Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics

Schmitz, Roland; Young, Ryan M; Ceribelli, Michele; Jhavar, Sameer; Xiao, Wenming; Zhang, Meili; Wright, George; Shaffer, Arthur L; Hodson, Daniel J; Buras, Eric; Liu, Xuelu; Powell, John; Yang, Yandan; Xu, Weihong; Zhao, Hong; Kohlhammer, Holger; Rosenwald, Andreas; Kluin, Philip; Müller-Hermelink, Hans Konrad; Ott, German

DOI:
10.1038/nature11378

Document Version
Peer reviewed version

Citation for published version (Harvard):
Schmitz, R, Young, RM, Ceribelli, M, Jhavar, S, Xiao, W, Zhang, M, Wright, G, Shaffer, AL, Hodson, DJ, Buras, E, Liu, X, Powell, J, Yang, Y, Xu, W, Zhao, H, Kohlhammer, H, Rosenwald, A, Kluin, P, Müller-Hermelink, HK, Ott, G, Gascoyne, RD, Connors, JM, Rimsza, LM, Campo, E, Jaffe, ES, Delabie, J, Smeland, EB, Ogwang, MD, Reynolds, SJ, Fisher, RI, Braziel, RM, Tubbs, RR, Cook, JR, Weisenburger, DD, Chan, WC, Pittaluga, S, Wilson, W, Waldmann, TA, Rowe, M, Mbulaiteye, SM, Rickinson, AB & Staudt, LM 2012, 'Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics', *Nature*, vol. 490, no. 7418, pp. 116-20. https://doi.org/10.1038/nature11378

Burkitt Lymphoma Pathogenesis and Therapeutic Targets from Structural and Functional Genomics

Roland Schmitz[1]*, Ryan M. Young[1]*, Michele Ceribelli[1]*, Sameer Jhavar[1]*, Wenming Xiao[2]*, Meili Zhang[1], George Wright[3], Arthur L. Shaffer[1], Daniel J. Hodson[1], Eric Buras[1], Xuelu Liu[2], John Powell[2], Yandan Yang[1], Weihong Xu[1], Hong Zhao[1], Holger Kohlhammer[1], Andreas Rosenwald[4], Philip Kluin[5], Hans Konrad Müller-Hermelink[4], German Ott[6], Randy D. Gascoyne[7], Joseph M. Connors[7], Lisa M. Rimsza[8], Elias Campo[9], Elaine S. Jaffe[10], Jan Delabie[11], Erlend B. Smeland[12], Martin D. Ogwang[13], Steven J. Reynolds[14], Richard I. Fisher[15], Rita M. Braziel[16], Raymond R. Tubbs[17], James R. Cook[17], Dennis D. Weisenburger[18], Wing C. Chan[18], Stefania Pittaluga[10], Wyndham Wilson[1], Thomas A. Waldmann[1], Martin Rowe[19], Sam M. Mbulaiteye[20], Alan B. Rickinson[19], and Louis M. Staudt[1]

[1]Metabolism Branch Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD, USA
[2]Bioinformatics and Molecular Analysis Section, Division of Computational Bioscience, Center for Information Technology, National Institutes of Health, Bethesda, MD, USA
[3]Biometric Research Branch, DCTD, National Cancer Institute, NIH, Bethesda, MD, USA
[4]Department of Pathology, University of Würzburg, Würzburg, Germany
[5]Department of Pathology and Medical Biology, Groningen University Medical Center, University of Groningen, Groningen, The Netherlands
[6]Department of Clinical Pathology, Robert-Bosch-Krankenhaus, and Dr. Margarete Fischer-Bosch Institute for Clinical Pharmacology, 70376 Stuttgart, Germany
[7]British Columbia Cancer Agency, Vancouver, British Columbia, Canada
[8]Department of Pathology, University of Arizona, Tucson, AZ, USA
[9]Hospital Clinic, University of Barcelona, Barcelona, Spain
[10]Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland 20892, MD, USA
[11]Pathology Clinic, Rikshospitalet University Hospital, Oslo, Norway
[12]Institute for Cancer Research, Rikshospitalet University Hospital and Center for Cancer Biomedicine, Faculty Division of the Norwegian Radium Hospital, University of Oslo, Oslo, Norway
[13]St. Mary's Hospital Lacor, Gulu, Uganda
[14]Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA
[15]James P. Wilmot Cancer Center, University of Rochester School of Medicine, Rochester, NY, USA
[16]Oregon Health and Science University, Portland, OR, USA
[17]Cleveland Clinic Pathology and Laboratory Medicine Institute, Cleveland, OH, USA
[18]Departments of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE, USA
[19]School of Cancer Sciences, Birmingham Cancer Research UK Centre, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.
[20]Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Department of Health and Human Services, Rockville, MD, USA
*RS, RMY, MC, SJ and WX contributed equally

Address correspondence to:
Louis M. Staudt, M.D., Ph.D.
Metabolism Branch, CCR, NCI
Building 10, Room 4N114, NIH
9000 Rockville Pike, Bethesda, MD USA
e-mail: lstaudt@mail.nih.gov

Burkitt lymphoma (BL) can often be cured by intensive chemotherapy, but the toxicity of such therapy precludes its use in the elderly and in patients with endemic BL in developing countries, necessitating new strategies[1]. The normal germinal center B cell is the presumed cell of origin for both BL and diffuse large B cell lymphoma (DLBCL), yet gene expression analysis suggests that these malignancies may utilize different oncogenic pathways[2]. BL is subdivided into a sporadic subtype (sBL) that is diagnosed in developed countries, the EBV-associated endemic subtype (eBL), and an HIV-associated subtype (hivBL), but it is unclear whether these subtypes employ similar or divergent oncogenic mechanisms. Here we used high throughput RNA sequencing and RNA interference screening to discover essential regulatory pathways in BL that cooperate with MYC, the defining oncogene of this cancer. In 70% of sBL cases, mutations affecting the transcription factor TCF3 (E2A) or its negative regulator ID3 fostered TCF3 dependency. TCF3 activated the pro-survival PI(3) kinase pathway in BL, in part by augmenting tonic B cell receptor signaling. In 38% of sBL cases, oncogenic *CCND3* mutations produced highly stable cyclin D3 isoforms that drive cell cycle progression. These findings suggest opportunities to improve therapy for patients with BL.

We performed RNA resequencing (RNA-seq) on 28 sBL patient biopsies and 13 BL cell lines and reanalyzed published RNA-seq data from 52 germinal center B cell-like (GCB) DLBCL cases and 28 activated B cell-like (ABC) DLBCL cases[3]. Elimination of known single nucleotide polymorphisms left a set of putative SNVs (pSNVs) of which 95% (495/518) were confirmed by Sanger sequencing (Supplementary Tables 1 and 2).

Mutations in many genes were more frequent in BL than in DLBCL, including *MYC* as well as many not previously implicated in this lymphoma subtype (Fig. 1; Supplementary Fig. 1a; Supplementary Table 1). Conversely, recurrently mutated genes in DLBCL[3-7] (*EZH2*, *SGK1*, *BCL2*, *CD79B*, *MYD88*) were rarely if ever mutated in BL. Several genes were mutated in BL and DLBCL (*TP53*, *GNA13*, *MKI67*, *CCND3*), although *TP53* mutations were more common in BL (Fig. 1; Supplementary Fig. 1b). This mutational survey suggests that BL is pathogenetically distinct from other germinal center-derived lymphomas.

Highly recurrent mutations in *TCF3* and its negative regulator *ID3* suggested that TCF3 plays a central role in BL pathogenesis, as it does in normal B cell development by

regulating the transcription of immunoglobulin and other B cell-restricted genes through E-box motifs[8,9]. *ID3* and/or *TCF3* mutations were present in sBL, hivBL, and eBL, in 70%, 67%, and 40% of samples, respectively, but these mutations were rare in other lymphoid cancers (Fig. 2a; Supplementary Table 3). In sBL, *ID3* mutations (58%) were more common than *TCF3* mutations (11%), and some tumors had mutations in both genes (13%). *ID3* mutations were usually bi-allelic whereas *TCF3* mutations were often mono-allelic (Supplementary Fig. 2a; Supplementary Table 3). A somatic origin was confirmed for 14 *ID3* mutations and 4 *TCF3* mutations (Supplementary Table 3). All *TCF3* mutations affected the basic helix-loop-helix (B-HLH) DNA binding and dimerization domain of one *TCF3* splice isoform (E47) but not the other (E12), suggesting a non-redundant role for E47 in BL pathogenesis. In cases with *TCF3* mutations, E47 was more highly expressed than E12, suggesting gain-of-function (Supplementary Fig. 2b).

Most *TCF3* mutations target 4 evolutionarily conserved residues in the B-HLH region (N551K, V557E/G, D561E/V/N, M572K; Supplementary Fig. 3a). The most common mutations affect V557 and D561, which are adjacent in the crystal structure and face away from DNA, suggesting a role in intermolecular interactions (Fig. 2b). The B-HLH domain may be distorted by mutations affecting M572 and L597, which are neighboring residues in crystal structure. N551 is a DNA contact residue[10], suggesting that N551K could alter TCF3 DNA binding.

A variety of nonsense and frameshift mutations inactivate *ID3* in BL tumors, suggesting a tumor suppressor mechanism (Supplementary Fig. 3b). Many missense mutations target the conserved loop region of ID3, potentially changing the tertiary structure of the B-HLH domain and impairing its ability to inhibit TCF3[11] (Fig. 2c). Numerous *ID3* missense mutations affect the HLH domain away from the interface of the two helices, possibly altering TCF3 interaction. Other mutations disrupt an *ID3* splice donor and force a cryptic splice donor to be used, thereby deleting residues V82-Q100 (Supplementary Figs. 2c, 3b).

An RNA interference screen revealed *TCF3* to be an essential gene in BL lines (Supplementary Fig. 2d, Supplementary Table 4), supporting the notion that the *TCF3* and *ID3* mutations in BL promote TCF3 action. TCF3 knockdown caused a time-

dependent toxicity in all BL lines, irrespective of *ID3/TCF3* mutations, but had no effect on DLBCL lines (Fig. 2d; Supplementary Figs. 2e, 4a, 4b). Wild-type TCF3 rescued BL lines from shTCF3 toxicity as could the TCF3 mutants, suggesting that they are not loss-of-function (Supplementary Fig. 2e). Introduction of wild-type ID3 into BL lines with *ID3* mutations was lethal, but BL-derived ID3 mutants had less or no toxicity, consistent with a tumor suppressor mechanism (Fig. 2e; Supplementary Figs. 2g, 4c).

The common ID3 and TCF3 mutants diminished their inhibitory heterodimerization. V557E and D561E TCF3 did not associate well with ID3 and failed to stabilize ID3 protein expression, unlike wild-type TCF3 (Fig. 2f; Supplementary Fig. 2h, 4d). Likewise, the ID3 mutant proteins were expressed less well than wild type ID3 and were less able to co-immunoprecipitate TCF3 (Fig. 2g; Supplementary Fig. 4e). However, N551K TCF3 behaved like wild-type TCF3 in these dimerization assays, suggesting a distinct mechanism.

We next used ChIP-seq analysis to gauge the ability of the

TCF3 mutants to interact with chromatin genome-wide. We engineered 2 BL lines to express biotinylated wild-type or mutant TCF3 isoforms ("TCF3-Biotag"; see Methods), allowing us to precipitate bound chromatin with streptavidin. For comparison, we used anti-TCF3 antibodies to precipitate chromatin in unmanipulated BL cells (Supplementary Table 5). Both the endogenous and TCF3-Biotag ChIP-seq peaks were enriched for E-box motifs (CAG(G/C)TG) and overlapped extensively (Supplementary Fig. 5a). In 25-bp bins bound by wild-type TCF3-Biotag, the V557E, D561E, and N551K isoforms had overlapping ChIP-seq tags (>7) in 98%, 98% and 92% of instances, respectively, but the overlap was only 10% for control Biotag ChIP-seq data ($p<10^{-300}$). Hence, all TCF3 mutants bound a large number of genomic targets equivalently.

Given the lower overlap between N551K and wild-type TCF3 chromatin binding, we identified genomic regions that had 4-fold greater (n=212) or lesser (n=139) association with wild-type TCF3 than N551K TCF3 ($p<10^{-10}$) (Supplementary Table 6, Supplementary Fig. 5b). In these binding regions, V557E and D561E TCF3 behaved like wild-type TCF3. The peaks bound preferentially by wild-type TCF3 contained multiple copies of the motif 5'-NNCACCTG-3' whereas the peaks bound preferentially by N551K were enriched for the sequence 5'-GGCAGCTG-3' (Fig. 2h). While both motifs match

the E-box consensus, these results suggest that N551K TCF3 is an altered specificity mutant that has somewhat different genomic targets than wild-type TCF3.

To gain insight into the biological processes controlled by TCF3 in BL, we profiled changes in gene expression following TCF3 knockdown or wild-type ID3 expression in ID3-mutant BL lines. We identified 139 "TCF3-upregulated" genes that were decreased in expression by both manipulations and 166 "TCF3-downregulated" genes that were increased in expression (FDR=0.017; Fig. 2i; Supplementary Figs. 6a, 2i). TCF3 ChIP-seq peaks were enriched among TCF3-upregulated genes (58%; $p=1.81x10^{-29}$) and among TCF3–downregulated genes (32%; $p=1.03x10^{-4}$) (Supplementary Fig. 6a). We will refer to such genes as "TCF3 direct targets".

Most TCF3-upregulated genes were more highly expressed in BL than in DLBCL whereas TCF3-downregulated genes were generally expressed at lower levels in BL ($p\leq0.001$; Fig. 2i; Supplementary Figs. 6a, b). BL tumors with *ID3* and/or *TCF3* mutations had higher expression of the TCF3-upregulated signature than tumors with wild-type *ID3* and *TCF3*, and the opposite was true for the TCF3-downregulated signature (p=0.0001; Supplementary Fig. 6c). Hence, the transcriptional influence of TCF3 on the BL phenotype appears to be accentuated by *ID3*/*TCF3* mutations. TCF3-upregulated genes were more highly expressed in germinal center B cells than in resting or activated blood B cells and the reverse was true for TCF3-downregulated genes (Fig. 2i; Supplementary Fig. 6a), suggesting that BL "inherits" the TCF3 gene expression program from its normal cellular counterpart.

Biological insights from this analysis include the fact that the negative regulators of TCF3 – *ID1*, *ID2*, and *ID3* – were direct targets of TCF3 transactivation, thereby creating a negative feedback loop (Fig. 2i; Supplementary Fig. 5a). By RNA-seq, *ID3* was 38-fold and 12-fold more highly expressed in BL than *ID1* and *ID2*, respectively, accounting for the preferential mutation of *ID3* in BL. TCF3 also positively regulated genes that play crucial roles in germinal center B cell biology (*POU2AF1*, *CXCR4*, *LTB*, *CCND3*). TCF3 upregulated *CCND3* and *E2F2* while downregulating *RB1*, thereby promoting cell cycle progression.

Two components of the B cell receptor, the immunoglobulin heavy and light chains, were both upregulated by TCF3 in BL, as in normal B cells[8,12] (Fig. 2i;

Supplementary Figs. 2i, j, 5a). In this regard, it was notable that knockdown of the BCR subunit CD79A was toxic for several BL lines in our RNA interference screen (Supplementary Fig. 7a). Two-thirds of BL lines were clearly BCR-dependent, based on a time-dependent decrease in their viability following knockdown of either CD79A or the BCR-associated kinase SYK (Fig. 3a). Unlike ABC DLBCL lines, which have a "chronic active" form of BCR signaling[4], BL lines do not require the NF-κB pathway for survival since they were not killed by an IκB kinase β inhibitor and had little or no dependence on CARD11[13], an adapter that engages NF-κB (Supplementary Fig. 7a, b, c). Rather, CD79A or SYK depletion in BL lines decreased AKT phosphorylation, a marker of PI(3) kinase signaling (Fig. 3b; Supplementary Fig. 7d), suggesting that the BCR-dependency in BL is akin to "tonic" BCR signaling[14], a phenomenon that engages pro-survival PI(3) kinase signaling more than NF-κB[15].

TCF3 knockdown decreased phospho-AKT levels in all BCR-dependent lines tested, as did ID3 overexpression (Fig. 3b, Supplementary Figs. 7d, e), perhaps due to decreased cell surface BCR expression following TCF3 depletion (Fig. 3c). In addition, a direct TCF3 target, *PTPN6*, encodes the phosphatase SHP-1, an inhibitor of BCR signaling (Supplementary Figs. 2j, 5a). TCF3 depletion increased SHP-1 mRNA and protein levels, indicating TCF3 repression (Fig. 3d; Supplementary Figs. 2i, 6a). Ectopic provision of SHP-1 decreased phospho-AKT in BCR-dependent BL lines, suggesting that TCF3 repression of SHP-1 may contribute to tonic BCR signaling and PI(3) kinase activation in BL (Fig. 3e).

A screen of a larger number of BL lines revealed that all had PI(3) kinase-dependent AKT phosphorylation and engagement of the mTOR pathway, as judged by phosphorylation of p70 S6 kinase (Fig. 3f). Treatment of BL lines with BKM120, a PI(3) kinase inhibitor in clinical trials, or rapamycin, an inhibitor of the mTORC1 complex, was toxic to most BL lines (Fig. 3g; Supplementary Fig. 7f). Of note, both BCR-dependent and –independent lines had constitutive PI(3) kinase signaling. Other mechanisms to activate PI(3) kinase in BL include *PTEN* mutations, which were infrequent (7%), and 10-fold overexpression (compared to DLBCL) of the MYC-dependent gene *MIR17HG*, which encodes a microRNA that inhibits PTEN expression[16] (Supplementary Fig. 7g). To judge whether the PI(3) kinase pathway may be active in

primary BL tumors, we identified genes that were significantly up- or down-regulated following rapamycin treatment of BL lines (FDR=0.0022; Supplementary Table 7). Among BL and GCB DLBCL biopsies, rapamycin down-regulated genes were generally more highly expressed in BL (p=0.026), whereas rapamycin up-regulated genes had the opposite enrichment pattern (p=0.007) (Fig. 3i), suggesting that PI(3) kinase-dependent mTORC1 activity is a consistent feature of BL tumors.

Another aspect of BL pathogenesis was revealed by recurrent mutations in the TCF3 direct target *CCND3*, encoding cyclin D3, a required regulator of the G1-S cell cycle transition in germinal center B cells[17,18]. *CCND3* mutations were frequent in sBL (38%) and hivBL (67%) but not eBL (1.8%), indicating a distinct genetic pathogenesis for this BL subtype (Figs. 4b). At a lower frequency, *CCND3* mutations were also present in ABC and GCB DLBCL[3,6,7]. Multiple nonsense and frame shift mutations removed up to 41 amino acids from the cyclin D3 C-terminus (Fig. 4a; Supplementary Table 8). Recurrent missense mutations affected threonine 283 (T283), known to be involved in D-type cyclin phosphorylation and stability[19], as well as nearby proline (P284) and isoleucine (I290) residues. These cyclin D3 residues were conserved in evolution, and similar residues are present in cyclin D1 and D2 (Fig. 4a). Most mutations were heterozygous and their somatic origin was confirmed in 5 cases (Supplementary Table 8).

To explore the function of the cyclin D3 mutants, we constructed fusion proteins linking green fluorescent protein (GFP) to either wild-type or mutant cyclin D3. All mutant isoforms accumulated to more than 10-fold higher levels than the wild-type isoform (Fig. 4c), and pulse-chase analysis showed that the mutant cyclin D3 isoforms have longer half lives (Supplementary Fig. 8a). To test the oncogenic potential of the cyclin D3 mutants, we transduced GFP-tagged wild-type or T283A cyclin D3 into lymphoma lines in which endogenous cyclin D3 was knocked down. Cells transduced with T283A cyclin D3 had a marked proliferative advantage over untransduced cells, but wild-type cyclin D3 had little effect (Fig. 4d). Separately, our RNA interference screen revealed that BL and GCB DLBCL lines depend on cyclin D3 and CDK6, a kinase that partners with D-type cyclins, irrespective of *CCND3* mutational status (Fig. 4e; Supplementary Figs. 8b, c, d, Supplementary Table 4). Hence, BL lines rely on cyclin D3/CDK6 for cell cycle progression, an effect augmented by oncogenic cyclin D3

mutations. The BL cell cycle is also deregulated by nonsense and frame shift mutations or homozygous deletions in *CDKN2A*, encoding the CDK6 inhibitor p16 (Supplementary Fig. 8e and Supplementary Table 8).

To explore this pathway as a therapeutic target, we treated BL, GCB DLBCL and mantle cell lymphoma (MCL) lines with a CDK4/6 inhibitor (PD 0332991) daily for 2 weeks. After an arrest in G1 phase, the BL and GCB DLBCL lines began to die by day 2, with a steady accumulation of apoptotic cells over time whereas the MCL line arrested in G1 phase but did not die (Fig. 4f). Treatment of a BL xenograft model after the establishment of tumors with PD 0332991 profoundly reduced tumor volume after 6 days, resulting in the virtual disappearance of tumor cells by day 10 (Fig. 4g, Supplementary Fig. 8f).

By merging functional and structural genomic data we have uncovered previously unappreciated pathways in BL pathogenesis, several of which are amenable to therapeutic attack (Fig. 4h). The majority of BL tumors acquire mutations that free TCF3 from ID3 inhibition. These mutations "hard-wire" a TCF3 transcriptional program that is characteristic of germinal center B cells and distinguishes BL from other aggressive lymphomas. BL lines require TCF3 for survival, in part because it augments pro-survival PI(3) kinase signaling by intensifying a tonic form of BCR signaling. The oncogenic synergy between the MYC and PI(3) kinase pathways that is suggested by our study is supported by the generation of BL-like tumors in mice in which these two pathways are deregulated[20]. Additionally, the key role of cyclin D3/CDK6 in BL pathogenesis is reinforced by the identification of cyclin D3 mutants in this mouse model.

While high-dose chemotherapy can often cure BL in younger patients from developed countries[1], these regimens are unsafe in older patients and cannot be deployed in less developed regions due to immune suppression and to logistical difficulties that preclude effective delivery[21]. Hopefully, the new insights into BL pathogenesis described herein will prompt clinical evaluation of drugs targeting the PI(3) kinase pathway, tonic BCR signaling, and cyclin D3/CDK6 in BL. Eventually, the rational combination of such targeted agents could provide more effective and less toxic treatment of BL worldwide.

Methods Summary

RNA-Seq was performed using established Illumina protocols on a HiSeq 2000 sequencer. RNA interference screening and cellular toxicity assays were conducted as described[5,13]. Gene expression profiling was performed using Agilent 4x44K microarrays. Detailed experimental and analytic procedures are presented in Supplementary Methods.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature

Acknowledgements

This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research, an NCI SPECS grant (UO1-CA 114778), and by the Foundation for NIH, through a gift from the Richard A. Lauderbaugh Memorial Fund, and by Cancer Research UK. This study was conducted under the auspices of the Lymphoma/Leukemia Molecular Profiling Project (LLMPP). R.S. was supported by the Dr. Mildred Scheel Stiftung für Krebsforschung (Deutsche Krebshilfe). D.J.H is a Kay Kendall Leukaemia Fund Intermediate research fellow. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Md. (http://biowulf.nih.gov). We thank Kathleen Meyer for help with the GEO submission, Tom Ellenberger for the E47 crystal structure coordinates, Bao Tran and the Center for Cancer Research Sequencing Facility and Kathleen Hartman in the NCI sequencing core. We thank the participants in the EMBLEM Study (http://emblem.cancer.gov/) in Uganda, the EMBLEM Study staff for collecting and processing the samples and data, and the Government of Uganda for allowing the study to be done and samples to be exported for research.

Author contributions

RS, RMY, MC, SJ, MZ, HK, ALS and DJH designed and performed experiments. TAW designed experiments. WX, YY, EB and HZ performed experiments. WX, GW, XL and JP analyzed data. AR, PK, HKM-H, GO, RDG, JMC, LMR, EC, ESJ, JD, EBS, RIF, RMB, RRT, JRC, DDW, WCC, SP, WW, MDO, SJR, SMM, MR and AR supplied BL patient samples or lines, and reviewed pathological and clinical data. LMS designed and supervised research and wrote the manuscript.

Gene expression profiling data have been submitted to GEO under accession number GSE35163

RNA-seq data has been deposited in NCBI Sequence Read Archive (SRA048058).

ChIP-seq data has been deposited in NCBI Sequence Read Archive (SRA052618).

Fig. 1: Recurrently mutated genes in aggressive lymphomas determined by RNA-seq. Shown are genes that were recurrently mutated in BL based on RNA-seq analysis (≥4/41 samples), as well as representative genes known to be recurrently mutated in DLBCL. Asterisks indicate differentially mutated genes ($p<0.05$; Supplementary Table 9).

Figure 2: *TCF3* is essential for Burkitt lymphoma viability. a, *TCF3* and *ID3* mutation frequencies in lymphoid cancers. MCL: mantle cell lymphoma; MM: multiple myeloma. **b,** Location of BL mutants in the crystal structure of the dimeric TCF3 E47 B-HLH domain[10]. **c,** Location of BL mutants in the crystal structure the ID3 HLH domain (pdb accession 2LFH). **d,** Selective toxicity of a TCF3 shRNA for BL lines. Shown is the fraction of GFP^+, shRNA-expressing cells relative to the GFP^-, shRNA-negative fraction at the indicated times, normalized to the day 0 values. Data are representative of 4 experiments. **e,** Toxicity of wild-type (WT) but not mutant ID3 isoforms for the ID3-mutant Namalwa BL line. Shown is the fraction of GFP^+, ID3-expressing cells relative to the GFP^-, ID3-negative cells, normalized to the day 0 values. Data are representative of 4 experiments. **f,** TCF3 mutants with reduced ability to bind ID3. WT or mutant TCF3 isoforms were coexpressed with WT ID3 in 293T cells. The indicated proteins were detected in total cellular extracts (input) or after anti-TCF3 immunoprecipitation (IP) (left). ID3 levels were quantified by densitometry and normalized to TCF3 E47 levels (right). **g,** BL-derived mutant ID3 proteins are less stable than WT ID3 and bind TCF3 less well. Mutant or WT ID3 isoforms were expressed in the ID3-deficient Namalwa BL line. The indicated proteins were detected in total cellular extracts (input) or after anti-TCF3 IP. **h,** N551K TCF3 is an altered specificity mutant. Shown at top are DNA base frequencies of the most enriched motifs in peaks bound > 4-fold more or less by N551K TCF3 compared to WT TCF3. The mean number (+/-s.e.m.) of the indicated motifs per differentially bound peak is plotted below. **i,** A TCF3 gene expression signature expressed in BL and normal germinal center B cells. Gene expression changes were profiled in ID3-mutant BL lines following TCF3 knockdown or WT ID3 overexpression. Shown are genes that were downregulated by at least 0.33 log2 in >70% of samples. Average expression of these genes in the indicated lymphoma subtypes based on

published data[2] and in B cell subpopulations based on RNA-seq is shown. *FS: frameshift, Δ: Deletion, *: nonsense.*

Fig. 3: **Tonic BCR signaling and PI(3) kinase activity in BL. a**, CD79A and SYK shRNAs are toxic for a subset of BL lines. Shown is the fraction of GFP^+, shRNA-expressing cells relative to the GFP^-, shRNA-negative fraction at the indicated times, normalized to the day 0 values. BCR-dependent BL lines are depicted using red colors. The BCR-dependent ABC DLBCL line TMD8[4] is also shown. Data are representative of 3 experiments. **b**, Knockdown of CD79A, SYK or TCF3 reduces PI(3) kinase activity. Following induction of the indicated shRNAs for 2 days, shRNA-expressing (GFP^+) cells were analyzed by FACS for phospho-S473-AKT as a measure of PI(3) kinase activity. **c,** TCF3 regulates surface BCR expression in BL. Following induction of the indicated shRNAs for 1 day, surface BCR expression (CD79B) was quantified by FACS in shRNA-expressing (GFP^+) cells. **d,** TCF3 suppresses *PTPN6* (SHP-1) expression. A TCF3 shRNA was induced in BL lines for 2 days, followed by immunoblotting for the indicated proteins. **e,** SHP-1 suppresses phospho-S473-AKT in BL lines. BL lines were transduced with a SHP-1 expression vector (+) or empty vector (–), whereupon the indicated proteins were analyzed by immunoblotting. **f**, BL lines have constitutively PI(3) kinase activity. The indicated proteins were analyzed by immunoblotting, before and after treatment with the PI(3) kinase inhibitor LY294002. **g**, PI(3) kinase inhibition is toxic to BL lines. Viable BL cells were quantified by MTS assay following treatment for 4 days with the indicated concentrations of the pan-class I PI(3) kinase inhibitor BKM120. **h**, A signature of rapamycin responsive genes is highly expressed in BL. Changes of gene expression were profiled over time in 2 BL lines following rapamycin (100 pM) treatment. Genes consistently downregulated in both lines were chosen (see Methods), and their expression in lymphoma biopsies[2] is shown based on the color scale. *PMBL: primary mediastinal B-cell lymphoma.* **i,** The rapamycin-upregulated and –downregulated signatures distinguish BL and GCB DLBCL. Genes are ranked according to their expression in BL vs. GCB DLBCL (T-statistic) and rapamycin signature genes are indicated with a green hash mark. Kolmogorov-Smirnov p-values are shown.

Fig. 4: Oncogenic *CCND3* mutations in Burkitt lymphoma a, Cyclin D3 residues affected by the indicated mutations in each lymphoma subtype. Amino acids 250-292 of NP_001751 are shown. *FL: follicular lymphoma; FS: frameshift* . **b,** Frequencies of *CCND3* mutations in different lymphoma subtypes. **c,** *CCND3* mutations increase protein stability. FACS analysis of the Gumbus BL line transduced with WT or mutant GFP-CCND3 fusion proteins. **d,** The T283A cyclin D3 mutant confers a proliferation advantage. Expression of endogenous *CCND3* was knocked down in Gumbus (BL) and BJAB (GCB DLBCL) cells and different GFP-CCND3 isoforms were ectopically expressed. The relative number of GFP-CCND3-expressing cells is plotted over time of shRNA and GFP-CCND3 induction, normalized to day 0. Data are representative of 3 experiments. **e,** *CCND3* shRNAs are selectively toxic for BL and GCB DLBCL lines. Shown is the fraction of GFP^+, shRNA-expressing cells relative to the GFP^-, shRNA-negative fraction at the indicated times, normalized to the day 0 values. Data are representative of 4 experiments. **f,** Cell cycle block in G1 phase is lethal to cyclin D3-mutant lymphoma lines. Lines were treated with the CDK4/6 inhibitor PD 0332991 (1 μM) over the indicated time course and analyzed for: viable cells in G1 phase, total viable cells and apoptotic cells. Data were normalized as indicated and are representative of 3 experiments. **g,** Therapeutic potential of PD 0332991 revealed using a BL xenograft model. Immunodeficient mice bearing established subcutaneous xenografts of the Gumbus BL line (engineered to express luciferase) were treated with PD 0332991 (150 mg/kg/day p.o.) for the indicated times. Tumor volumes were estimated by luciferase luminescence. Error bars are s.e.m. (n=3). **h,** Schematic of recurrent oncogenic pathways in Burkitt lymphoma. Gain-of-function and loss-of-function aberrations are indicated by plus signs and by X signs, respectively. Grey boxes indicate drugs that block these deregulated pathways.

References

1 **Yustein, J.T. & Dang, C.V., Biology and treatment of Burkitt's lymphoma. *Curr Opin Hematol* 14 (4), 375-381 (2007).**

2 **Dave, S.S. *et al.*, Molecular diagnosis of Burkitt's lymphoma. *N Engl J Med* 354 (23), 2431-2442 (2006).**

3 **Morin, R.D. *et al.*, Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* (2011).**

4 **Davis, R.E. *et al.*, Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* 463 (7277), 88-92 (2010).**

5 **Ngo, V.N. *et al.*, Oncogenically active MYD88 mutations in human lymphoma. *Nature* 470 (7332), 115-119 (2011).**

6 **Pasqualucci, L. *et al.*, Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet* (2011).**

7 **Lohr, J.G. *et al.*, Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc Natl Acad Sci U S A* 109 (10), 3879-3884 (2011).**

8 **Murre, C., McCaw, P.S., & Baltimore, D., A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 56 (5), 777-783 (1989).**

9 **Kee, B.L., E and ID proteins branch out. *Nat Rev Immunol* 9 (3), 175-184 (2009).**

10 **Ellenberger, T., Fass, D., Arnaud, M., & Harrison, S.C., Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev* 8 (8), 970-980 (1994).**

11 **Pesce, S. & Benezra, R., The loop region of the helix-loop-helix protein Id1 is critical for its dominant negative activity. *Mol Cell Biol* 13 (12), 7874-7880 (1993).**

12 **Murre, C. *et al.*, Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* 58 (3), 537-544 (1989).**

13 **Ngo, V.N. *et al.*, A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* 441 (7089), 106-110 (2006).**

14 **Lam, K.P., Kuhn, R., & Rajewsky, K., In vivo ablation of surface immunoglobulin on mature B cells by inducible gene targeting results in rapid cell death. *Cell* 90 (6), 1073-1083 (1997).**

15 **Srinivasan, L. *et al.*, PI3 kinase signals BCR-dependent mature B cell survival. *Cell* 139 (3), 573-586 (2009).**

16 **Xiao, C. *et al.*, Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. *Nat Immunol* 9 (4), 405-414 (2008).**

17 **Cato, M.H., Chintalapati, S.K., Yau, I.W., Omori, S.A., & Rickert, R.C., Cyclin D3 is selectively required for proliferative expansion of germinal center B cells. *Mol Cell Biol* (2010).**

[18] **Peled, J.U. *et al.*, Requirement for cyclin D3 in germinal center formation and function. *Cell Res* 20 (6), 631-646 (2010).**

[19] **Diehl, J.A., Cheng, M., Roussel, M.F., & Sherr, C.J., Glycogen synthase kinase-3beta regulates cyclin D1 proteolysis and subcellular localization. *Genes Dev* 12 (22), 3499-3511 (1998).**

[20] **Sander, S. *et al.*, Synergy between PI3K signalling and MYC in Burkitt lymphomagenesis. *Cancer Cell* DOI:10.1016/j.ccr.2012.06.012 (2012).**

[21] **Orem, J., Mbidde, E.K., & Weiderpass, E., Current investigations and treatment of Burkitt's lymphoma in Africa. *Trop Doct* 38 (1), 7-11 (2008).**

Figure 1

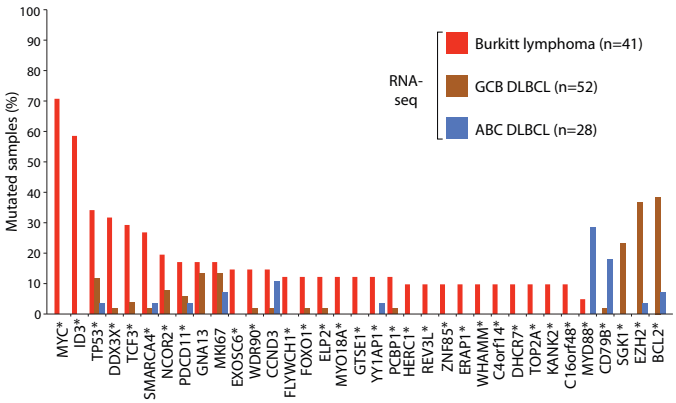

Mutated samples (%)
100
90
80
70
60
50
40
30
20
10
0
RNA-seq
Burkitt lymphoma (n=41)
GCB DLBCL (n=52)
ABC DLBCL (n=28)
MYC*
ID3*
TP53*
DDX3X*
TCF3*
SMARCA4*
NCOR2*
PDCD11*
GNA13
MKI67
EXOSC6*
WDR90*
CCND3
FLYWCH1*
FOXO1*
ELP2*
MYO18A*
GTSE1*
YY1AP1*
PCBP1*
HERC1*
REV3L*
ZNF85*
ERAP1*
WHAMM*
C4orf14*
DHCR7*
TOP2A*
KANK2*
C16orf48*
MYD88*
CD79B*
SGK1*
EZH2*
BCL2*

Figure 2

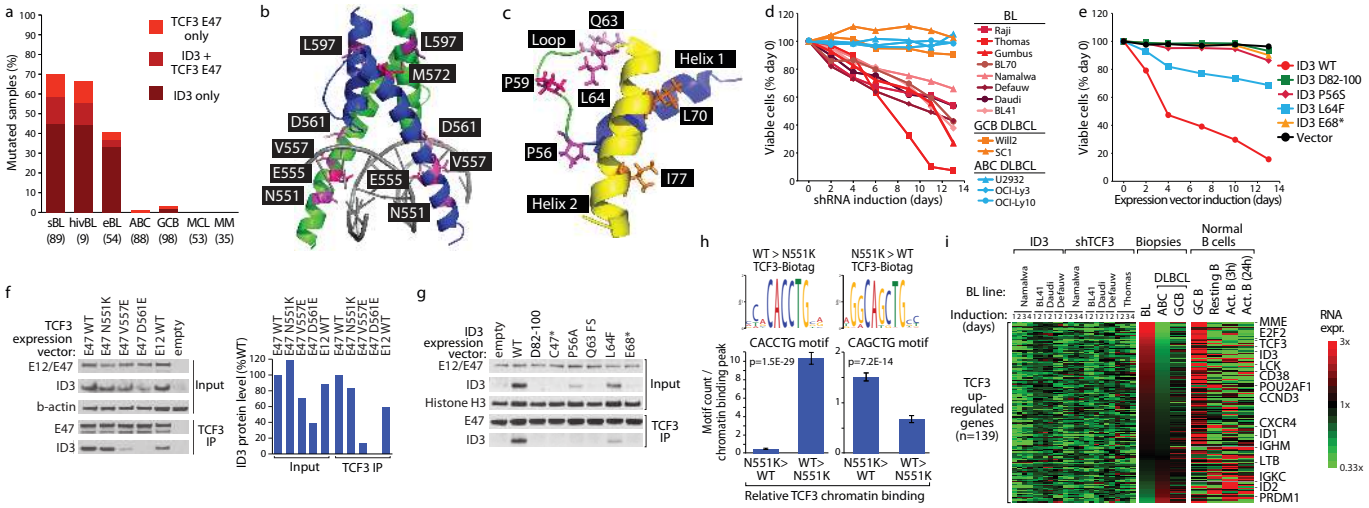

a
Mutated samples (%)
TCF3 E47 only
ID3 + TCF3 E47
ID3 only
sBL (89)
hivBL (9)
eBL (54)
ABC (88)
GCB (98)
MCL (53)
MM (35)
b
L597
L597
M572
D561
D561
V557
V557
E555
E555
N551
N551
c
Loop
Q63
Helix 1
P59
L64
L70
P56
I77
Helix 2
d
Viable cells (% day 0)
shRNA induction (days)
BL
Raji
Thomas
Gumbus
BL70
Namalwa
Defauw
Daudi
BL41
GCB DLBCL
Will2
SC1
ABC DLBCL
U2932
OCI-Ly3
OCI-Ly10
e
Viable cells (% day 0)
Expression vector induction (days)
ID3 WT
ID3 D82-100
ID3 P56S
ID3 L64F
ID3 E68*
Vector
f
TCF3 expression vector:
E12/E47
ID3
b-actin
E47
ID3
Input
TCF3 IP
ID3 protein level (%WT)
Input
TCF3 IP
g
ID3 expression vector:
E12/E47
ID3
Histone H3
E47
ID3
Input
TCF3 IP
h
WT > N551K TCF3-Biotag
N551K > WT TCF3-Biotag
CACCTG motif
CAGCTG motif
p=1.5E-29
p=7.2E-14
Motif count / chromatin binding peak
N551K> WT
WT> N551K
N551K> WT
WT> N551K
Relative TCF3 chromatin binding
i
ID3
shTCF3
Biopsies
Normal B cells
BL line:
Induction: (days)
DLBCL
BL
ABC
GCB
GC B
Resting B
Act. B (3h)
Act. B (24h)
TCF3 up-regulated genes (n=139)
MME
E2F2
TCF3
ID3
LCK
CD38
POU2AF1
CCND3
CXCR4
ID1
IGHM
LTB
IGKC
ID2
PRDM1
RNA expr.
3x
1x
0.33x

Figure 3

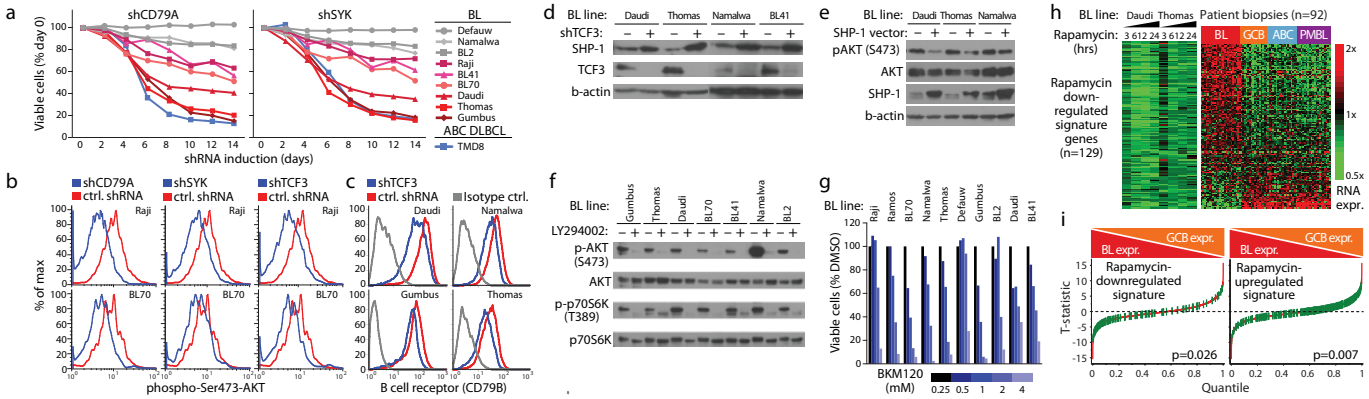
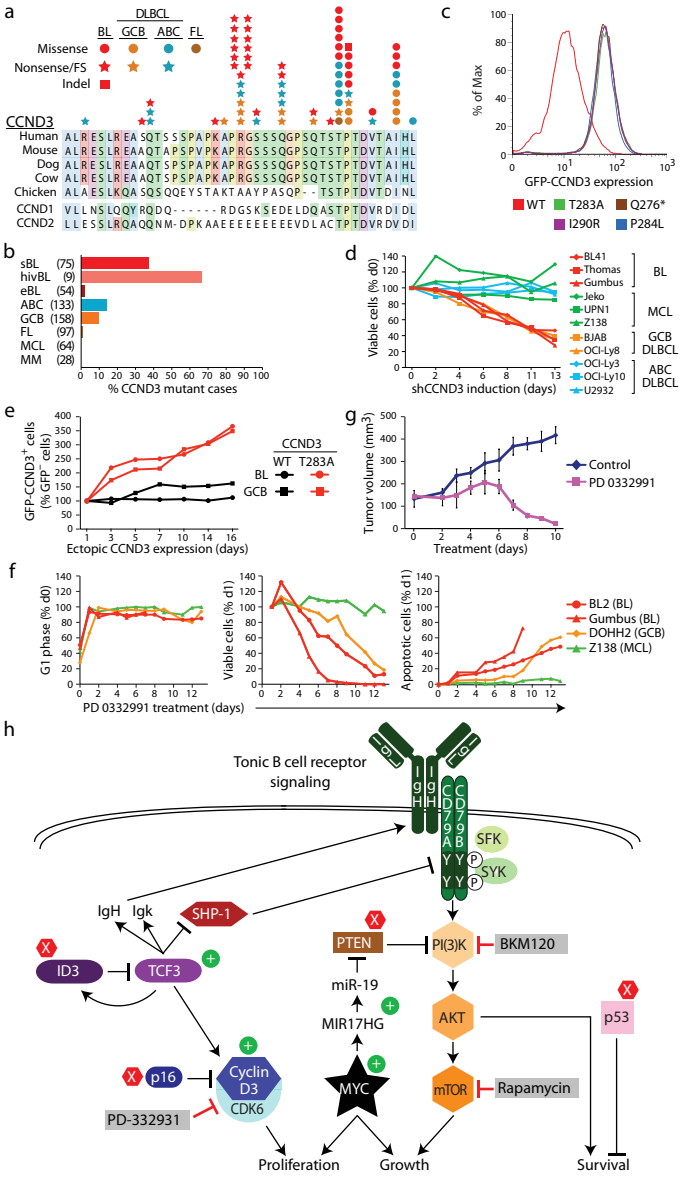

a
shCD79A
shSYK
Viable cells (% day 0)
shRNA induction (days)
BL
Defauw
Namalwa
BL2
Raji
BL41
BL70
Daudi
Thomas
Gumbus
ABC DLBCL
TMD8
b
shCD79A
shSYK
shTCF3
ctrl. shRNA
Raji
BL70
% of max
phospho-Ser473-AKT
c
shTCF3
ctrl. shRNA
Isotype ctrl.
Daudi
Namalwa
Gumbus
Thomas
B cell receptor (CD79B)
d
BL line:
Daudi
Thomas
Namalwa
BL41
shTCF3:
SHP-1
TCF3
b-actin
e
BL line:
Daudi
Thomas
Namalwa
SHP-1 vector:
pAKT (S473)
AKT
SHP-1
b-actin
f
BL line:
Gumbus
Thomas
Daudi
BL70
BL41
Namalwa
BL2
LY294002:
p-AKT (S473)
AKT
p-p70S6K (T389)
p70S6K
g
BL line:
Raji
Ramos
BL70
Namalwa
Thomas
Defauw
Gumbus
BL2
Daudi
BL41
Viable cells (% DMSO)
BKM120 (mM)
0.25 0.5 1 2 4
h
BL line:
Daudi
Thomas
Patient biopsies (n=92)
Rapamycin: (hrs)
3 6 12 24 3 6 12 24
BL
GCB
ABC
PMBL
Rapamycin down-regulated signature genes (n=129)
2x
1x
0.5x
RNA expr.
i
GCB expr.
BL expr.
Rapamycin-downregulated signature
Rapamycin-upregulated signature
T-statistic
p=0.026
p=0.007
Quantile

Figure 4


a
DLBCL
BL GCB ABC FL
Missense
Nonsense/FS
Indel
CCND3
Human
Mouse
Dog
Cow
Chicken
CCND1
CCND2
c
% of Max
GFP-CCND3 expression
WT T283A Q276*
I290R P284L
b
sBL (75)
hivBL (9)
eBL (54)
ABC (133)
GCB (158)
FL (97)
MCL (64)
MM (28)
% CCND3 mutant cases
d
Viable cells (% d0)
shCCND3 induction (days)
BL41
Thomas
Gumbus
Jeko
UPN1
Z138
BJAB
OCI-Ly8
OCI-Ly3
OCI-Ly10
U2932
BL
MCL
GCB DLBCL
ABC DLBCL
e
GFP-CCND3+ cells (% GFP- cells)
Ectopic CCND3 expression (days)
CCND3
WT T283A
BL
GCB
g
Tumor volume (mm3)
Treatment (days)
Control
PD 0332991
f
G1 phase (% d0)
Viable cells (% d1)
Apoptotic cells (% d1)
PD 0332991 treatment (days)
BL2 (BL)
Gumbus (BL)
DOHH2 (GCB)
Z138 (MCL)
h
Tonic B cell receptor signaling
IgH IgK
SHP-1
ID3
TCF3
SFK
SYK
PTEN
PI(3)K
BKM120
miR-19
MIR17HG
AKT
p53
p16
Cyclin D3
CDK6
PD-332931
MYC
mTOR
Rapamycin
Proliferation
Growth
Survival