JOURNAL OF
EMERGING INVESTIGATORS

Article

Discovery of novel targets for diffuse large B-cell lymphoma

Vivian Xiao[1], Yu Liu[1]
[1] Poolesville High School, Poolesville, Maryland

SUMMARY

The newly emerging CD19-targeted Chimeric Antigen Receptor T-cells (CAR-T) therapy has revolutionized the treatment of advanced diffuse large B-cell lymphoma (DLBCL) and achieved a functional cure in some patients. However, a majority of patients suffer relapse. A major pathway for tumor relapse is the loss of the target antigen, CD19, on the cancer cell surface, essentially making them invisible to the therapy. Consequently, new targets are required for the continued treatment of relapsed patients. Facing this dilemma, we hypothesized that a systematic search of the human genome may reveal more novel targets of DLBCL with characteristics similar to CD19. These include high expression in DLBCL and other B-cell tumors, but minimal expression in normal tissues. To test this hypothesis, we used several well-annotated public genomic and proteomic databases and carried out in-depth screening to pinpoint optimal new targets for DLBCL. Our study uncovered several new targets that have not been extensively pursued in the public domain, such as CD79A and TNFRSF13C. Based on our refined criteria for target evaluation, we finally concluded that CD79A represents the most ideal novel target for DLBCL and potentially other B-cell tumor types. This target should be explored as the antigen of next-generation CAR-T therapy that has the potential to save and extend the lives of patients.

INTRODUCTION

Hematological malignancy, commonly known as blood cancer, represents ~10% of all human cancers (1, 2). The majority of them are derived from B-lymphocytes and can be classified into B-cell lymphoma, including DLBCL, and B-cell leukemia (1, 2). The main difference is that lymphoma is mainly found in lymph nodes whereas leukemia generally occurs in the blood. Recently, progress has been made on treating B-cell tumors based on novel therapeutic platforms such as CAR-T and Antibody-Drug Conjugate (ADC) against well-established targets, such as CD19, the prototypical B-cell target broadly expressed by most B-cell malignancies. Some previously untreatable patients have been essentially cured by these revolutionary modern medicines (3-5).

CAR-T therapy is an emerging therapy that is fundamentally different from conventional cancer drugs such as chemotherapy. Unlike chemotherapy which indiscriminately kills tumor and normal cells alike, CAR-T specifically targets the tumor cells while sparing the healthy cells. CAR-T represents a "live-cell" therapy where the drug is in the form of engineered human T-cells (3, 4). In this process, T-cells are harvested from the patients' blood, engineered outside the body with a recombinant gene encoding the CAR construct. The T-cells are expanded greatly and then infused back to the corresponding patients, where they will seek out the tumors and destroy them in a precision-driven manner.

CAR-T is a fusion construct consisting of an antibody-based extracellular binding element that recognizes the tumor antigen, linked to an intracellular signaling domain capable of activating the T-cells through a transmembrane region (3, 4). When the extracellular domain binds the tumor antigen, it causes the intracellular domain to trigger activation of the T-cells, transforming them into highly voracious "serial-killers" of tumor cells (3, 4, 5).

CAR-T is a very powerful drug and has strict requirements for its targets: 1) cell-surface membrane protein with significantly exposed extracellular domain (so that it can be recognized by the antibody-part of CAR-T); 2) substantial tumor expression; and 3) minimal normal tissue expression, especially absent in vital organs.

Despite success in curing some patients, the majority of patients with DLBCL only derive temporary benefits from the CD19 CAR-T and will eventually relapse (5, 6). Interestingly, most of the relapses were caused by CD19 loss or downregulation, which represents an effective way for tumors to evade CD19-targeted therapies (6-9).

In response to this emerging mechanism of resistance, additional B-cell-specific antigens have been proposed as new targets to treat relapsing tumors that have lost *CD19* expression. The primary example is *CD22*, another well-established B-cell lineage marker. However, recent clinical trial data indicated that patients showed transient response to CD22-targeted CAR-T before relapsing again through a similar process of target down-regulation, invoking the same escape pathway (10, 11). This unfortunate reality indicates that we need more novel and compelling targets for B-cell tumors.

To address this challenge, we used a comprehensive and in-depth investigation of the human genome to find additional targets for DLBCL and other B-cell tumors with similar tissue expression patterns as CD19, such as high expression in B-cell tumors but restricted expression in normal tissues specifically to the B-cell lineage. These novel targets may also qualify as compelling targets for CAR-T. As a canonical B-cell target, CD19 satisfies these two criteria well as it is highly and broadly expressed in B-cell-derived cancers and its normal tissue expression is restricted to the B-cell lineage only. These properties help ensure that CD19 targeted drugs will not only achieve good anti-tumor efficacy, but also

minimize the toxicity in healthy tissues (3, 4).

With this paradigm in hand, we set out to discover and explore potentially useful online databases that are freely available to the public: Gene Expression Profiling Interactive Analysis (GEPIA), Human Protein Atlas (HPA), and UniProt (12-14).

GEPIA turns out to be the most valuable tool. It not only captures the target expression pattern in tumors, but also its normal tissue expression in the same dataset, enabling a direct comparison of RNA expression levels in tumors and normal tissue. The larger the "window" (differential) between the tumor and normal tissue expression levels, the better the target. We established a screening algorithm that first identified targets that display the most similar tissue expression pattern to CD19, then assessed the various hits for their strength of tumor expression and extent of normal tissue expression. In conclusion, we selected the best target, CD79A, due to its strongest DLBCL tumor expression and the most restricted normal tissue exposure.

RESULTS

We started the analysis by searching for CD19 in GEPIA and HPA in order to generate its tumor and normal tissue expression data (**Figure 1**). Among all tumor types, *CD19* is uniquely and highly expressed in DLBCL (**Figure 1A**). We also found that *CD19* expression profile in normal tissues is highly restricted to B-cells but mostly absent from the other immune tissue types like T-cells, natural killer (NK) cells and myeloid cells (neutrophils, monocytes, etc.) (**Figure 1B**).

Next, we identified the most similar genes to *CD19* by searching its "Most Similar Genes" in GEPIA (**Table 1**). The Top hits were tabulated in the following manner: gene symbols in column-1, gene IDs in column-2, and ranking (by

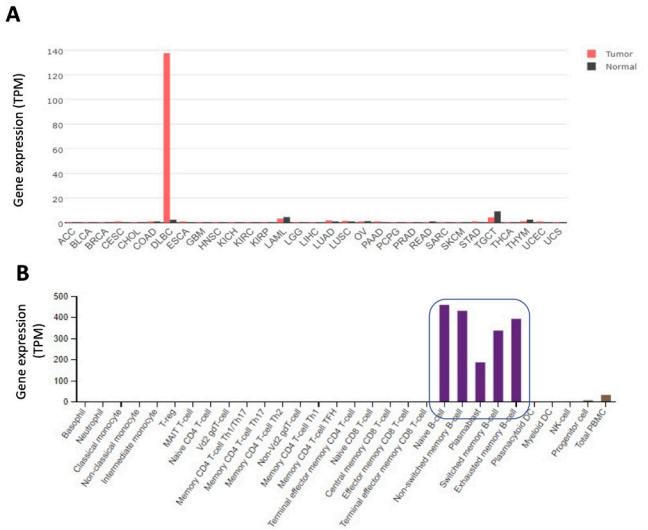

Figure 1: CD19 expression profiling in tumor and normal tissues. A: CD19's expression pattern in different tumor types and their corresponding normal tissues was established by performing a search with the target's name in the GEPIA database as described in Materials and Methods. B: CD19's expression pattern in different immune cell types was established by performing a search with the target's name in the HPA (Human Protein Atlas) database as described in Materials and Methods. The box denotes specific restriction of CD79A expression to B-cell types.

Gene Symbol	Gene ID	CD19 similarity score
PAX5	ENSG00000196092.12	0.78
TNFRSF13C	ENSG00000159958.4	0.78
CD79A	ENSG00000105369.9	0.77
BLK	ENSG00000136573.12	0.73
MS4A1	ENSG00000156738.17	0.72
SP140	ENSG00000079263.18	0.72
CD72	ENSG00000137101.12	0.71
PTPRCAP	ENSG00000213402.2	0.68
CD79B	ENSG00000007312.12	0.68

Table 1: Cell surface localization scores for treatment target candidates.

their similarity scores to *CD19*) in column-3. These hits were screened in UniProt to identify cell surface membrane protein targets and the resultant hits were highlighted in red (**Table 1**).

Of the top hits, we identified five that are cell surface membrane proteins: *TNFRSF13C*, *CD79A*, *MS4A1*, *CD72*, and *CD79B*. *MS4A1* is better known as *CD20* and is among the oldest and most well-known B-cell markers. Since we are most interested in discovering novel targets, we excluded this molecule from further triage.

Then, we studied the remaining 4 targets in GEPIA to ascertain their DLBCL tumor expression levels (**Figure 2**). All four showed high specific expression in DLBCL, but a much lower expression in other tumor types as well as normal tissues. Judging by the expression levels (TPM on the Y-axis), we found *CD79A* and *CD79B* were highly and robustly expressed in DLBCL, whereas *CD72* and *TNFRSF13C* had the weakest expression.

Subsequently, we investigated the normal tissue expression profiles of these targets by searching the HPA dataset (**Figure 3**). Strikingly, all targets demonstrated selective enrichment in B-cell types, indicating they are qualified B-cell lineage markers. In-depth interrogation showed some subtle differences: while *CD79A* is mostly specific to B-cells and displays minimal exposure in other immune cell

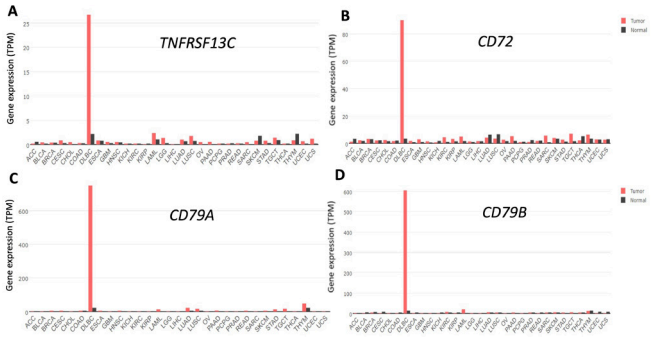

Figure 2: Top-4 cell surface membrane proteins with most similar expression profiles to CD19. Genes most similar to CD19 were individually screened in the Uniprot database for their subcellular localization profiles. Genes encoding cell surface membrane proteins were identified and their tumor expression patterns revealed by GEPIA search.

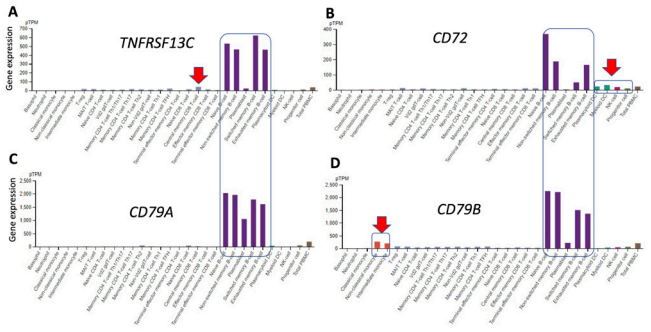

Figure 3: Expression profiles of the top targets across different immune cell types. The top-4 targets were screened in HPA for their expression patterns in normal immune cell types to verify whether they are restricted B-cell lineage markers. The big boxes denote that the major expression of the targets is restricted to B-cell types; the red arrows denote the targets' additional expression in non-B-cell types.

types (**Figure 3C**), the other three targets all showed modest expression in non-B cell types (**Figure 3A,B,D**, red arrows). For example, *CD79B* has low but significant expression in monocytes (**Figure 3D**: both non-classical monocyte and intermediate monocytes); *CD72* displays some expression in myeloid dendritic cells, plasmacytoid dendritic cells, and NK cells (**Figure 3B**).

Combining the results on both tumor and normal tissue expression, we concluded that *CD79A* is the most ideal novel DLBCL target and hence recommended for the next-generation CAR-T therapy. This is supported by the following facts: it has the highest expression in the tumor tissues (**Figure 2**), and it has the most restricted normal tissue expression specific to B-cells only (**Figure 3**).

As the last step of target validation, we ascertained the protein expression profile of CD79A in HPA, offered in addition to its gene expression dataset. This is a necessary step because our screening process so far is mainly driven by RNA expression analysis, but the actual target of the CAR-T therapy is the translated protein product from the RNA. Even though RNA expression generally correlates with the encoded protein, it is always prudent to verify that this is the case.

HPA's protein expression data is based on immunohistochemistry (IHC), a special protein visualization technology that relies on tissue staining by an antibody that specifically recognizes the target protein. The CD79A IHC profile across 20 different cancer types, ranging from glioma to lymphoma, was downloaded from HPA and tabulated (**Table 2**).

For each cancer type, the total number of samples stained and the number of positively stained tumor samples were displayed. Percentage of positive samples were calculated by dividing the number of positive samples by the total tumor samples. Lymphoma was highlighted in red to indicate it is the only tumor type with significant CD79A protein expression (**Table 2**).

There is a striking correlation with the gene expression profile in GEPIA; among all cancer types, only lymphoma, to which DLBCL is a major component, displayed strong staining. Among a total of 12 lymphoma tissues, 10 of them showed high or medium staining for CD79A, resulting in 83% positivity. In contrast, other cancer types (mostly solid tumors) did not show any meaningful staining.

Tumor type	Number of positive samples	Number of total samples	% of positive samples
Glioma	0	12	0
Thyroid	0	4	0
Lung	0	12	0
Colorectal	0	12	0
Head and neck	0	4	0
Stomach	0	12	0
Liver	1	12	8
Carcinoid	0	4	0
Pancreatic	0	12	0
Renal	0	12	0
Urothelial	0	12	0
Prostate	0	11	0
Testis	0	12	0
Breast	0	9	0
Cervical	0	12	0
Endometrial	0	9	0
Ovarian	0	10	0
Melanoma	0	12	0
Skin	0	11	0
Lymphoma	10	12	83

Table 2: CD79A IHC profile across 20 different cancer types.

DISCUSSION

Despite great strides made in cancer therapies, most cancers are incurable. CAR-T is the prime example of a novel and highly promising precision cancer therapy. Before the advent of CAR-T, patients with relapsed or refractory B-cell lymphoma, such as DLBCL, were viewed as untreatable. By conferring a nearly 50% complete remission rate, the CD19-targeted CAR-T transformed the field and lifted the *de facto* death sentence that these patients faced. However, although a small number of patients do achieve long-term cure, for most patients, the initial benefits are generally followed by disease recurrence. This unfortunate event is frequently driven by the loss of the target, CD19, on tumor cells which essentially makes them invisible to the CAR-T.

Realizing the urgent need to find new targets to enable the retreatment of the tumors after the downregulation of CD19, we hypothesized that additional targets with similar tissue expression pattern as CD19 can be discovered in the human genome. By leveraging public databases including HPA, GEPIA and UniProt, that allow systematic gene expression analysis across a variety of tumor and normal tissues, we designed an efficient screening algorithm that rapidly identified promising new targets from over 28,000 genes in the human genome.

HPA represents a hybrid database with mixed gene expression (RNA) and protein expression data. Specifically, it illustrates the expression profile of any gene across most normal tissues. It also provides the proteomic expression data, in the form of IHC, of any target in small collections of both normal tissues as well as tumor tissues. In contrast, GEPIA consists of purely RNA expression data, but it compiles the most comprehensive data by providing both the normal tissue and tumor expression result of any gene in the same dataset, hence enabling a direct comparison of a target's relative expression in tumor vs. normal tissues. This is a very useful feature allowing rapid target triage. UniProt, on the other hand, is not an expression database but represents the most authoritative evaluation tool to help you determine whether a protein is a cell-surface membrane protein (hence targetable

by antibody-based therapy like CAR-T) or not.

We narrowed down the initial "hit" list by verifying the cellular localization of the target. We relied on UniProt to identify the targets with cell surface localization profile. This trimmed down the list to 4: CD72, CD79A, CD79B, and TNFRSF13C (**Table 1**).

Applying two more criteria, namely the strength of tumor (DLBCL) expression (**Figure 2)** and restricted extent of normal tissue expression (**Figure 3**), we concluded that *CD79A* represents the most promising target due to its high level of DLBCL tumor expression and narrow restriction to the B-cell lineage among all normal tissues. Then, we leveraged HPA to confirm that the gene expression profile of CD79A is highly concordant with its protein expression pattern, and abundant CD79A protein can be detected in most of the malignant lymphoma samples. In aggregate, our hypothesis-driven analysis shows that CD79A is likely the most qualified new target for DLBCL and potentially other B-cell tumor types.

Our findings suggest CD79A-targeted therapy will likely show the highest anti-tumor potency by being the most abundant target on the tumor cells. At the same time, it may also display the best safety profile: its normal tissue exposure is highly restricted to B-cells, whose depletion is tolerated well by humans based on previous clinical experience with CD19. CD19 marks all B-cells, and CD19-targeted CAR-T efficiently depletes B-cells in the patients, and this was found to be well-tolerated and manageable. Therefore, we conclude that B-cell depletion caused by CD79A-targeting may also be as tolerable.

Our analysis had several limitations. Although we demonstrated the abundant expression of CD79A in both mRNA and protein levels in DLBCL lymphoma samples, and CD79A is a verified cell surface membrane protein based on UniProt annotation, our *in-silico* analysis stopped short of proving that CD79A can be successfully targeted by an antibody. To accomplish this goal, lab-based research is needed to first generate a monoclonal antibody that recognizes the extracellular portion of CD79A protein, and then demonstrate that this antibody can bind to lymphoma cells using an experimental procedure called FACS (fluorescence-activated single-cell sorting). A positive signal from this future assay would provide the final validation of CD79A. Another pitfall is related to the limited tumor coverage of the GEPIA database. B-cell-derived tumors include many different sub-types other than DLBCL, like B-cell leukemia, mantle cell lymphoma, follicular lymphoma, and Burkitt's lymphoma. Because GEPIA only contains expression data for DLBCL and not the other subtypes, we could not establish CD79A's expression profile beyond DLBCL. This issue can only be addressed by expanded tumor data.

Target quality is the most critical determinant of the success of a drug, especially for CAR-T therapy. The most important driver for the success of CD19-targeted CAR-T is that CD19 is a high-quality target. CD19 is widely expressed across most B-cell tumors, while narrowly restricted to the B-cell lineage among healthy tissues. This expression profile ensures that the drug would not attack any normal tissues and incur any safety issues beyond the depletion of normal B-cells, which is known to be tolerated by humans and easily manageable. Hence, we have established a solid rationale in starting our genome-wide search by identifying targets with the most similar tissue expression profile as CD19.

Despite the similarity in tissue expression pattern, CD79A has a distinct biological role than CD19. CD79A is one of the most essential components of the B-cell receptor (BCR) complex which is required for B-cell development, differentiation, and activation upon antigen binding (15-17). The BCR signaling pathway is a key driver of B-cell malignancies, especially DLBCL, and CD79A is a critical component of BCR, Therefore it is reasonable to speculate that tumor cells may not easily escape CD79A-targeted therapy by simply losing or downregulating the expression of an essential gene like CD79A, unlike earlier experiences with CD19 or CD22 (18-20). These rationales suggest that a CD79A-targeted CAR-T may confer a more durable response and longer-lasting patient benefits than CAR-Ts against CD19 or CD22.

CD79A CAR-T could be utilized as a stand-alone therapy. Alternatively, CD79A CAR-T could be combined with pre-existing ones like CD19 CAR-T for more effective and durable cancer control. This combination scheme may prevent or delay the emergence of cancer relapse due to the loss of CD19. Downregulation of CD19 may help tumors acquire resistance to CD19 CAR-T, but these tumor cells are likely still positive for CD79A expression (83% based on HPA proteomic data: **Table 2**). These tumor cells may still be efficiently killed by the CD79A-targeted CAR-T, leading to a more durable response and longer-lasting benefit.

By successfully pinpointing an ideal new target for DLBCL treatment, we demonstrated that 1) quality cancer research and drug discovery can be conducted by high-school students with a keen interest in human health and some working knowledge of public databases, and 2) a properly conceived hypothesis and a smartly designed study plan can go a long way towards maximizing the chance of meaningful and impactful discovery. We are hopeful that the compelling new target we have identified here can be exploited to enable the next-generation CAR-T therapy.

MATERIALS AND METHODS

New target screening design and flow-chart

The key criteria of a good target for CAR-T are captured in **Figure 4**, which also illustrated the corresponding public database that can be used to verify whether a target has satisfied a particular criterion.

To ascertain whether a protein (the translated product of a gene) is a surface membrane protein or not, we used UniProt, the most commonly accepted protein database with a clear annotation of this property. To investigate the normal tissue

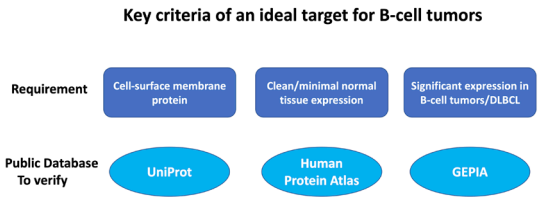

Figure 4: The three key criteria of a good cancer target for CAR-T therapy. First requirement is a cell surface membrane protein; second requirement is a clean target with minimal or restricted normal tissue expression; third requirement is significant high expression in tumors of interest (in this case B-cell tumors such as DLBCL).

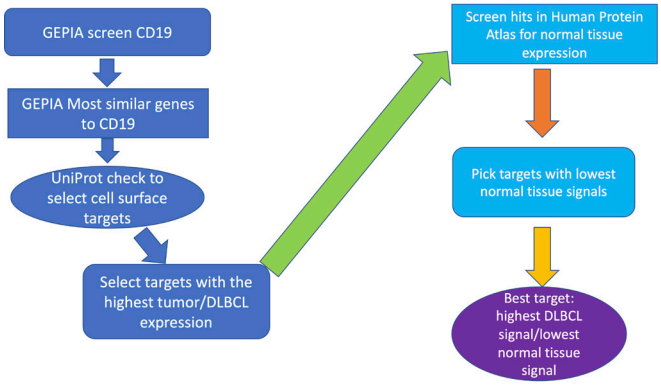

Figure 5: New DLBCL target screening flow-chart. The left side represents the initial screen in GEPIA for CD19 to capture the "most similar genes" to CD19 (with similar tissue expression patterns), the resultant hits were then filtered by UniProt to identify targets that are cell surface membrane proteins. Top leads are the ones with the highest tumor/DLBCL expression. These targets are further triaged in HPA to select the ones with the lowest/most restricted normal tissue expression. The final targets will have both significant tumor expression as well as low/minimal normal tissue expression.

expression profile of a target, we used Human Protein Atlas/HPA. The third database we used was GEPIA, a reputed database that describes the expression profiles of any human gene across most tumor and normal tissue types in the same dataset, enabling instant visualization of its potential overexpression window in tumor vs. normal tissues.

To test our hypothesis, we followed the protocol outlined in **Figure 5** to search for novel targets. We first searched CD19 in GEPIA, then looked for genes most similar to CD19. Afterwards, we used UniProt to identify the targets in this list that are cell surface membrane proteins, then screened these potential targets in GEPIA and HPA to investigate their expressions in DLBCL tumors and normal tissues. Finally, we determined the ideal target(s), with the strongest tumor expression and lowest/narrowest normal tissue expression.

Public databases used in this study

GEPIA: Gene Expression Profiling Interactive Analysis (12)
HPA: Human Protein Atlas (13)
UniProt: UniProt Consortium (14)

Received: April 14, 2021
Accepted: September 2, 2021
Published: November 16, 2021

REFERENCES

1. "Lymphoma." American Cancer Society. https://www.cancer.org/cancer/lymphoma.html
2. "FACTS AND STATISTICS OVERVIEW." Leukemia & Lymphoma Society. https://www.lls.org/facts-and-statistics/facts-and-statistics-overview
3. Schultz, L, and Crystal Mackall. "Driving CAR T cell translation forward." Sci Transl Med. 27;11(481), 2019. doi: 10.1126/scitranslmed.aaw2127
4. June, Carl H, *et al*. "CAR T cell immunotherapy for human cancer." Science. Vol 359, number 6382, 2019, pp:1361-1365. doi: 10.1126/science.aar6711
5. Lim, WA and Carl H. June. "The Principles of Engineering Immune Cells to Treat Cancer." Cell. Vol 9, number 168(4), 2017, pp: 724-740. doi: 10.1016/j.cell.2017.01.016
6. Majzner, R, and Crystal Mackall. "Clinical lessons learned from the first leg of the CAR T cell journey." Nat Med. 25(9), 2019, pp: 1341-1355. doi: 10.1038/s41591-019-0564-6
7. Shah, N and Terry Fry. Mechanisms of resistance to CAR T cell therapy. Nat Rev Clin Oncol. 16(6), 2019, pp: 372-385. doi: 10.1038/s41571-019-0184-6
8. Walsh, Z, *et al*. "Multi-Specific CAR Targeting to Prevent Antigen Escape." Curr Hematol Malig Rep. 14(5), 2019, pp: 451-459. doi: 10.1007/s11899-019-00537-5
9. Qin, H, *et al*. "Preclinical Development of Bivalent Chimeric Antigen Receptors Targeting Both CD19 and CD22." Mol Ther Oncolytics. 11, 2018, pp: 127-137. doi: 10.1016/j.omto.2018.10.006
10. Fry, T, *et al*. "CD22-targeted CAR T cells induce remission in B-ALL that is naive or resistant to CD19-targeted CAR immunotherapy." Nat Med. 24(1), 2018, pp: 20-28. doi: 10.1038/nm.4441
11. Shah, Nirali. "The one-two punch (of CAR T cells)." Blood. 135(5), 2020, pp: 303-304. doi: 10.1182/blood.2019004272
12. Tang, Z. *et al*. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 10.1093, 2017/nar/gkx247. doi: 10.1093/nar/gkx247
13. Uhlén, M, *et al*. "Tissue-based map of the human proteome." Science (2015) PubMed: 25613900 DOI: 10.1126/science.1260419. doi: 10.1126/science.1260419
14. The UniProt Consortium. "UniProt: the universal protein knowledgebase in 2021." Nucleic Acids Res. 49:D1 (2021). doi: 10.1093/nar/gkaa1100
15. Weiss, A, and D Littman. "Signal transduction by lymphocyte antigen receptors." Cell. 1994 Jan 28; 76(2):263-74. doi: 10.1016/0092-8674(94)90334-4.
16. Ezequiel, M, *et al*. "Analysis of the Individual Contributions of Igα (CD79a)- and Igβ (CD79b)-Mediated Tonic Signaling for Bone Marrow B Cell Development and Peripheral B Cell Maturation." J Immunol December 1, 2006, 177 (11) 7913-7922. doi: 10.4049/jimmunol.177.11.7913
17. Kelly, A Pike, *et al*. "The cytoplasmic domain of Ig alpha is necessary and sufficient to support efficient early B cell development." J Immunol. 2004 Feb 15;172(4):2210-8. doi: 10.4049/jimmunol.172.4.2210.
18. Niemann, Carsten, and Adrian Wiestner. "B-cell receptor signaling as a driver of lymphoma development and evolution." Semin Cancer Biol. 2013 Dec;23(6):410-21. doi: 10.1016/j.semcancer.2013.09.001.
19. Davis, R. Eric *et al*. "Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma." Nature. 2010 Jan 7;463(7277):88-92. doi: 10.1038/nature08638.
20. Havranek, O, *et al*. "Tonic B-cell receptor signaling in diffuse large B-cell lymphoma." Blood. 2017 Aug 24;130(8):995-1006. doi: 10.1182/blood-2016-10-747303