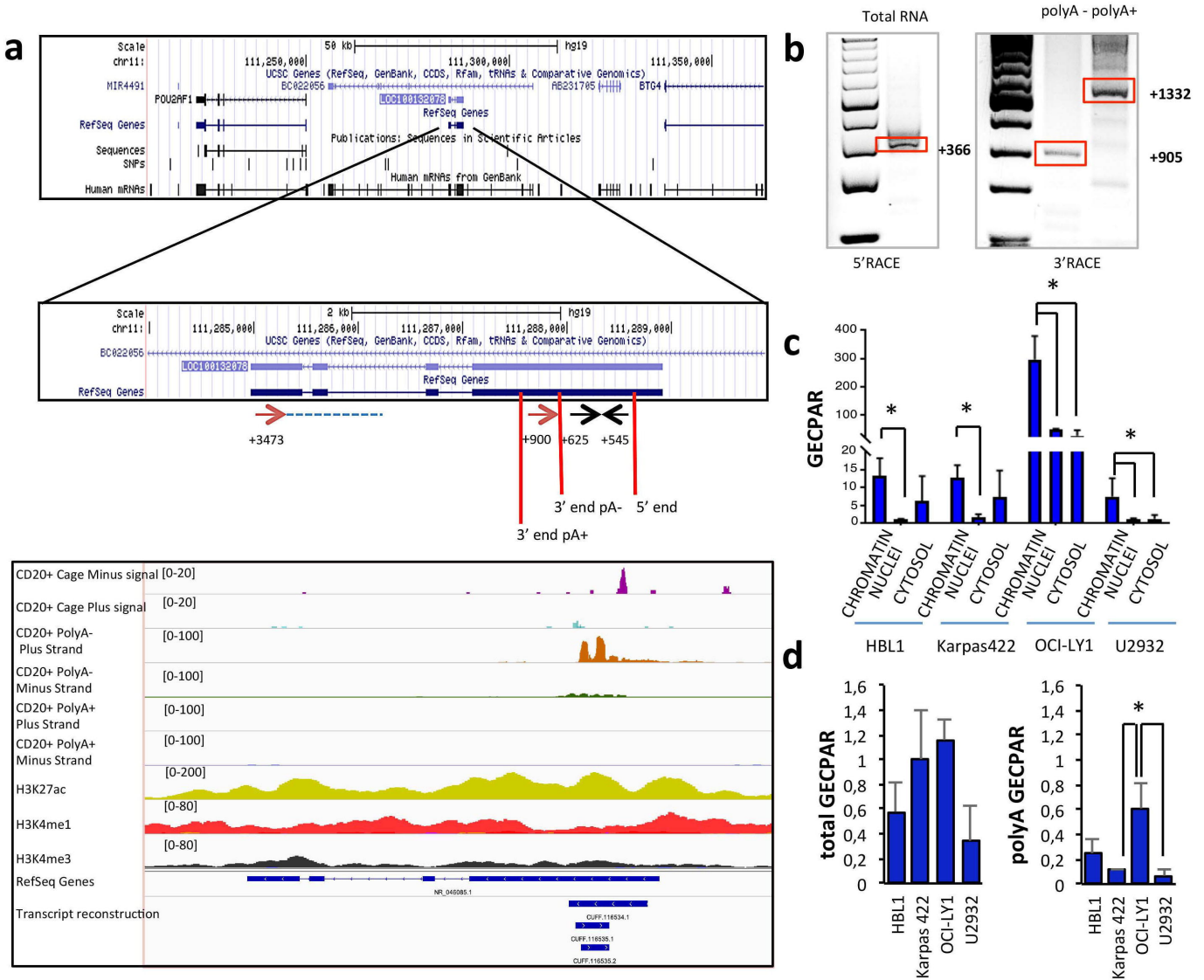
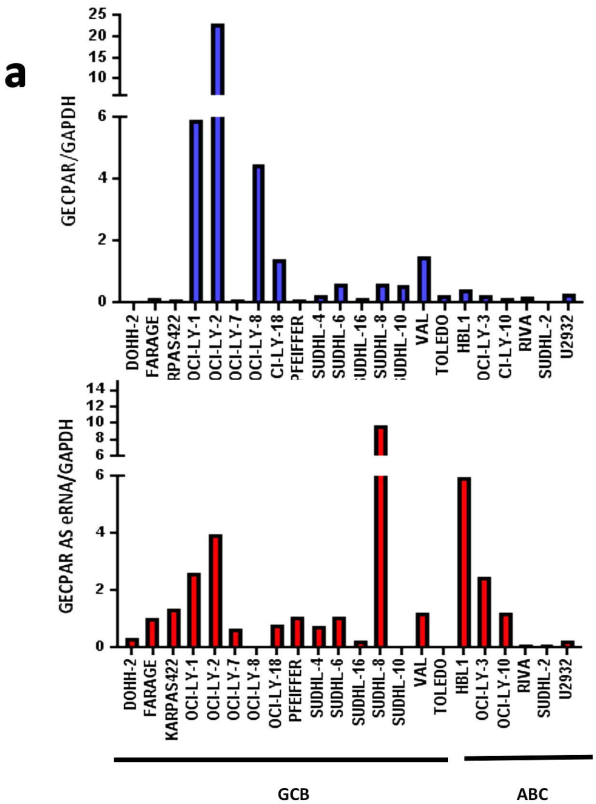
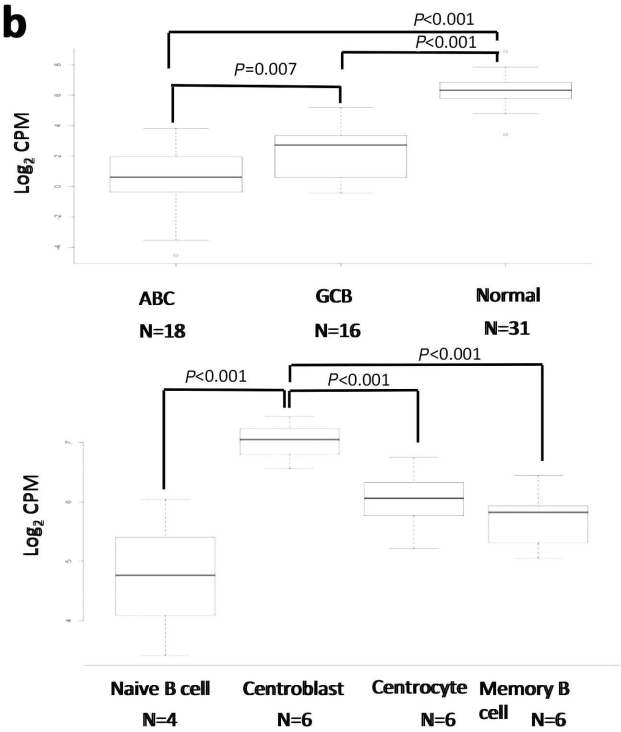
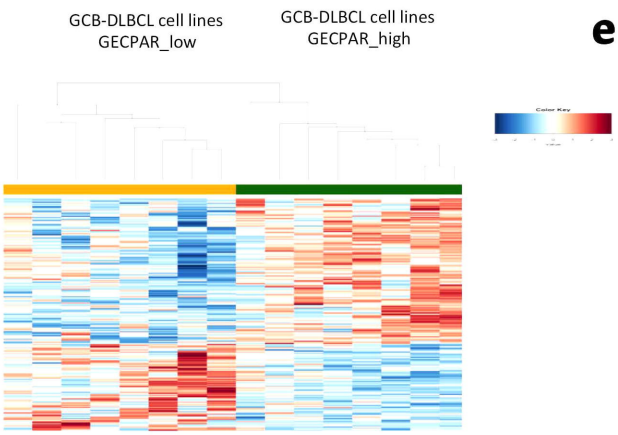
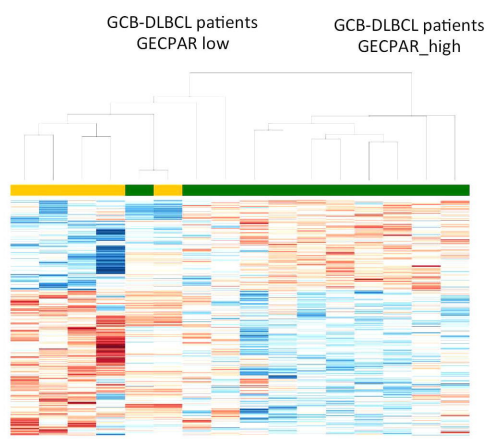

a
Scale
chr11:
50 kb
hg19
111,250,000
111,300,000
111,350,000
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
MIR4491
BC022056
AB231705
BTG4
POU2AF1
LOC100132078
RefSeq Genes
Sequences
SNPs
Publications: Sequences in Scientific Articles
Human mRNAs
Human mRNAs from GenBank
2 kb
111,285,000
111,286,000
111,287,000
111,288,000
111,289,000
+3473
+900
+625
+545
3' end pA-
5' end
3' end pA+
b
Total RNA
polyA - polyA+
+1332
+366
+905
5'RACE
3'RACE
c
GECPAR
CHROMATIN
NUCLEI
CYTOSOL
HBL1
Karpas422
OCI-LY1
U2932
d
total GECPAR
polyA GECPAR
HBL1
Karpas 422
OCI-LY1
U2932
CD20+ Cage Minus signal [0-20]
CD20+ Cage Plus signal [0-20]
CD20+ PolyA- Plus Strand [0-100]
CD20+ PolyA- Minus Strand [0-100]
CD20+ PolyA+ Plus Strand [0-100]
CD20+ PolyA+ Minus Strand [0-100]
H3K27ac [0-200]
H3K4me1 [0-80]
H3K4me3 [0-80]
RefSeq Genes
Transcript reconstruction

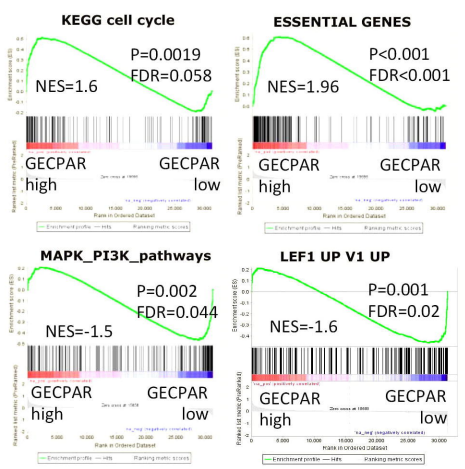
Fig.2


a
GECPAR/GAPDH
GECPAR AS eRNA/GAPDH
DOHH-2
FARAGE
KARPAS422
OCI-LY-1
OCI-LY-2
OCI-LY-7
OCI-LY-8
OCI-LY-18
PFEIFFER
SUDHL-4
SUDHL-6
SUDHL-16
SUDHL-8
SUDHL-10
VAL
TOLEDO
HBL1
OCI-LY-3
OCI-LY-10
RIVA
SUDHL-2
U2932
GCB
ABC


b
Log2 CPM
P<0.001
P<0.001
P=0.007
ABC
N=18
GCB
N=16
Normal
N=31
P<0.001
P<0.001
P<0.001
Naive B cell
N=4
Centroblast
N=6
Centrocyte
N=6
Memory B cell
N=6


c
GCB-DLBCL cell lines
GECPAR_low
GCB-DLBCL cell lines
GECPAR_high
Color Key


e
GCB-DLBCL patients
GECPAR low
GCB-DLBCL patients
GECPAR_high


d
KEGG cell cycle
P=0.0019
FDR=0.058
NES=1.6
ESSENTIAL GENES
P<0.001
FDR<0.001
NES=1.96
MAPK_PI3K_pathways
P=0.002
FDR=0.044
NES=-1.5
LEF1 UP V1 UP
P=0.001
FDR=0.02
NES=-1.6
GECPAR high
GECPAR low

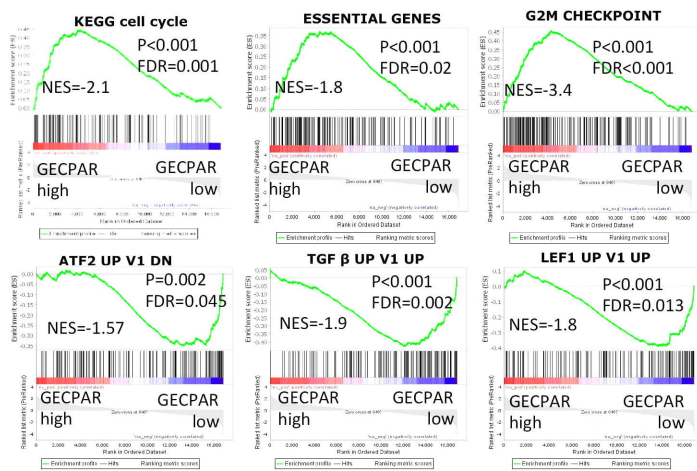

f
KEGG cell cycle
P<0.001
FDR=0.001
NES=-2.1
ESSENTIAL GENES
P<0.001
FDR=0.02
NES=-1.8
G2M CHECKPOINT
P<0.001
FDR<0.001
NES=-3.4
ATF2 UP V1 DN
P=0.002
FDR=0.045
NES=-1.57
TGF β UP V1 UP
P<0.001
FDR=0.002
NES=-1.9
LEF1 UP V1 UP
P<0.001
FDR=0.013
NES=-1.8
GECPAR high
GECPAR low

Fig 3.

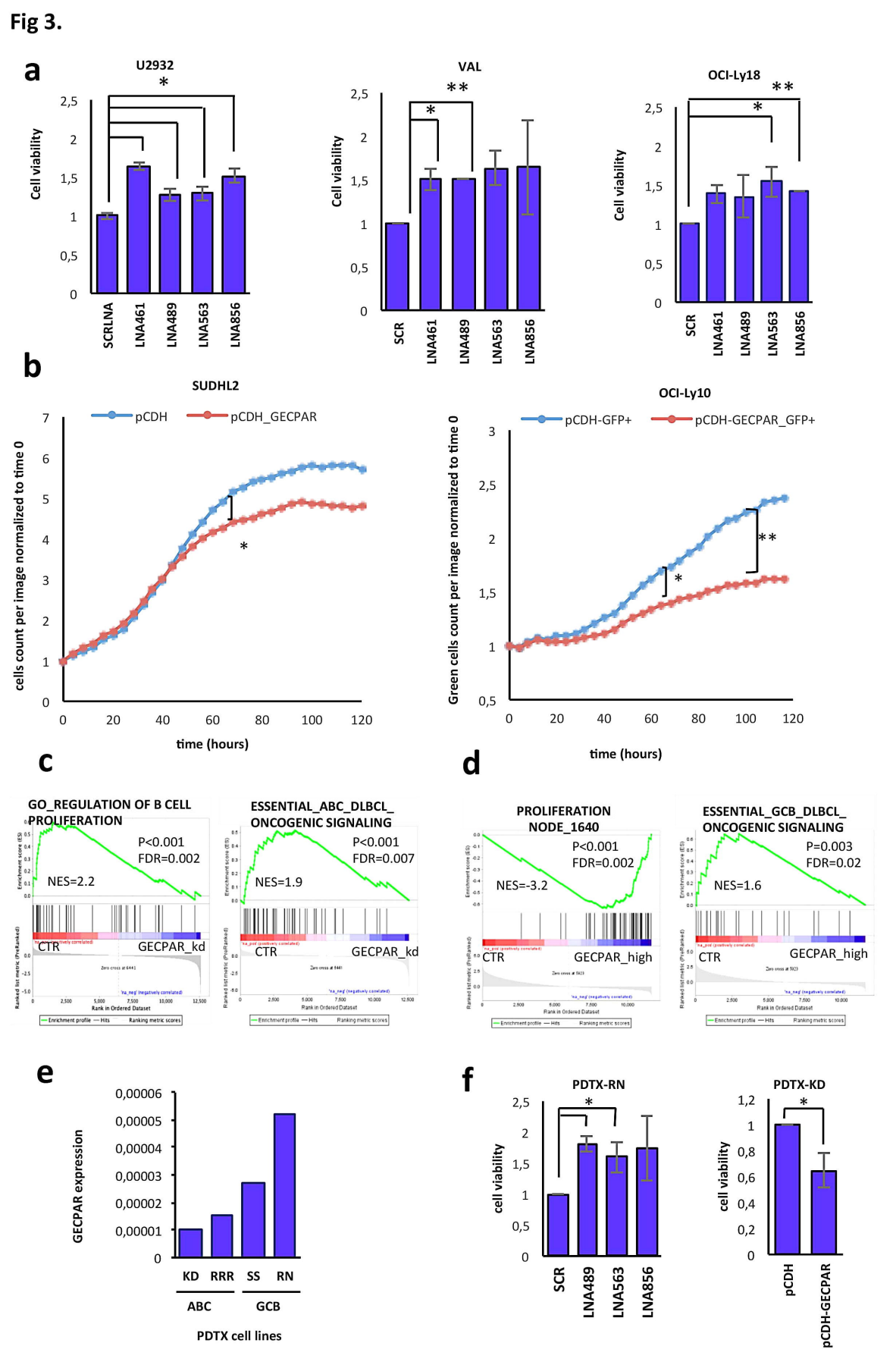
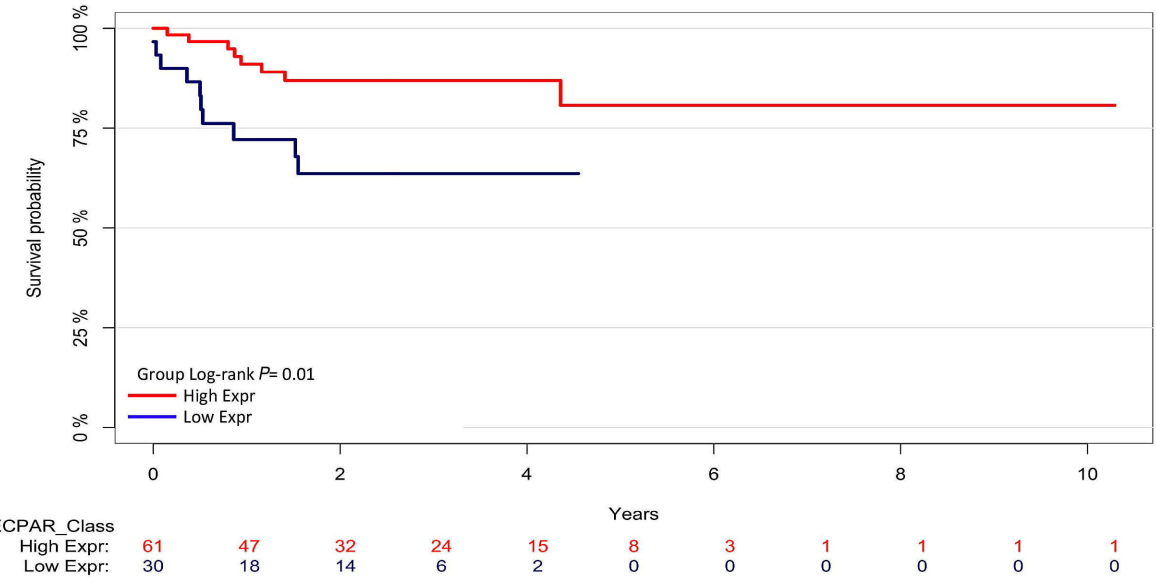

a
U2932
VAL
OCI-Ly18
Cell viability
SCRLNA
SCR
LNA461
LNA489
LNA563
LNA856
b
SUDHL2
pCDH
pCDH_GECPAR
cells count per image normalized to time 0
time (hours)
OCI-Ly10
pCDH-GFP+
pCDH-GECPAR_GFP+
Green cells count per image normalized to time 0
c
GO_REGULATION OF B CELL PROLIFERATION
P<0.001
FDR=0.002
NES=2.2
CTR
GECPAR_kd
ESSENTIAL_ABC_DLBCL_ ONCOGENIC SIGNALING
P<0.001
FDR=0.007
NES=1.9
d
PROLIFERATION NODE_1640
P<0.001
FDR=0.002
NES=-3.2
GECPAR_high
ESSENTIAL_GCB_DLBCL_ ONCOGENIC SIGNALING
P=0.003
FDR=0.02
NES=1.6
Rank in Ordered Dataset
e
GECPAR expression
KD
RRR
SS
RN
ABC
GCB
PDTX cell lines
f
PDTX-RN
PDTX-KD
cell viability
pCDH
pCDH-GECPAR

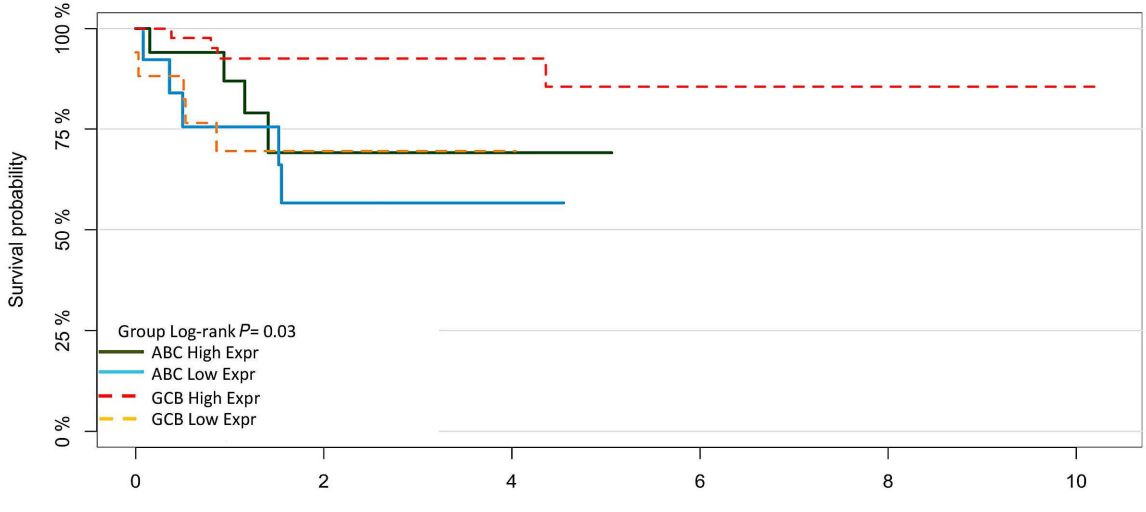
Fig 4.

a



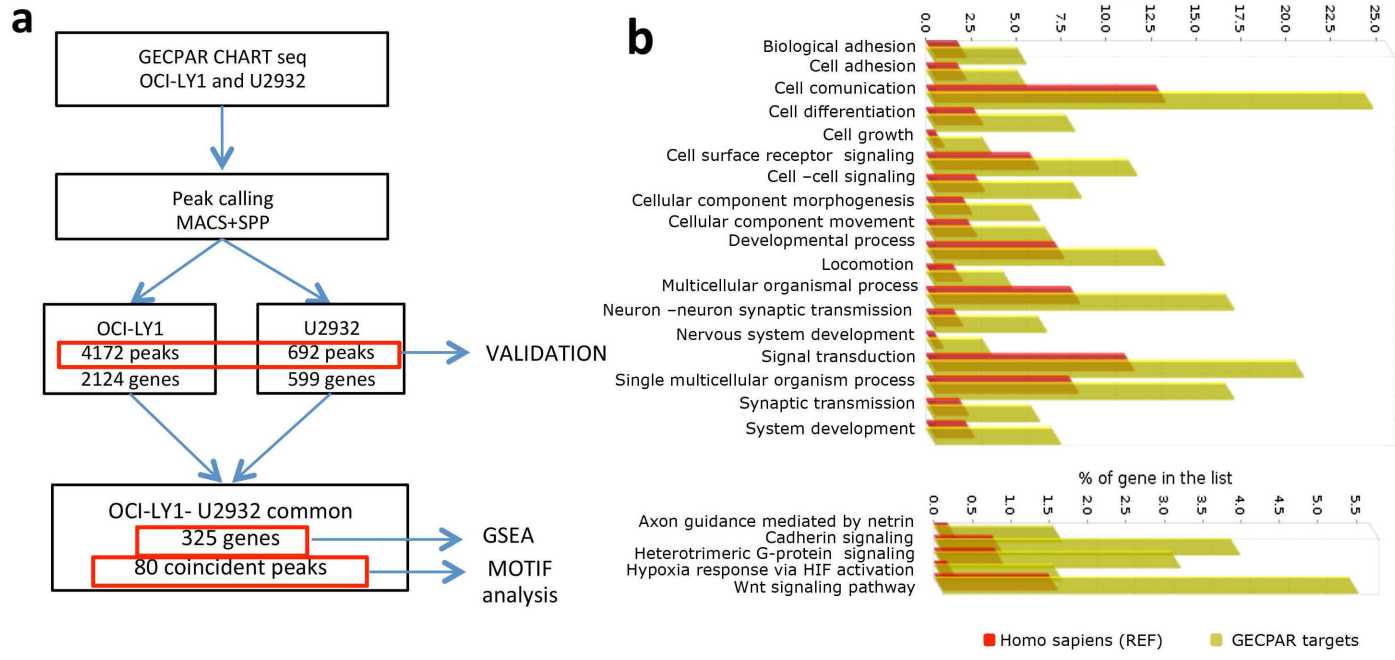
GECPAR_Class											
High Expr:	61	47	32	24	15	8	3	1	1	1	1
Low Expr:	30	18	14	6	2	0	0	0	0	0	0

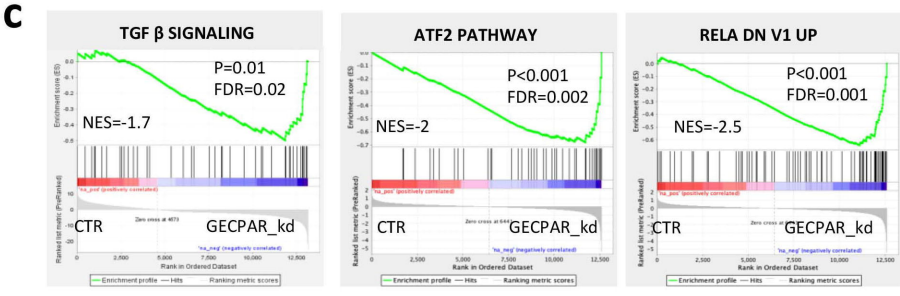
b

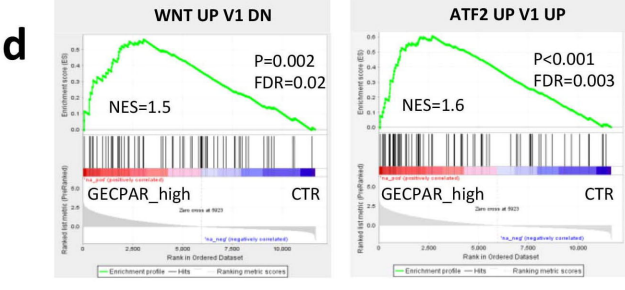


Group											
ABCHigh Expr:	17	12	4	1	1	0	0	0	0	0	0
ABCLow Expr:	13	8	6	4	2	0	0	0	0	0	0
GCBHigh Expr:	44	35	28	23	14	8	3	1	1	1	1
GCBLow Expr:	17	10	8	2	0	0	0	0	0	0	0

Fig 5.


a
GECPAR CHART seq
OCI-LY1 and U2932
Peak calling
MACS+SPP
OCI-LY1
4172 peaks
2124 genes
U2932
692 peaks
599 genes
VALIDATION
OCI-LY1- U2932 common
325 genes
80 coincident peaks
GSEA
MOTIF
analysis
b
% of gene in the list
Biological adhesion
Cell adhesion
Cell comunication
Cell differentiation
Cell growth
Cell surface receptor signaling
Cell –cell signaling
Cellular component morphogenesis
Cellular component movement
Developmental process
Locomotion
Multicellular organismal process
Neuron –neuron synaptic transmission
Nervous system development
Signal transduction
Single multicellular organism process
Synaptic transmission
System development
% of gene in the list
Axon guidance mediated by netrin
Cadherin signaling
Heterotrimeric G-protein signaling
Hypoxia response via HIF activation
Wnt signaling pathway
Homo sapiens (REF)
GECPAR targets


c
TGF β SIGNALING
P=0.01
FDR=0.02
NES=-1.7
CTR
GECPAR_kd
ATF2 PATHWAY
P<0.001
FDR=0.002
NES=-2
CTR
GECPAR_kd
RELA DN V1 UP
P<0.001
FDR=0.001
NES=-2.5
CTR
GECPAR_kd


d
WNT UP V1 DN
P=0.002
FDR=0.02
NES=1.5
GECPAR_high
CTR
ATF2 UP V1 UP
P<0.001
FDR=0.003
NES=1.6
GECPAR_high
CTR

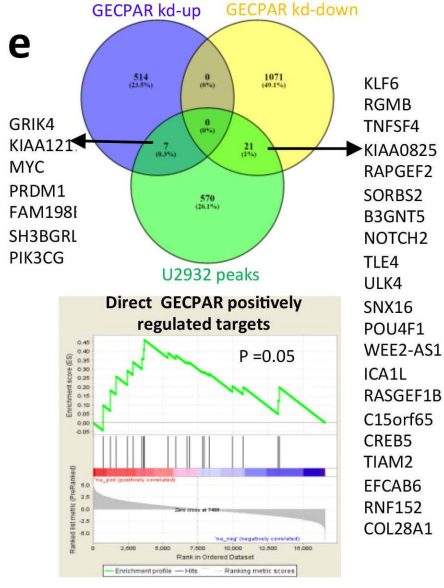
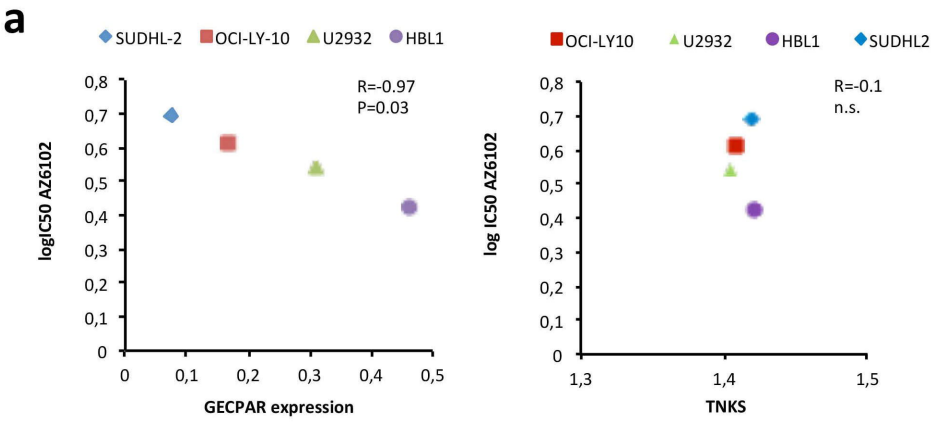
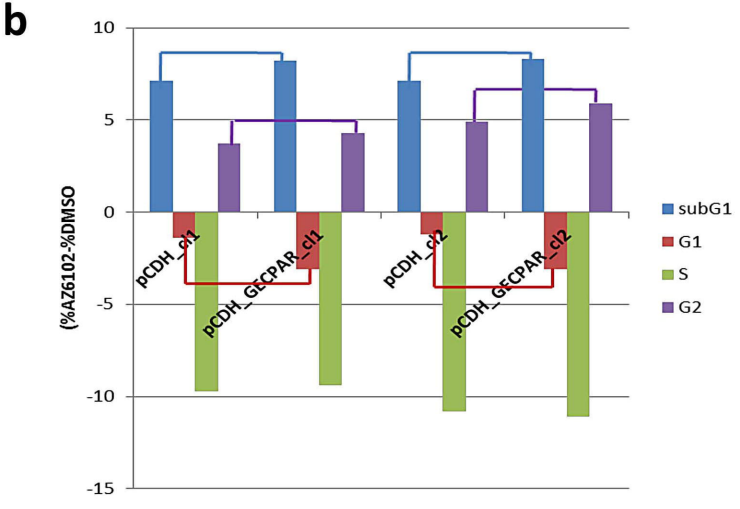

e
GECPAR kd-up
GECPAR kd-down
514
0
1071
7
21
570
U2932 peaks
GRIK4
KIAA121
MYC
PRDM1
FAM198I
SH3BGRL
PIK3CG
KLF6
RGMB
TNFSF4
KIAA0825
RAPGEF2
SORBS2
B3GNT5
NOTCH2
TLE4
ULK4
SNX16
POU4F1
WEE2-AS1
ICA1L
RASGEF1B
C15orf65
CREB5
TIAM2
EFCAB6
RNF152
COL28A1
Direct GECPAR positively regulated targets
P =0.05

Fig 6.


a
SUDHL-2
OCI-LY-10
U2932
HBL1
R=-0.97
P=0.03
logIC50 AZ6102
GECPAR expression
OCI-LY10
U2932
HBL1
SUDHL2
R=-0.1
n.s.
log IC50 AZ6102
TNKS


b
(%AZ6102-%DMSO
pCDH_cl1
pCDH_GECPAR_cl1
pCDH_cl2
pCDH_GECPAR_cl2
subG1
G1
S
G2

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

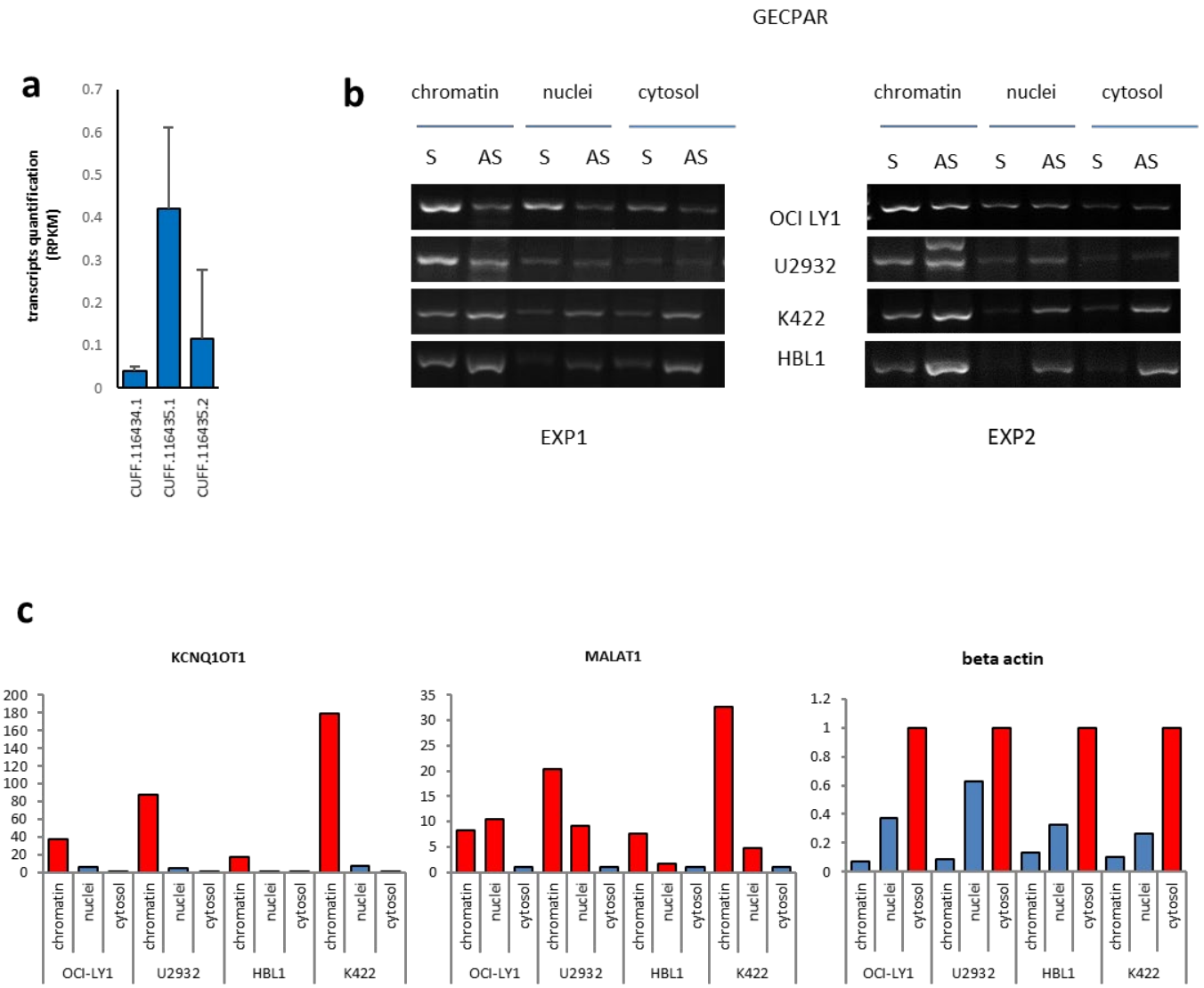
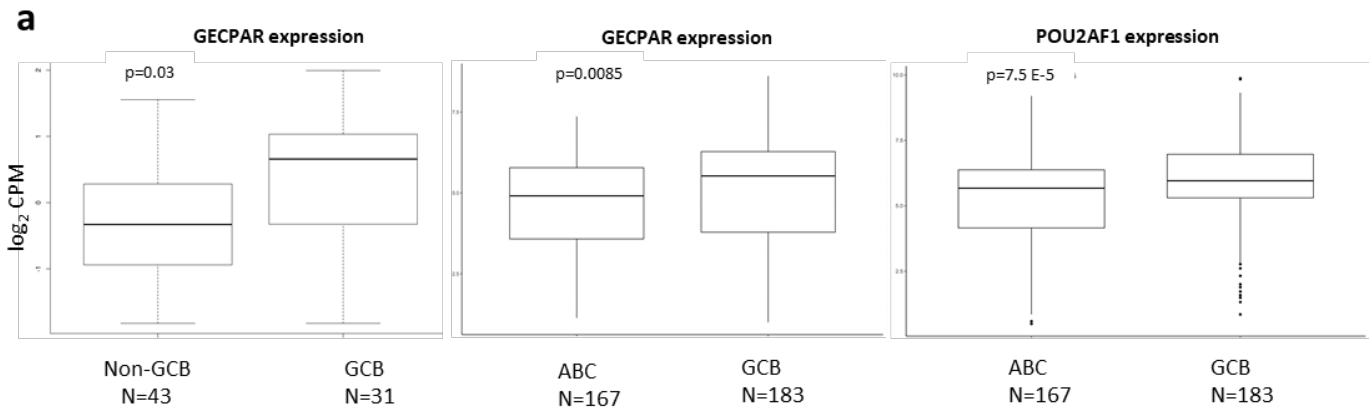
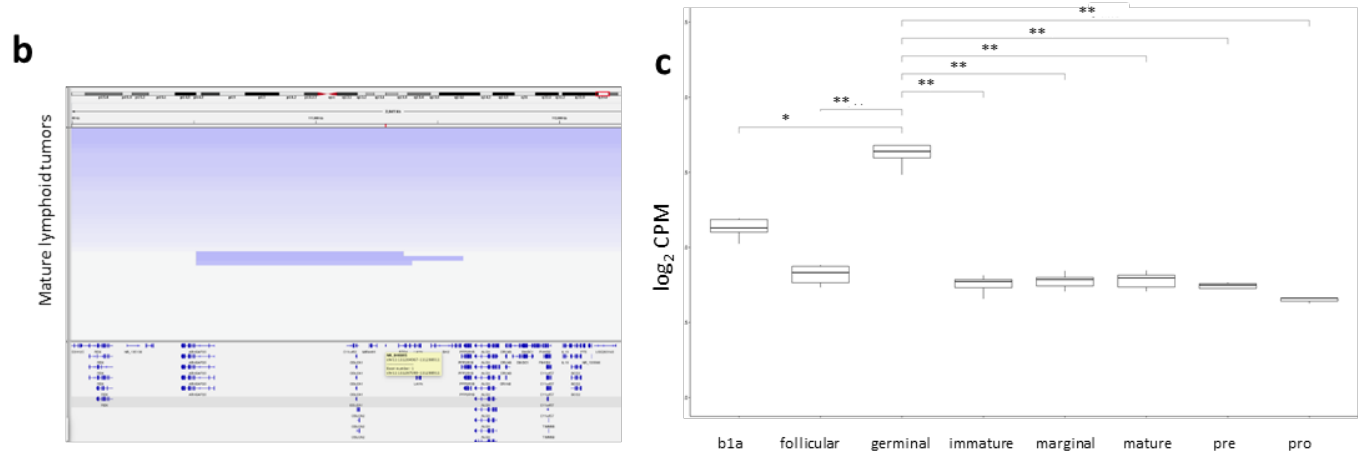


Fig. S1 a, Quantification of De Novo reconstructed transcripts in CD20+ RNAseq in correspondence of LOC100132078 transcript **b**. Directional semiquantitative RT-PCR of two independent experiments of subcellular fractionation of GECPAR and its antisense transcript. **c**, qRT-PCR of KCNQ1OT1 as a positive 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

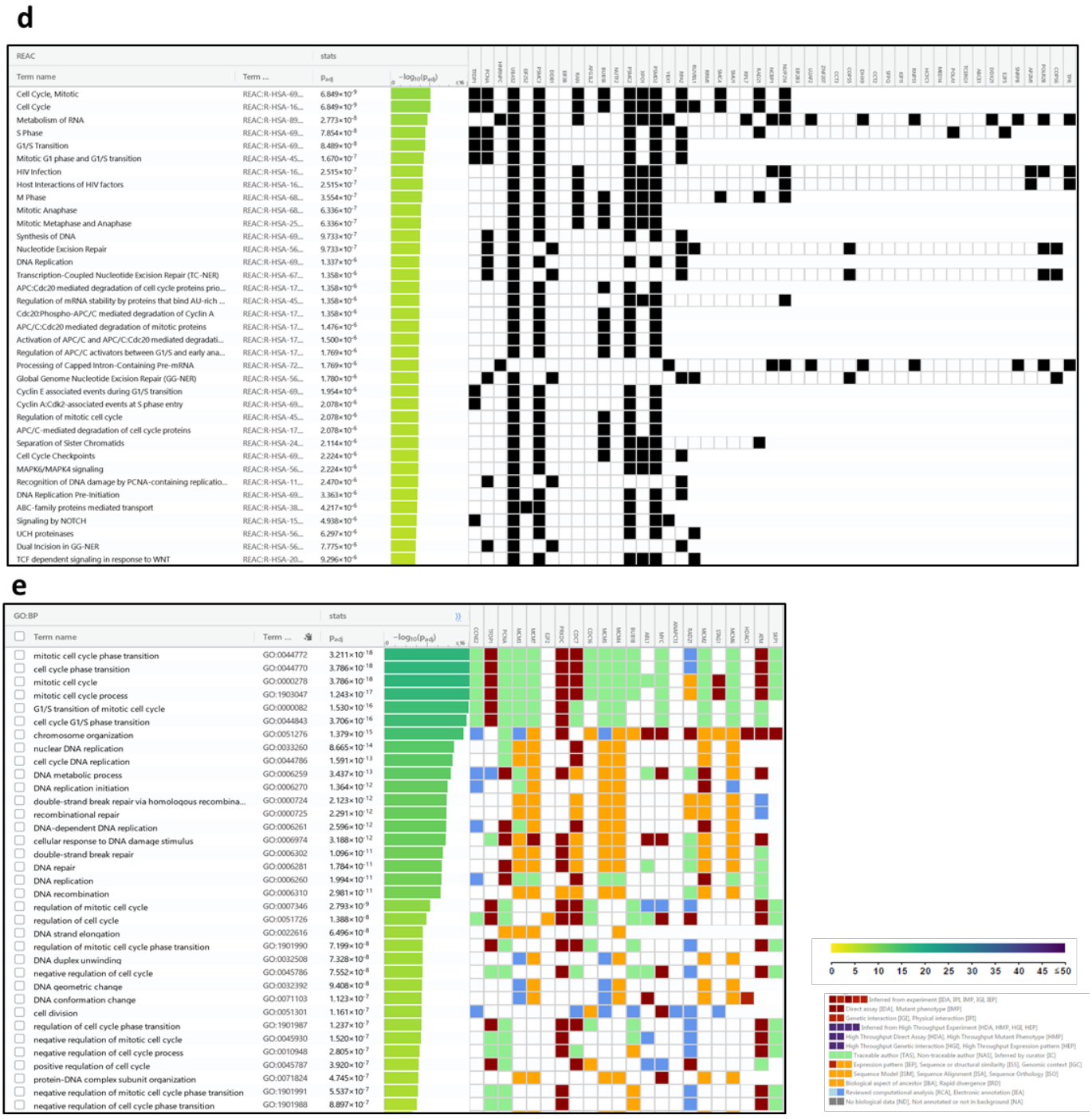
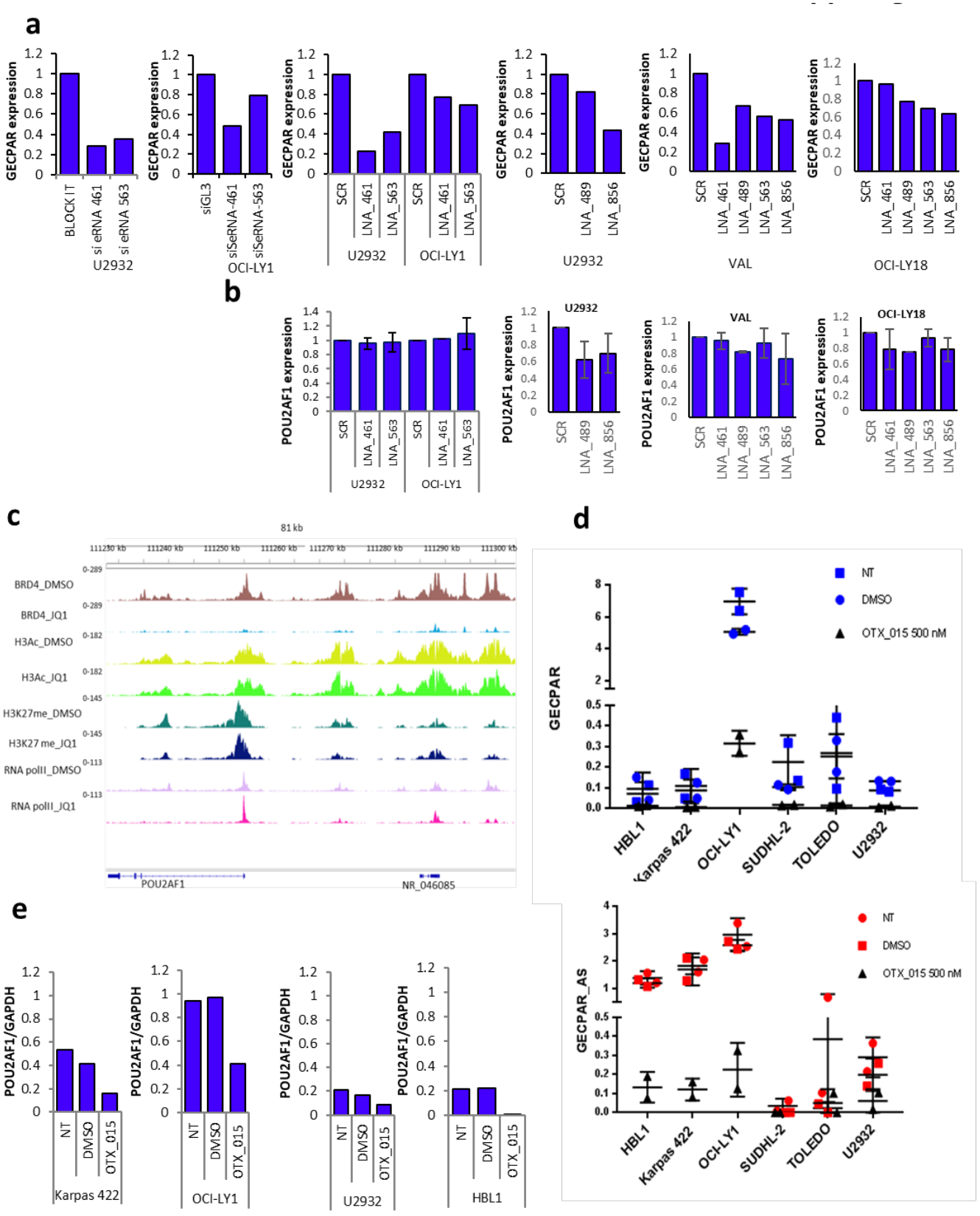


Fig. S2 a, Box plots of GECPAR expression quantified by total RNA seq in GCB or ABC DLBCL patients in a validation cohort (left), box plots of GECPAR (middle) and POU2AF1 (right) expression quantified by microarray in a large validation cohort of GCB or ABC DLBCL patients. **b**, Copy number alterations 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

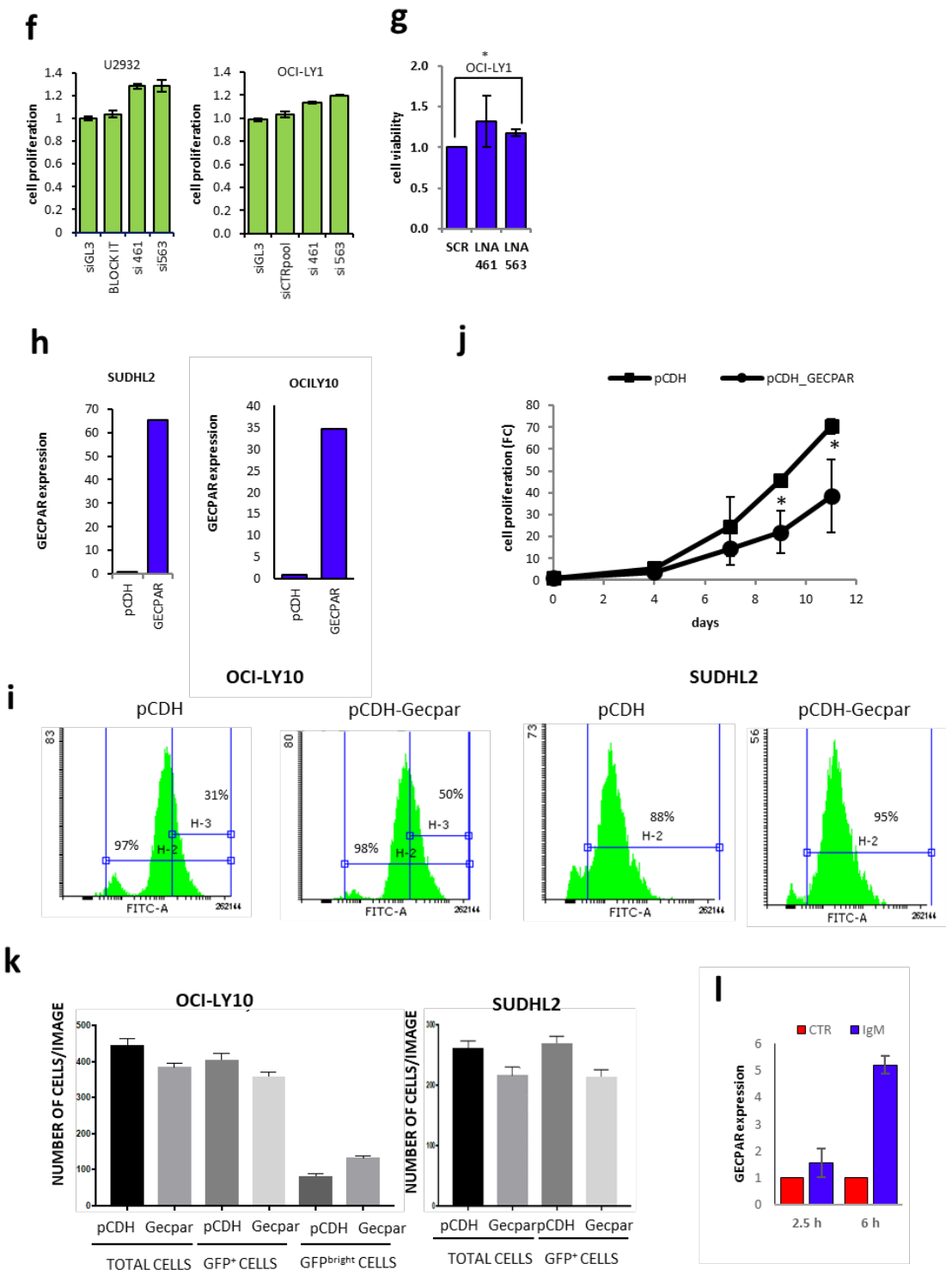


Fig. S3 a,GECPAR expression 24h after interference with two different siRNA in U2932 and OCI-Ly1 and with four different LNA antisense oligonucleotides in U2932, OCI-Ly1, VAL and OCI-Ly18. GECPAR expression is normalized to samples transfected with negative controls. Numerical codes associated to siRNA 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.

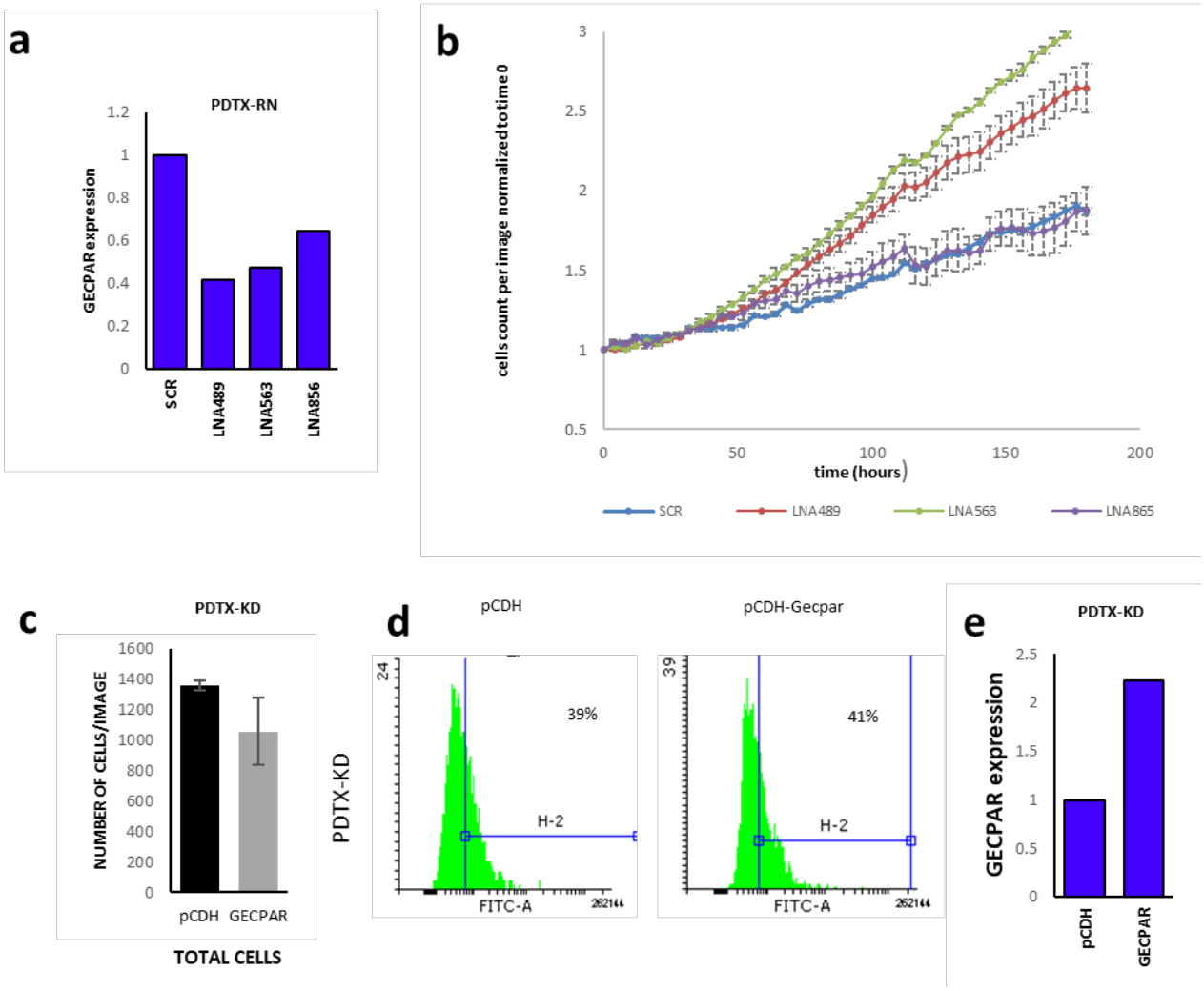


Fig. S4 a, GECPAR expression 48h after interference with four different LNA antisense oligonucleotides in PDTX-RN. **b,** Proliferation assay performed with Incucyte instrument in PDTX-RN nucleofected with negative control (SCR) and three different GECPAR specific LNA antisense oligonucleotides and followed 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.

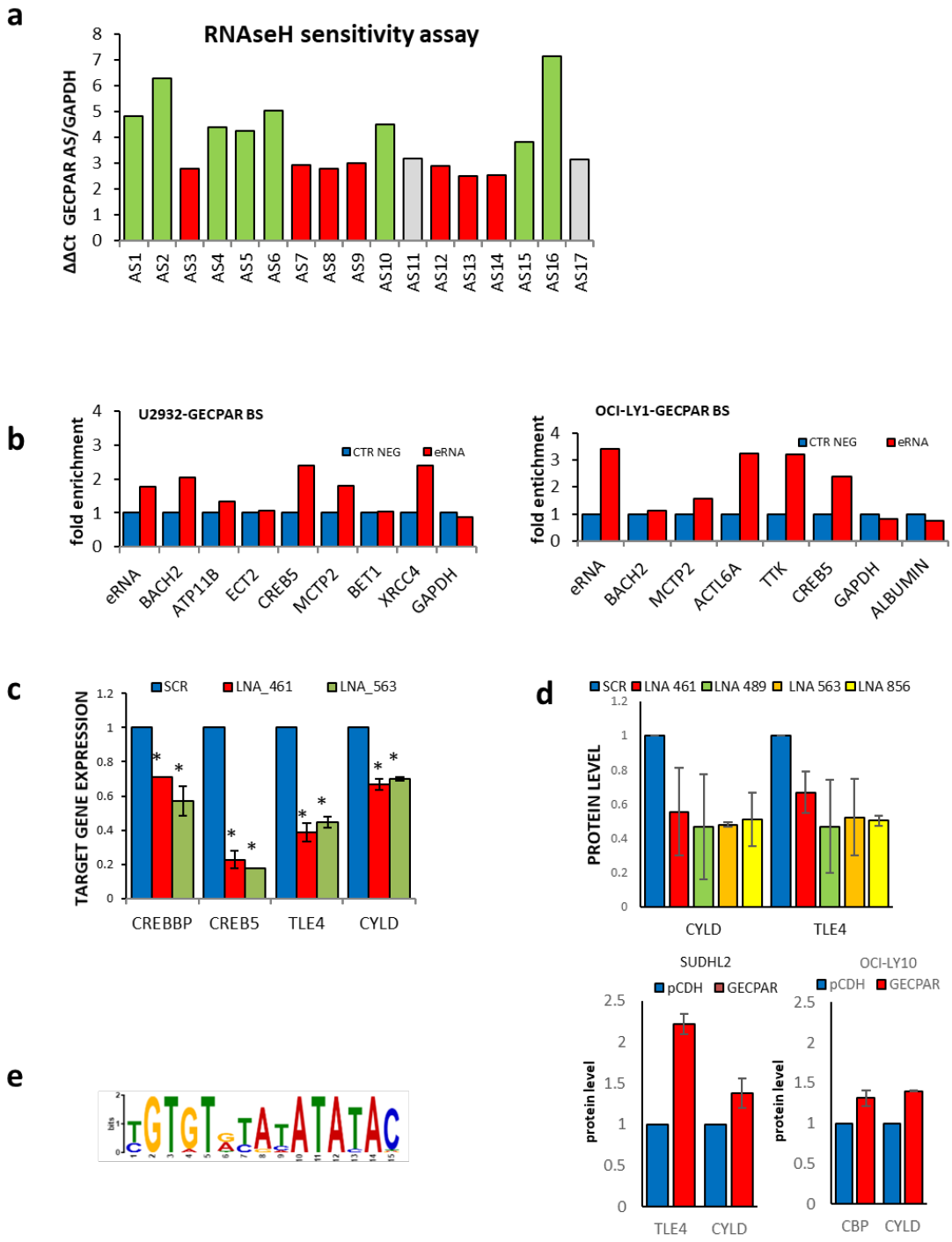
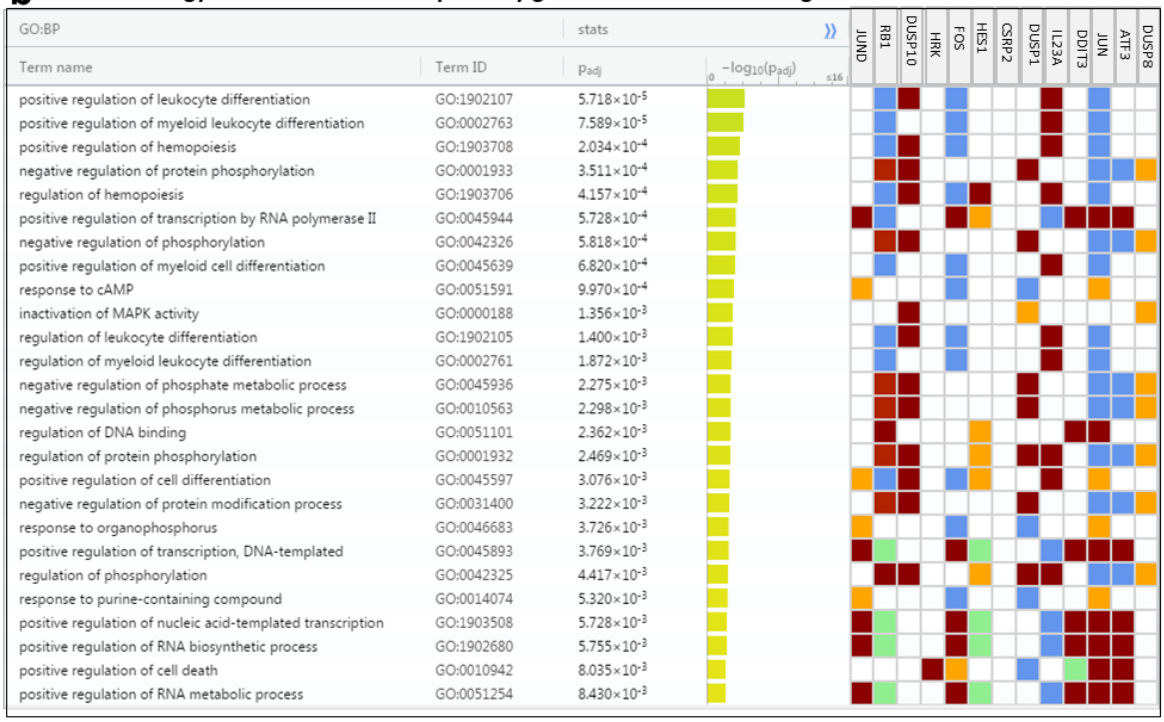


Fig. S5 a, GECPAR level in OCI-LY1 RNA extracted from chromatin after incubation with 17 different antisense oligonucleotides designed to bind GECPAR and treatment with RNAse H. **b**, DNA enrichment after GECPAR pulldown in U2932 (left) or OCI-LY1 (right), concordant with representative peaks from CHARTseq. **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**

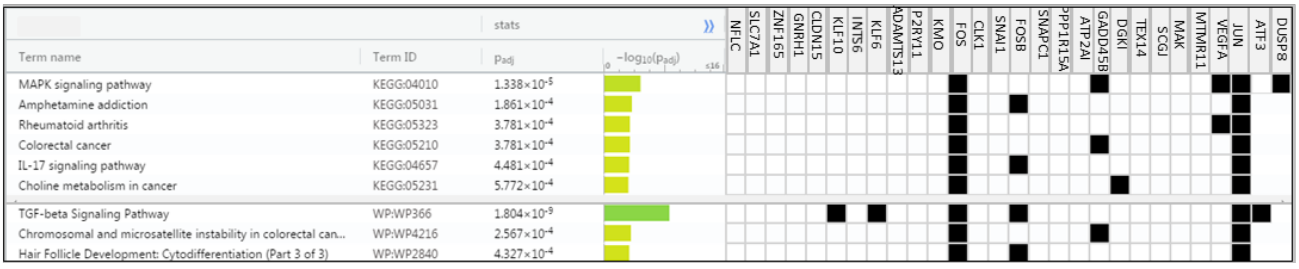
GO:BP		stats	TGFBR1	SMURF1	SMURF2	SMAD7	SKI	KLF10	FNTA	ARID4B	SKIL	LTBP2	SLC20A1	PPP1R15A	TJP1
Term name	Term ID	p_{adj}													
transforming growth factor beta receptor signaling pathway	GO:0007179	1.876×10^{-12}													
cellular response to transforming growth factor beta stimulus	GO:0071560	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.

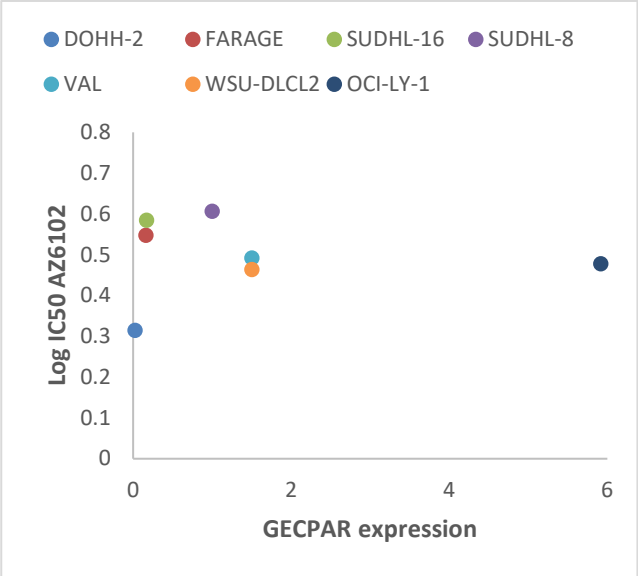


Fig. S7. GECPAR expression and Log IC50 of AZ6102 in 7 GCB-DLBCL cell lines tested for tankyrase inhibitor sensitivity.