A Systems Biology Analysis of Chronic Lymphocytic Leukemia

Giulia Pozzati[1*], Jinrui Zhou[2*], Hananel Hazan[3], Giannoula Lakka Klement[4], Hava T. Siegelmann[5], Edward A. Rietman [5,6†], Jack A. Tuszynski[1,7,8†]

[1] Dipartimento di Ingegneria Meccanica e Aerospaziale (DIMEAS), Politecnico di Torino, I-10129 Turin, Italy; giulia.pozzati30@gmail.com (GP)
[2] Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA; sherry.jinrui.zhou@gmail.com (JZ)
[3] Allen Discovery Center at Tufts University, Medford, MA, 02155, USA; hananel@hazan.org.il (HH)
[4] CSTS Healthcare, Toronto, ON, Canada, giannoula@aiomics.com (GKL)
[5] Manning College of Information and Computer Science, University of Massachusetts, Amherst, MA 01003, USA; erietman@gmail.com (EAR), hava.siegelmann@gmail.com (HTS)
[6] Applied Physics, 477 Madison Ave. 6th floor, New York, NY, 10022, USA; ed@appliedphysics.org (EAR)
[7] Department of Data Science and Engineering, The Silesian University of Technology, Gliwice, Poland
[8] Department of Physics, University of Alberta, Edmonton, AB T6G 2E9, Canada; jackt@ualberta.ca (JAT)

* These authors contributed equally.
† Correspondence: jackt@ualberta.ca, erietman@gmail.com

Abstract

Whole-genome sequencing has revealed that TP53, NOTCH1, ATM, SF3B1, BIRC3, ABL, NXF1, BCR, ZAP70 are often mutated in CLL, but not consistently across all CLL patients. This paper employs a statistical thermodynamics approach in combination with the systems biology of the CLL protein-protein interaction networks to identify the most significant participant proteins in the cancerous transformation. Betti number (a topology of complexity) estimates highlight a protein hierarchy, primarily in the Wnt pathway known for aberrant CLL activation. These individually identified proteins suggest a network-targeted strategy over single-target drug development. The findings advocate for a multi-target inhibition approach, limited to several key proteins to minimize side effects, thereby providing a foundation for designing therapies. This study emphasizes a shift towards a comprehensive, multi-scale analysis to enhance personalized treatment strategies for CLL, which could be experimentally validated using siRNA or small molecule inhibitors. The result is not just the identification of these proteins but their rank-order, offering a potent signal amplification in the context of the 20,000 proteins produced by the human body, thus providing a strategic basis for therapeutic intervention in CLL, underscoring the necessity for a more holistic, cellular, chromosomal, and genome-wide study to develop tailored treatments for CLL patients.

Keywords: CLL; Systems biology; PPI network; Gibbs homology; Betti number; Wnt pathway

Author Summary

Chronic Lymphocytic Leukemia (CLL) is a unique and slowly progressing cancer affecting white blood cells, and research on CLL has highlighted the inconsistency of gene mutations across patients. Using a novel approach that merges statistical thermodynamics and systems biology, this research examines the CLL protein-protein interaction networks to pinpoint proteins integral to the onset of the disease. Betti number (a topology of complexity) estimates, which measure the importance of individual proteins when removed from the network, helped identify numerous potential therapeutic targets, notably within the Wnt signaling pathway, a pathway implicated in various cellular processes and known for its defective expression in CLL. The finding advocates for a multi-target inhibition approach, focusing on several key proteins to minimize side effects, thereby laying a foundation for designing more

effective therapies for CLL. This paper emphasizes the potential benefits of a comprehensive study, spanning cellular to genome-wide scales, to design personalized treatments for CLL patients.

Introduction

Chronic lymphocytic leukemia (CLL) is a type of cancer that affects white blood cells and tends to progress slowly over many years. It is a chronic lymphoproliferative disorder characterized by increased production of morphologically mature but immunologically dysfunctional B lymphocytes. As a result, these cells are unable to fight infections as well as normal white blood cells do [1].

The disease starts developing in the bone marrow, since here leukemia cells survive longer and eventually outnumber normal cells. Then, cells further grow and may spread to other parts of the body including the spleen, the lymph nodes and the liver [1]. Since the growth of leukemia cells is slow, CLL may remain latent for many years before it causes symptoms, and it is usually harder to cure than acute leukemias [1].

From the genetic perspective, CLL is a unique disease with multiple gene signatures. One cohort of patients can exhibit a different gene-signature set than another cohort. Whole-genome sequencing has revealed that TP53, NOTCH1, ATM, SF3B1, BIRC3, ABL, NXF1, BCR, ZAP70 are often mutated in CLL; but, not consistently across all CLL patients [2]–[4]. For example, NOTCH1 is mutated in about 10% of newly diagnosed patients, and in about 15% to 20% of progressive ones. Similarly, SF3B1 is mutated in about 10% of newly diagnosed CLL patients, and about 17% in late-stage disease [2]. Just because a gene is mutated does not mean it will be strongly expressed. One of the goals of our study is to show a molecular thermodynamics approach to determine the most energetically significant pathways supporting a given patient's CLL initiation and progression. This new molecular systems approach may shed light on optimal treatment for each patient – essentially personalized therapy. Before we present this new methodology, we provide an overview of the known biomarkers for CLL and then a survey of the current treatment options as well as experimental drugs in development.

It is of fundamental importance to obtain information about the patient's status and prognosis to define therapeutic strategy. There exist several laboratory-based prognostic markers, such as high levels of serum beta-2 microglobulin (B2M) and the absolute lymphocyte count (ALC). However, chromosomal aberrations detected using Fluorescent In Situ Hybridization (FISH) serve as the main prognostic tools. The most common aberrations detected in CLL patients are [5]:

- Deletions on the long arm of chromosome 13 (del(13q)): in patients with this aberration the disease progresses slowly.
- Deletions on the long arm of chromosome 11 (del(11q)): this usually occurs among young males and tends to manifest with bulky lymph nodes. It is associated with rapid disease progression and short survival. 11q chromosome contains Ataxia-Telangiectasia mutated gene (ATM) and ATM kinase is responsible for inhibited cell cycle progression in case of DNA damage. Furthermore, ATM kinase acts on p53 by phosphorylating it in order to induce apoptosis. Therefore, when 11q is deleted, this phosphorylation does not occur, and the cell damage cannot be repaired [6].
- Deletion on the short arm of chromosome 17 (del(17p)): results in the loss of TP53 which is the most important prognostic marker in CLL. It is associated with rapid disease progression and resistance to fludarabine chemoimmunotherapy. In addition to the role of TP53 as a prognostic marker in CLL, it is also fundamentally a predictive marker for chemo-immunotherapy, guiding treatment decisions and potentially influencing the response to specific therapies such as fludarabine chemoimmunotherapy [7].
- Immunoglobulin heavy-chain variable region gene (IGHV) mutational status. For prognosis and therapy choice it is important to detect IGHV mutational status since the unmutated state is correlated with low survival.
- Other markers, which are present in a low percentage of newly diagnosed CLL patients, but whose incidence increases in patients who are refractory to fludarabine chemotherapy, are mutations of NOTCH1, SF3B1 and BIRC3. Finally, combining genetics, clinical parameters and biochemistry, the CLL International Prognostic Index (CLL-IPI) is a tool to predict the status of the disease [8].

Wnt signaling is a network of interacting protein pathways which control processes such as cell differentiation, cell cycle regulation, proliferation, apoptosis, cytoskeletal rearrangement, cell polarity, adhesion, motility, migration and invasion and the interaction with the microenvironment [9]. Wnt signaling is correlated with haematopoiesis and is linked with leukaemogenesis of cancers such as CLL [9]. Two Wnt signaling pathways are associated with CLL, namely the Wnt/β-catenin dependent and independent pathways. The Wnt/β-catenin is associated with cell proliferation, homeostasis, cell cycle regulation and thus its malfunction indicated a hallmark of many cancers. Regarding the Wnt/β-catenin independent pathway, the Wnt/PCP (Planar Cell Polarity) is the most important one and it takes place in regulation of cell polarity, migration and invasion. Wnt pathways play a role in CLL pathogenesis and response to treatment. Moreover, the expression of Wnt signaling molecules from Wnt/β-catenin and Wnt/PCP pathways is defective in CLL. For example, ROR1 (receptor tyrosine kinase-like orphan receptor), a Wnt-5 (a Wnt protein) dedicated receptor in the Wnt/PCP pathway, is expressed on the surface

of CLL cells and not on the healthy B-cells. Therefore, ROR1 is a sensitive marker of a possible relapse of patients with a more aggressive form of the disease.

The scope of this study extends beyond the traditional single-target silver bullet approach in drug development, acknowledging the intricate network of proteins that drive the pathological transformation of CLL. A systems biology perspective indicates that targeting a manageable group of 5-6 network nodes could be more effective for combination therapy design, considering the potential for serious side effects due to overlapping off-target interactions. The statistical thermodynamics method applied here aims to identify and hierarchize such targets, which could be inhibited by existing approved or investigational drugs, setting the stage for a more nuanced and personalized treatment approach in CLL.

Current Treatment Options and Experimental Drug Candidates

Unfortunately, currently available treatments may relieve CLL patients from their symptoms and extend their survival, but still CLL remains incurable [6]. For patients without "active disease" who are asymptomatic or those with early-stage disease, the treatment consists of just a simple observation during which blood counts are performed every three months [6]. For patients with "active disease", before choosing therapy, the clinical status must be evaluated in terms of general health, characteristics such as TP53 abnormalities or adverse cytogenetics or relapsed disease [6]. Standard treatment has been chemoimmunotherapy with fludarabine, cyclophosphamide and rituximab (FCR). However, it has demonstrated lack of efficacy and it leads to numerous side effects, especially in patients with TP53 or NOTCH1 mutations, unmutated IGHV, deletion of 17p or 11q [8].

Target agents are small molecules that have greater efficacy in patients harboring TP53 mutation or del(17p) whose examples include [6]:

- Bruton tyrosine kinase inhibitors (ibrutinib, acalabrutinib),
- BCL-2 inhibitor (venetoclax),
- Purine analogs (fludarabine, pentostatin),
- Alkylating agents (cyclophosphamide, chlorambucil, bendamustine),
- Monoclonal antibodies (rituximab, ofatumumab, obinutuzumab),
- PI3K inhibitor (idelalisib).

The most common chemotherapy medications used are listed below, together with their main mode of action:

- Fludarabine: a purine analogue and an antineoplastic agent.
- Cyclophosphamide: an alkylating agent.
- Rituximab: a monoclonal antibody which targets the B-lymphocyte antigen CD20 expressed on the surface of B cells.

These three together (FCR) constitute a chemoimmunotherapy treatment:

- Bendamustine: an alkylating agent used along with Rituximab (BR) to form another combination chemoimmunotherapy treatment.
- Chlorambucil: an alkylating agent.
- Ibrutinib is a Bruton tyrosine kinase (BTK) inhibitor. BTK, an enzyme which works for B cell survival and growth, helps delay the progression of cancer. It inhibits CLL cell migration, proliferation and survival [10]. Unfortunately, it presents some side effects such as pneumonia, upper respiratory tract infection, atrial fibrillation, sinusitis, headaches, nausea and many more [10].
- Acalabrutinib is a more selective irreversible BTK inhibitor since it acts just like ibrutinib, but without the side effects involving other kinases [10]. Its most common side effects are headaches, tiredness, low red blood cells, low platelets and low white blood cells [10].
- PI3K (Phosphatidylinositol-3-kinase) inhibitors such as Idelalisib [11], which was FDA approved in 2014 for use in combination with rituximab for treating relapsed CLL [12]. However, Idelalisib is also toxic with nearly 40% of patients having had to interrupt the therapy due to rash or 3-4 grade transaminitis, and pulmonary infections. A PI3Kδ inhibitor called TGR 1202 has better selectivity compared to Idealisib. It was approved for medical use in the USA in February 2021. TGR 1202 reduces the phosphorylation of AKT in lymphoma and leukemia cells.
- Venetoclax binds and inhibits the antiapoptotic protein B-cell lymphoma 2 (BCL-2) [9]. In CLL, inhibition of this pathway has been considered an optimal therapeutic strategy [13]. Use of Venetoclax was approve by FDA in 2016 [13]. Its side effects usually include low levels of white and red blood cells, respiratory infections, diarrhea, nausea, tiredness and tumor lysis syndrome (TLS).
- Sotorasib **(AMG510)** is a highly selective and irreversible inhibitor which binds at an allosteric pocket leading to the trapping of KRAS (Kirsten rat sarcoma virus) in inactive GDP bound state. Note that KRAS transmits signals for growth, division and differentiation to the nucleus of the cell form the outside. KRAS mutations are among the most oncogenic events in carcinomas, including CLL, and the

majority of them consist of missense mutation of the 12th codon (glycine). It was approved by FDA in May, 2021. Some of its side effects are diarrhea, nausea and muscles or bone pain [14].

- Adagrasib **(MRTX849)** is an irreversible covalent inhibitor of G12C KRAS mutation that makes a covalent bond to cysteine and binds in the switch-II pocket of KRAS in its inactive GDP state. It demonstrated improved antitumor activity when in combination with vistusertib (an mTOR inhibitor). In clinical trials some patients experienced pneumonitis and heart failure, which led to the interruption of the treatment. Others experienced nausea, fatigue and anemia. This inhibitor is still in clinical trials together with numerous other experimental drugs under development [14].
- Experimental drug candidates also include AKT pathway allosteric inhibitors: ARQ092/miransertib; BAY1125976; MK2206, TAS-117 [15].
- ATP-competitive AKT inhibitors: capivasertib and ipatasertib showed a favorable safety profile along with signs of activity in phase I monotherapy trials [14]. Other AKT inhibitors include the following compounds: Afuresertib **(GSK2110183),** Uprosertib **(GSK2141795, GSK795**) and Ordidonin **(NSC-250682**) [16].

Additionally, various drug candidates are in development with MYC pathway inhibition profiles [17]:

- Compound 361 **(MYCi361, NUCC-0196361**) [18].
- Compound 975 (**MICi975, NUCC-0200975**) [19].
- **MYCMI-6** [20].
- **KSI-3716** and **MYRA-A** [20].
- **KI-MS2-008** [20].
- **L755507** [17].

Systems Biology Background

The conceptual framework for understanding the thermodynamics and energetics of the molecular biology of human diseases from a network biology perspective has been developed over the past decade. Various studies were undertaken to quantify different signaling and metabolic pathways in various cancer types and other diseases using metrics such as network entropy and the Gibbs free energy applied to each specific case [21]–[26]. Here, we will give only a brief summary of this approach. The transcriptome and other -omic (e.g., proteomic, genomic, etc.) measures can represent the energetic state of a cell. Using the word "energetic", we mean from a thermodynamics perspective. There is a chemical potential between interacting molecules in a cell, and the chemical potential of all the proteins that interact with each other can be imagined forming a rugged landscape, not dissimilar to Waddington's epigenetic landscape [27], [28].

The method we propose uses mRNA transcriptome data or RNA-seq data as a surrogate for protein concentration. This assumption is largely valid. Kim et al. [29] and Wihelm et al. [30] have shown an 83% correlation between mass spectrometry-generated proteomic information and transcriptomic information for multiple tissue types. Further, Guo et al. [31] found a Spearman correlation of 0.8 in comparing RNAseq and mRNA transcriptome from TCGA human cancer data [32].

Given a set of transcriptome data, a representative of protein concentration, we overlay that on the human protein-protein interaction network from BioGrid [33]. This means we assign to each protein on the network, the scaled (between 0 and 1), transcriptome value (or RNAseq value). From that we can compute the Gibbs free energy of each protein-protein interaction using the mapping relation:

$$G_i = c_i \ln \frac{c_i}{\sum_j c_j} \qquad [1]$$

where c_i is the "concentration" of the protein i, normalized, or rescaled, to be between 0 and 1. The sum in the denominator is taken over all protein neighbors of i, and including i. Therefore, the denominator can be considered a degree-entropy. Carrying out this mathematical operation essentially transforms the "concentration" value assigned to each protein to a Gibbs free energy. Thus, we replace the scalar value of transcriptome to a scalar function – the Gibbs free energy.

The above equation is derived from a well-known concept in chemical thermodynamics [34]. A biological cell, or a group of cells (a tumor) exist in a complex chemical balance produced by a network of interacting molecular species ranging from small molecules to some very large molecules on the order of hundreds to thousands of Daltons. The molecular concentration-balance in this network is the Gibbs free energy G. This thermodynamic quantity is typically expressed in the context of systems kept at a constant temperature and pressure, where the system can exchange molecules with the environment. For an arbitrary molecular system, the Gibbs function is given as a molar difference [35] in Equation [2]:

$$\delta G = \mu \delta n. \qquad [2]$$

where μ symbolizes the chemical potential, δG is the Gibbs energy and δn is the molar difference (essentially concentration difference). Typically, one writes the chemical potential as:

$$\mu_i = \left(\frac{\delta G}{\delta n_i}\right)_{P,T,j,k,etc.} \qquad [3]$$

The Equation [3] above assumes that the molar concentrations of other molecular components (other than i) are held constant along with constant temperature and pressure. Using equation [1] and given a network of interacting chemical species, or proteins, and given their concentration, we can compute the Gibbs free energy for a single protein in the PPI.

The Gibbs free energy is a negative number, so associated with each protein on the network is a negative energy well. This results in a rugged energy landscape represented schematically in Figure 1. If we use what is referred to as a topological filtration on this landscape, we essentially move a filtration plane up from the deepest energy well. As the filtration plane is moved up, larger-and-larger energetic subnetworks are captured. For convenience we stop the filtration at energy threshold 32 – meaning 32 nodes in the energetic subnetwork. We call these subnetworks Gibbs-homology networks.

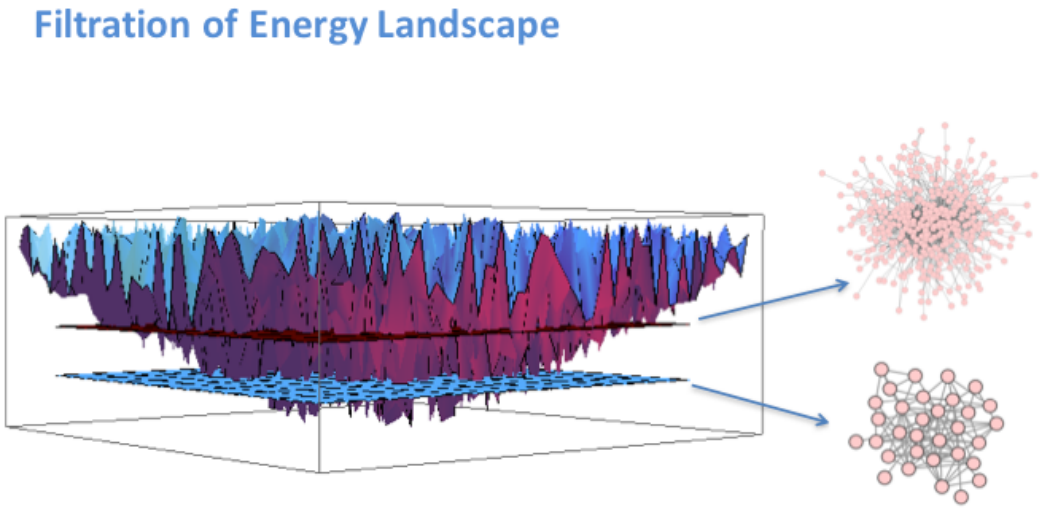


Figure 1: As the "filtration plane" moves up from the bottom, more-and-more nodes are captured in larger-and-larger energetic subnetworks for protein-protein interaction set.

We now compute the Betti centrality, a topological measure, on the 32-node energetic networks as described in Benzekry et al. [23]. The concept is easily described. In networks, there are holes, or rings, of various sizes. In these energetic pathways, protein-protein interaction networks, the proteins form interaction rings. In densely connected, but not fully connected, networks the rings, or holes, may consist of triangles and larger rings of interaction. To find the Betti centrality we ask ourselves: Which protein when removed from the network will change the overall total number of rings the most? The total number of rings is called the Betti number. Given a network G consisting of edges e and vertices v, the Betti centrality is given by Equation [4]:

$$B(v_i) = B(G) - B(G - \{v_i\}) \qquad [4]$$

Hence, the difference from the total Betti number B(G) and the Betti number of the network after removing node i, gives the Betti centrality for node i. We compute this for all nodes in the threshold-32 energetic network. Often, there will be two or more proteins in the network that have equivalent Betti centrality.

Methodology and Datasets

We report on a meta-analysis of 1001 samples from CLL patients and cancer cell lines. This study used online data from GEO [36]: GSE10139, GSE28654, GSE31048, GSE39671, GSE49896, GSE50006, and GSE69034. The data were mRNA expression numbers, all collected using Affymetrix Human Genome Array, HG-U133_2. We also used the human protein-protein interaction network from Biogrid [33]. In particular, we used the dataset downloaded from the BIOGRID-ORGANISM: homo_sapiens-3.5.172.2.

Reiterating the method, we collected the GSE expression datasets, and then the expression value for each gene was overlaid on the human protein-protein interaction network for each protein or node in the network. For each node in the network, we then applied Equation [1] which resulted in the Gibbs energy for that node. This resulted in a rugged landscape similar to Figure 1. Then, the procedure consisted in doing a filtration and computing the Betti number for zero nodes removed and then removing a node and recomputing the Betti number and replacing the node. This removal-computation-replacement procedure resulted in a list of nodes that had the largest impact on complexity of the Gibbs homology network. We finally ranked significant nodes in a Pareto chart for each patient. Pareto charts were prepared at several filtration thresholds, 32, 48, 64, 96.

Results

Our discussion of the results is presented below, and it follows an analysis of the individual datasets and the research publication associated with it (if present) prior to presenting the meta-analysis Pareto chart and the network graphs.

1. [37] (GSE10137) "A genomic approach to improve prognosis and predict therapeutic response in chronic lymphocytic leukemia", by Friedman et al. 2009. This was one of the papers with a large table in the Supplementary section. The table consisted of upregulated and downregulated probes indicative of progressive disease; upregulated and downregulated probes indicating chlorambucil resistance; upregulated and downregulated probes indicative of Pentostatin, Cyclophosphamide, and Rituximab signature. An important quote from the paper states that: "Others have previously noted the prognostic significance of cytoskeletal genes and the tumor necrosis factor in CLL. Notably, probes for ZAP-70 did not constitute this genomic signature, although mean expression for ZAP-70 probes in samples from patients with progressive disease was higher than those from patients with stable disease." The table of genes was parsed from the PDF document and used in our subsequent analysis (discussed below).
2. [3] (GSE28654) "Gene expression profiling identifies ARSD as a new marker of disease progression and sphingolipid metabolism as a potential novel metabolism in chronic lymphocytic leukemia" Trojani, et al. 2012 [3]. A table in the manuscript lists about 65 genes that were selected as being differentially expressed in two cohorts of CLL patients. Of those genes the authors selected 19 genes for PCR analysis because of their significance. Those genes are: ZAP70, ARSD, LPL, ADAM29, AGPAT2, CRY1, MBOAT1, YPEL1, NRIP1, RIMKLB, P2RX1, EGR3, TGFBR3, APP, DCLK2, FGL2, ZNF667, CHPT1, FUT8. An important quote from the paper states that: "In the literature, lists of differentially expressed genes obtained using high-throughput microarray by different laboratories and research centers have often limited overlap [[38], [39]. These differences are matters of important scientific discussions and are imputed, among other causes, to dataset dimensions: small number of subjects (some tens) with respect to the number of variables (tens of thousands of genomic probes in human). … Notably and reassuringly the gene set list (65 genes) emerged from this study showed a substantial (but not quantitated) overlap with results from previously published microarray studies [40]–[43]." The list of 65 genes was incorporated in our subsequent analysis.
3. [40] (GSE31048) "Somatic mutation as a mechanism of Wnt/β-Catenin pathway activation in CLL" [40]. In the Supplement to this paper were two large tables listing genes. One table listed from their own study (Wnt pathway) and the other table listed Wnt genes from literature and Websites. Both tables were combined for the study. A quote from the paper: "… our data demonstrate that altered gene expression is indistinguishable between samples with and without mutations."
4. [41] (GSE39671) "Subnetwork-based analysis of chronic lymphocytic leukemia identifies pathways that associate with disease progression" Chuang et al. 2012 [41]. The Supplemental data only included figures and graphs. No table of gene list. Of note is the quote: "Furthermore, the marker sets identified by different research groups often share few genes in common. Two landmark studies, Rosenwald and colleagues [42] and Klein and colleagues [43] each identify approximately 100 genes that were expressed differentially by CLL cells that use mutated versus unmutated IGHV genes. However, only 4 marker genes were identified in common between these studies."
   [44] (GSE49896) "miR-150 influences B-cell receptor signaling in chronic lymphocytic leukemia by regulating expression of GAB1 and FOXP1", Mraz, et al. 2014 [44]. The following is quoted from their paper: "We identified miR-150 as being the most abundantly expressed miRNA in CLL. However, we observed significant heterogeneity in the expression levels of this miRNA among CLL cells of different patients. Low-level expression of miR-150 associated with unfavorable clinicobiological and prognostic markers, such as expression of ZAP-70 or use of unmutated IGHV ($P < .005$). Additionally, our data suggest that the

levels of methylation of the upstream region of 1000 nt proximal to miR-150 associate with its expression. We demonstrated that GAB1 and FOXP1 genes represent newly defined direct targets of miR-150 in CLL cells. We also showed that high-level expression of GAB1 and FOXP1 associates with relatively high sensitivity of CLL cells to surface immunoglobulin ligation. High levels of GAB1/FOXP1 and low levels of miR-150 associate with a greater responsiveness to BCR ligation in CLL cells and more adverse clinical prognosis."

5. GSE50006 – no manuscript.
6. GSE69034 – no manuscript.

We created a master list of all genes cited and/or given in the tables associated with the above manuscripts. This list is in the Appendix 2. There were 515 genes total. The list of genes is inputted into the DAVID platform for functional annotation analysis [45], and only 208 genes are found, which indicates that DAVID's database has annotations for only 208 of those genes. The missing genes might be due to them being less well-characterized, newer discoveries not yet integrated into DAVID's database, or they might be represented differently in the user's list compared to DAVID's nomenclature. To identify genes relevant for a generic condition like "leukemia", the KEGG [46] and OMIM [47] databases are used to filter and analyze the results such that both are integrated into DAVID. These databases contain curated information about genes related to specific pathways or diseases. By cross-referencing the 208 identified genes with "leukemia" in both KEGG and OMIM, genes whose expression or mutation is linked with the onset, progression, or other aspects of leukemia are pinpointed, aiming in narrowing down potential targets for research, therapeutic development, or further molecular study. Searching that file resulted in the following list AKT1, CTBP1, CTBP2, CTBPA, SMAD4, HDAC1, LEF1, RARA, TCF3, TCF7, TCF7L1, TCF7L2, MYC.

Comparing the PublishedGeneList with our CLLnet96 list, only four were found: MYC, HDAC1, CTNNB1, APP. Two of those, MYC and HDAC1 are known to participate in leukemia. The CLLnet96 list is assembled from all 1001 patients at Gibbs threshold 96. To reiterate the concept of threshold. For any given patient the deepest well in the landscape is, usually the same for all thresholds; but there may be differences based on the expression, and this gives rise to differences in the Gibbs homology network. An energy threshold of, say 32, will result in a network of 32 nodes that are the largest negative energy values. This is called a topological filtration. Using this technique, we can produce one of these 32 threshold networks for each patient. If we do that, and then concatenate the entire list of nodes for each of the patients at this threshold, followed by sorting and discarding redundant nodes in the list, the result will be what we call, CLLnet32 list. By the nature of the filtration, CLLnet32 $\subset$ CLLnet48 $\subset$ CLLnet64 $\subset$ CLLnet96. In words, CLLnet32 is a proper subset of CLLnet48, etc. So, taking the list CLLnet96 will by definition incorporate all others. Comparing our CLLnet96 with the superset of published genes (i.e. PublishedGeneList in the Appendix 2), we find only four that were both lists, MYC, HDAC1, CTNNB1, and APP.

After comparing the PublishedGeneList and the CLLnet96 superset, we then used DAVID, an online bioinformatics resource that allows one to submit a list of genes (or other biological components, e.g. proteins) and it returns important information such as KEGG pathway or OMIM associated with that gene. There are 98 genes in the superlist of CLL96net list. From that analysis we find the following genes to be associated with leukemia (various types): KRAS, GRB2, HDAC1, NPM1, TP53, MYC. While in that superlist, CUL1, TP53, and CTNNB1 is associated with the Wnt signaling pathway.

The published gene lists consisted of two parts. Keeping in mind that although the papers cited above included GSE expression data, most of them did not include tables of genes they identified from their analysis as being important. Instead, they were looking for prognostic markers for disease progression. So, Part 1 of the published gene list consisted of selections identified by the authors from, GSE10137, GSE28654. The combined list consisted of 320 genes. Of those 320 genes, 22 were found in DAVID. Only CEBPA and MYC were found to be associated with any form of leukemia. And CSNK2A1 and MYC were found to be associated with Wnt signaling pathway. When we expand the published list to include GSE321048, which was a focused study on the Wnt pathway and CLL [40], the list expands to 515. Naturally, a huge number of genes were flagged by DAVID as being in the Wnt pathway (78 total). And a smaller subset was found to be associated with some form of leukemia: AKT1, CTBP1, CTBP2, CEBPA, SMAD4, HDAC1, LEF1, RARA, TCF3, TCF7, TCF7L1, TCFL2, MYC. Looking for common genes between the expanded published and our larger list of 96 threshold, we find KRAS, GRB2, HDAC1, NPM1, TP53, MYC, APP, CTNNB1.

At threshold 32, CTNNB1 is best Betti target once out of 1001 patients, but it is present in the threshold 32 networks 326 times. Keep in mind anything found in the 32 threshold is energetically important. So, we find it in 32.5% of the population as a potentially good target for CLL (at 48 threshold 37.9%; at 64 threshold 44.1%; at 96 threshold 58.9%). CTNNB1 is an important gene involved in CLL. It is also an important node in the Wnt pathway [48].

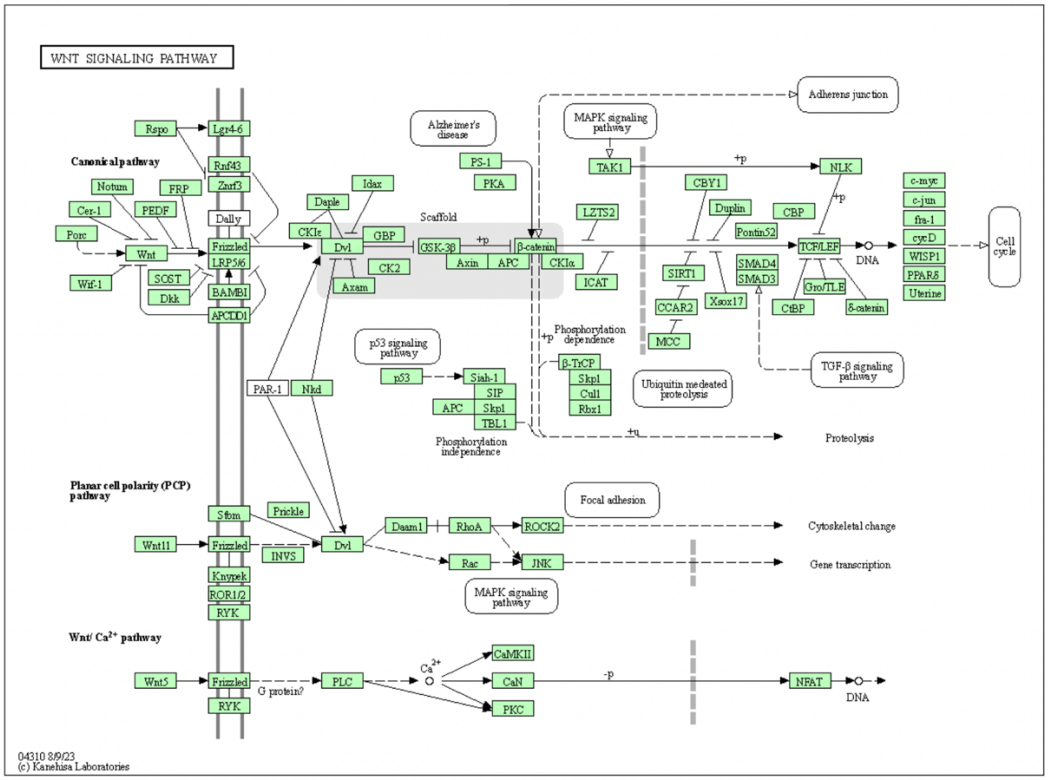


Figure2. Wnt signaling pathway from KEGG https://www.genome.jp/pathway/hsa04310 [46].

Results and discussion: Wnt pathway

It is interesting that so many of the authors of the papers cited above did not find overlap among their gene list and other investigators. There was little overlap between those author's lists until we included the dataset from GSE321048, the Wnt pathway. We speculate that the reason our Gibbs analysis of expression data did not overlap well with other expression data, is that the Gibbs function includes a measure of network entropy (denominator in Equation [1]). Further, many of the genes that are highly expressed, as reported in the literature, are not necessarily mutated based on whole genome sequencing.

Figure 2, shows the Wnt pathway from KEGG. After using the online R-script KEGGraph at Bioconductor it was converted to an edge-list of relevant protein-protein interactions [49].

The resulting edge-list was plotted using Cytoscape 3.7 [50]. The PPI network is shown in Figure 3. Two nodes are highlighted. MYC is highlighted and connected to: LEF1, TCF7, TCF7L1 TCF7L2. MYC, as we will see is an important player in Wnt pathway. Also, CTNNB1 has 24 neighboring interactions and has a betweenness of 0.3155, the highest in this network. It also is an important player in the Wnt pathway.

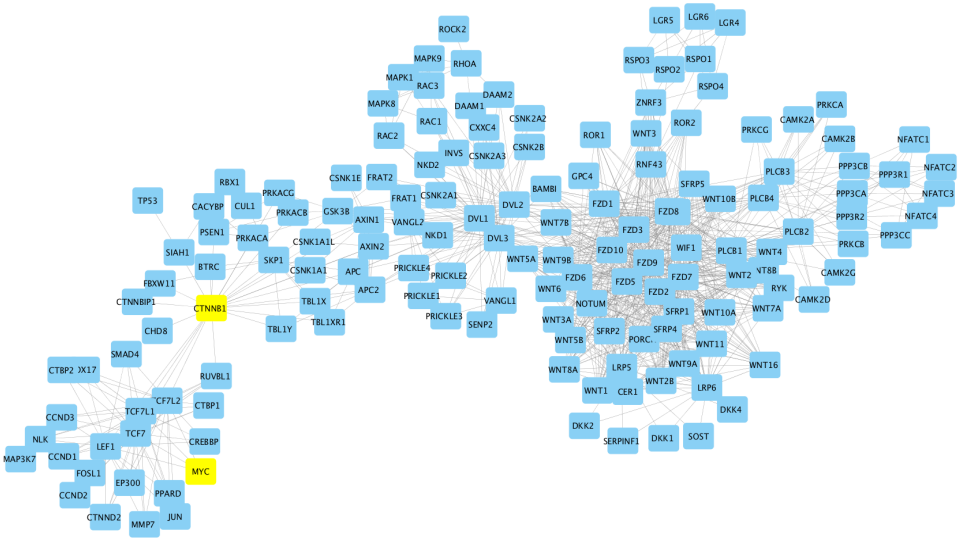


Figure 3. The PPI of Wnt pathway.

As described above, we computed the Betti centrality for the Gibbs homology networks. Figure 4 shows a Pareto chart for the Betti centrality nodes at threshold-48. In our analysis of the 1001 expression samples, CTNNB1 was present as a key Betti centrality node in three samples. Whereas MYC was not present as a key Betti centrality node at threshold-48, but at threshold-32 MYC was present 24 times; 12 times (50%) it was found in dataset GSE30671 which is associated the manuscript by Chuang et al. [41]. This again shows the inconsistency in gene expression values from samples of CLL patients. Of key importance is the fact that RPS15 is a Betti centrality node in 3 patients and RPS15A is a Betti centrality node in 5 patients at threshold-48. This is shown in Figure 4.

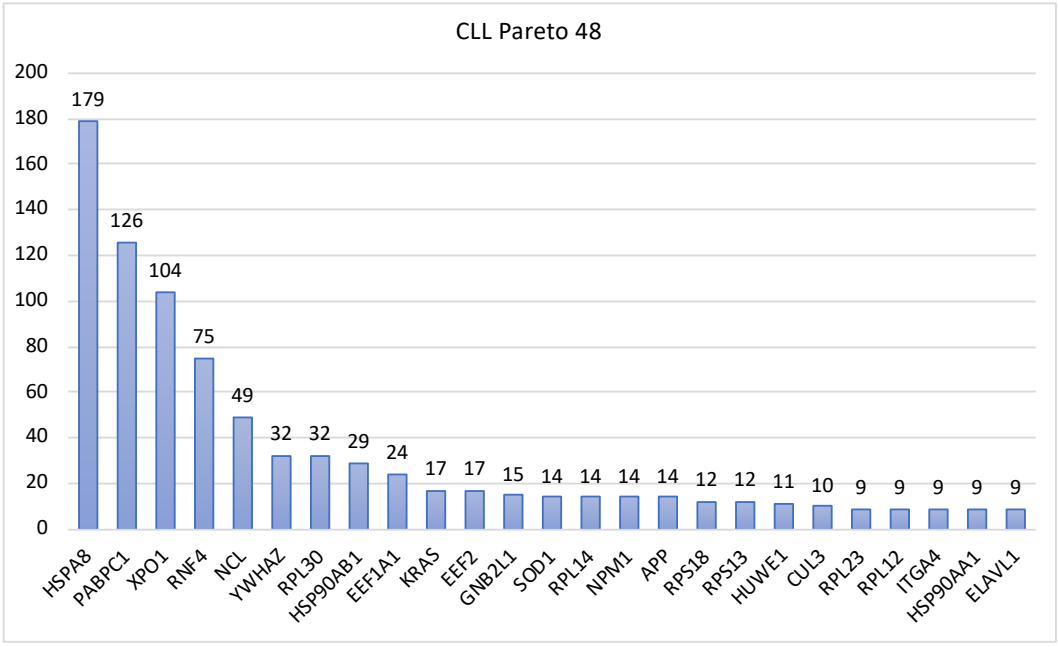


Figure 4. Pareto chart for Betti centrality at Gibbs-homology threshold-48, showing only those with nine or more occurrences.

As shown Figure 3, MYC is an important node in the Wnt pathway. It is directly connected to LEF1, TCF7, TCF7L1, and TCF7L2. Except for LEF1, which is a lymphoid enhancer binding factor, the others are transcription factors. Figure 5 shows a Gibbs homology network at threshold-48 for patient (GSM787065 part of GSE31048 [40]) in which RPS15 is the Betti centrality node. In the network diagram the nodes are in a degree sorted order starting at the bottom with MYC as the highest degree (48) and going around counterclockwise.

RPS15 and MYC have been pulled out of the network for easy locating, and MYC with all its first connections have been highlighted in yellow.

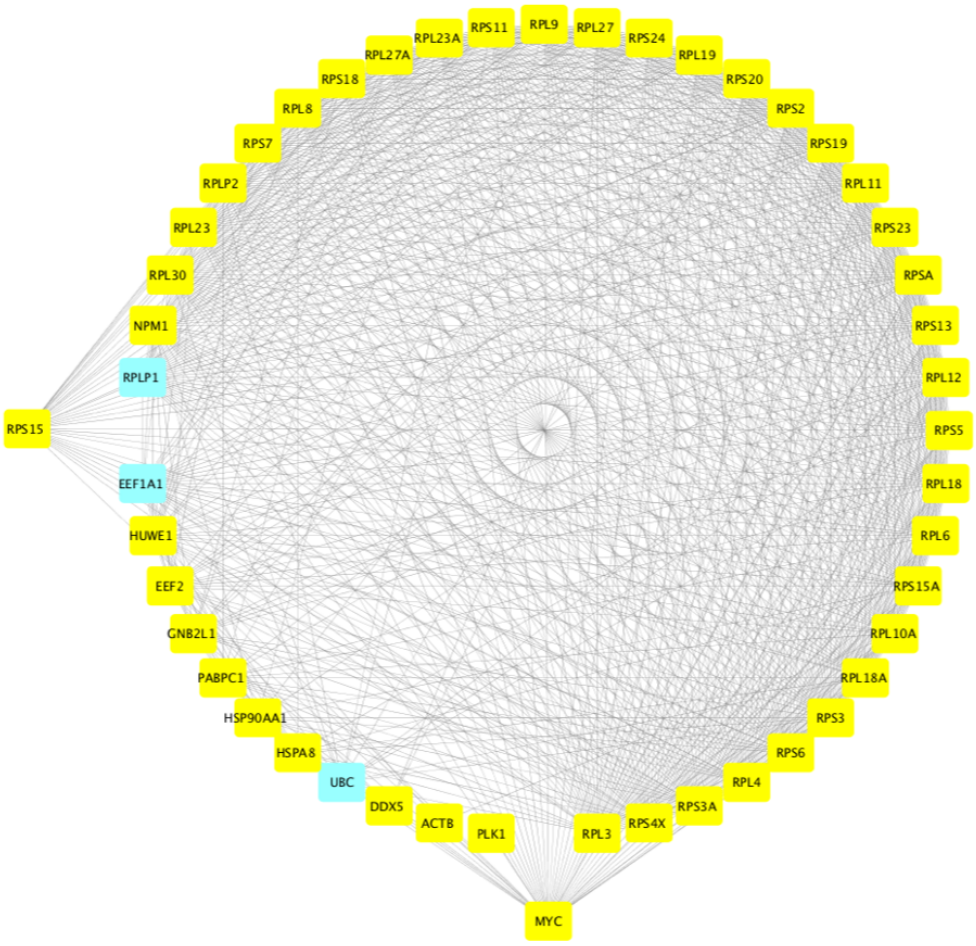


Figure 5. The Gibbs homology network for a patient in which RPS15 has the highest Betti centrality. RPS15 and MYC have been pulled out for easy location. MYC and all its first neighbors are highlighted in yellow.

As we pointed out above, a gene can be mutated, and yet not over-expressed or under-expressed relative to normal. This is likely the main cause for differences in reported transcriptome data from various investigators. What is clear from the literature (e.g. Wang et al.) [40] is that the Wnt pathway is highly important and over-expressed genes in that pathway is often indicative of cancer. MYC is a regulator of ribosome protein synthesis [51] and has been shown to be a key regulator in supporting and maintaining tumorigenesis [52]. For example, Wu, et al. [52] found that inactivation of MYC resulted in some tumors undergoing regression, and mutated RPS15 were identified in almost 20% of CLL patients who relapsed after FCR treatment. These mutations are associated with clinical aggressiveness in CLL along with the mutant RPS15 displaying defective regulation of endogenous p53, which indicates a novel molecular mechanism underlying CLL pathobiology [52]. RPS15 and RPS15A are often overexpressed in CLL [53] and our results confirm this with all (1001 patients) Gibbs-homology subnetworks at threshold-96 showing RPS15 or RPS15A as being an energetically important node.

Conclusions

We have shown in this study that the genes, BTK, NFkB, JAK/STAT, NOTCH1, BCL2, EEF2, among others play a significant role in the support of CLL. Yet some of them are only rarely studied in the literature because they are not strongly expressed; our research confirms this. Just because a gene or two is mutated does not mean it will be strongly expressed. For example, consider the trisomy of chromosome 7 in colorectal cancers [54]. Using genome-wide chromosome conformation capture (Hi-C), RNA sequencing and protein profiling, along with FISH the authors show that indeed chromosome 7 shows a fair number of upregulated genes relative

to healthy cells. However, more strongly they found that chromosome 9 had regions that were very strongly expressed. So, trisomy 7 results in global gene expression changes for colorectal cancer. Might we have something similar going on with CLL patients?

Hi-C analysis is a technique that explores the functional organization of the human nucleome – the folding and unfolding of the chromosomes as the cell goes through its normal processes, including mitosis [55]. Gene topology or arrangement in 3D space affects gene expression, and the 3D topology affects where mutation will occur [56] – [58]. The mathematical technique of processing the Hi-C data involves finding the Laplacian of the "contact matrix" followed by the calculation of the Fiedler vector [59]. The Fiedler vector parses the chromosome into two components: A, the active component, and B, the inactive component. From this it is possible to state, in a high-level way, something about the structure of the chromosome and its function [60]. Since copy number variation closely follows the expression, the function can be deduced from RNA sequencing or transcriptome mapping [61]–[62].

As pointed out above, gene expression found inconsistent sets of genes that were highly expressed and that had high Gibbs-energy. We find the same inconsistent results when we were looking at Hi-C results for CLL patients. Speedy et al. [64] found that BCL2 was strongly implicated in the disease. They also found a disruption at the NFkB-binding site; but, other genes, such as, JAK/STAT, BTK, EEF2 were not mentioned in their manuscript. Beekman et al. [65] found only NOTCH1; Puiggros, et al. [66] found only NOTCH1 and SF3B1 as candidates for high risk of mutation; and Kiefer et al. [67] found NOTCH1 for trisomy 12.

Though the actual causal agent of CLL is not well known, we can speculate that if there is some molecular agent (e.g. herbicide) or an energetic EM signal (e.g. X-ray) it will typically impact the cell only during a specific phase of cell cycle [68]. There are regions of the genome that are more sensitive to alterations due to some specific energy level in the overall molecular network, we call a cell. These mutations are driven by the relevant chemical potential, stereochemistry and Gibbs free energy. We argue that the locations of the relevant genes in the chromosome and the 4D dynamics of the nucleome may suggest a more holistic molecular and cellular approach to understanding CLL and therefore new therapeutic strategies [69]. Building on this notion, the insights from the 4D Nucleome Network [69] elucidate the intricacies of genome organization in space and time. The project underlines the critical role of the genome's three-dimensional organization in gene regulation. In the context of CLL, the spatial dynamics of chromatin can have a profound impact on gene expression patterns, emphasizing the importance of the genome's spatial and temporal dynamics in understanding and potentially treating the disease [70]. Elucidating this idea further, the work conducted by Sawh et al. in 2022 reveals that the eukaryotic genome is a multilayered entity, exhibiting intricate organization levels that range from nucleosomes to larger chromosomal scales [71]. These layers undergo significant remodeling across different tissues and developmental stages in C. elegans. It's noteworthy that advancements in C. elegans research, both imaging-based and sequencing-based, have unveiled the influence of histone modifications, regulatory elements, and broader chromosome configurations in this 4D organization. Specific revelations, such as the physiological implications of topologically associating domains and compartment variability during initial developmental phases, underscore the depth of genome dynamics. These insights provide compelling evidence that understanding such 4D genome organization nuances is crucial for decoding complex diseases like CLL. Interestingly enough, NONE of the genes described in Appendix 1 are in chromosome 13 which often has deletions in about 50% of CLL patients [72]. In Appendix 1 we support our argument for a larger view that CLL genes are widely spread throughout the whole genome and different chromosomes.

In conclusion, our study challenges the conventional single-target paradigm in CLL therapy, advocating for a higher-level, network-oriented strategy. The identification and hierarchical ranking of 20-30 significant proteins, amidst the roughly 20,000 synthesized by the human organism, represent a leap in signal detection and amplification [73]. This nuanced profiling, achieved via a statistical thermodynamics approach, underscores the potential of targeting a selective array of 5-6 network nodes. This selectivity is crucial to mitigate the risk of adverse effects caused by overlapping off-target interactions commonly seen with broader therapeutic targets. The proteins highlighted in our research, notably within the Wnt signaling pathway, are not merely isolated entities but components of a complex network that drives the CLL pathology. Therefore, our proposed method does not end at the identification of these proteins but extends to rank-ordering them in terms of therapeutic relevance. The next step for validating the findings involves experimental assays using siRNA [74] or small molecule inhibitors, which will provide the empirical backbone for our theoretical model. Such an approach may revolutionize the current treatment regimens by transitioning from a one-size-fits-all model to a more customized, patient-specific strategy. This could be especially beneficial given the genetic variability among CLL patients, as indicated by the inconsistent mutation patterns observed in whole-genome sequencing. By incorporating the principles of systems biology and acknowledging the network dynamics of protein interactions, we can begin to envision a more effective, personalized therapeutic landscape for CLL. This, in turn, may pave the way for similar strategies in other cancers, marking a paradigm shift in oncological treatment towards precision medicine.

Appendix 1

Notes on the genomic location of key genes "involved" in CLL (from GeneCards.org)

BTK is found in the X chromosome.

Genomic Locations for BTK Gene

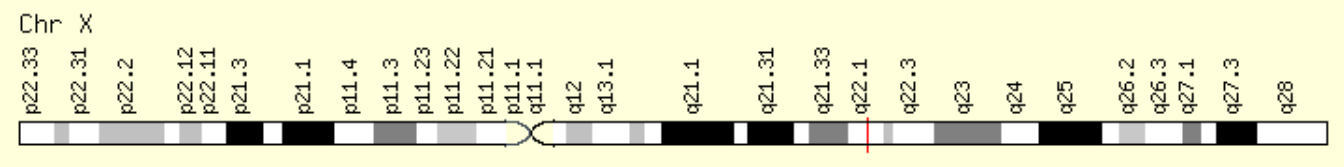
chrX:101,349,447-101,390,796
(GRCh38/hg38)
Size:
41,350 bases
Orientation:
Minus strand

chrX:100,604,435-100,641,212
(GRCh37/hg19)
Size:
36,778 bases
Orientation:
Minus strand

Genomic View for BTK Gene

Genes around BTK on UCSC Golden Path with GeneCards custom track

Cytogenetic band:
Xq22.1 by HGNC
Xq22.1 by Entrez Gene
Xq22.1 by Ensembl BTK Gene in genomic location: bands according to Ensembl, locations according to GeneLoc (and/or Entrez Gene and/or Ensembl if different)



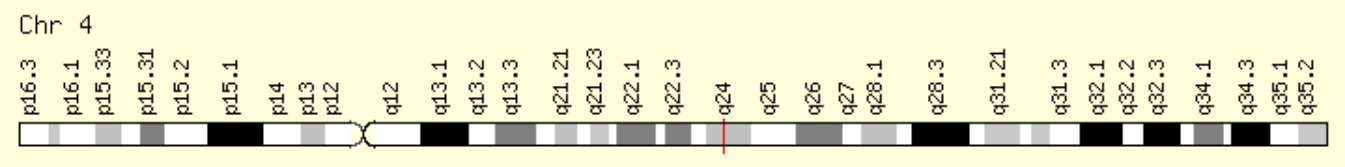
NFKB1 is in chromosome 4

Genomic Locations for NFKB1 Gene

chr4:102,501,329-102,617,302
(GRCh38/hg38)
Size:
115,974 bases
Orientation:
Plus strand

chr4:103,422,486-103,538,459
(GRCh37/hg19)
Size:
115,974 bases
Orientation:
Plus strand

Genomic View for NFKB1 Gene



NFKB2 is on chromosome 10

Genomic Locations for NFKB2 Gene

chr10:102,394,110-102,402,529
(GRCh38/hg38)
Size:
8,420 bases
Orientation:
Plus strand
chr10:104,153,867-104,162,281
(GRCh37/hg19)
Size:
8,415 bases
Orientation:
Plus strand

Genomic View for NFKB2 Gene

Genes around NFKB2 on UCSC Golden Path with GeneCards custom track

JAK1 is on chromosome 1

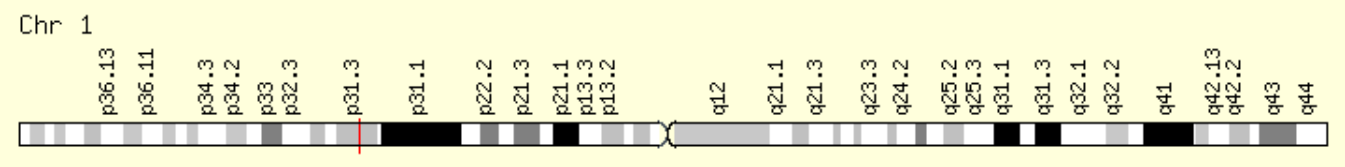
Genomic Locations for JAK1 Gene

chr1:64,833,223-65,067,754
(GRCh38/hg38)
Size:
234,532 bases
Orientation:
Minus strand

chr1:65,298,906-65,432,187
(GRCh37/hg19)
Size:
133,282 bases
Orientation:
Minus strand

Genomic View for JAK1 Gene

Genes around **JAK**1 on UCSC Golden Path with GeneCards custom track



JAK2 is on chromosome 9

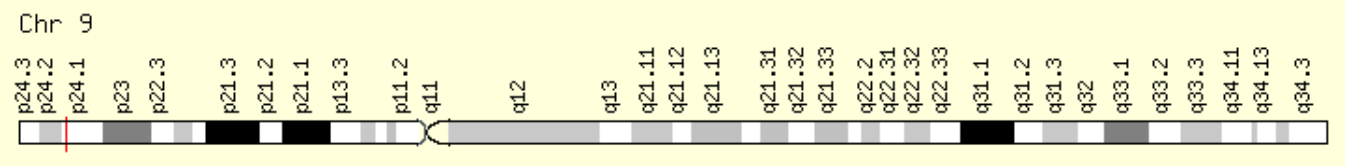
Genomic Locations for JAK2 Gene

chr9:4,984,390-5,128,183
(GRCh38/hg38)
Size:
143,794 bases
Orientation:
Plus strand

chr9:4,985,033-5,128,183
(GRCh37/hg19)
Size:
143,151 bases
Orientation:
Plus strand

Genomic View for JAK2 Gene

Genes around **JAK**2 on UCSC Golden Path with GeneCards custom track



JAK3 is on chromosome 19

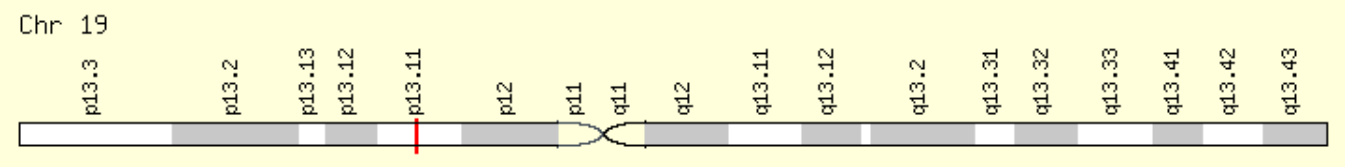
Genomic Locations for JAK3 Gene

Genomic Locations for **JAK**3 Gene
chr19:17,824,780-17,848,071
(GRCh38/hg38)
Size:
23,292 bases
Orientation:
Minus strand

chr19:17,935,589-17,958,880
(GRCh37/hg19)
Size:
23,292 bases
Orientation:
Minus strand

Genomic View for JAK3 Gene

Genes around **JAK**3 on UCSC Golden Path with GeneCards custom track



STAT1 is on chromosome 2

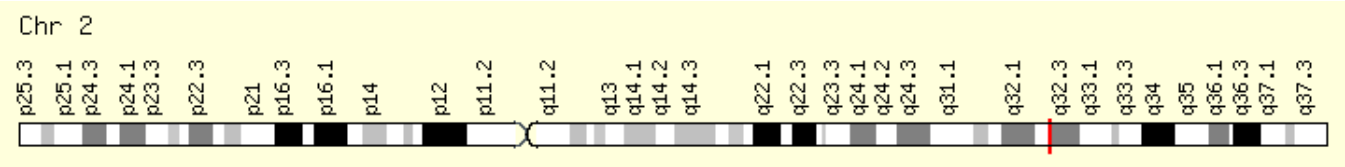
Genomic Locations for STAT1 Gene

Genomic Locations for STAT1 Gene
chr2:190,964,358-191,020,960
(GRCh38/hg38)
Size:
56,603 bases
Orientation:
Minus strand

chr2:191,829,084-191,885,686
(GRCh37/hg19)
Size:
56,603 bases
Orientation:
Minus strand

Genomic View for STAT1 Gene

Genes around STAT1 on UCSC Golden Path with GeneCards custom track



STAT2 is on chromosome 12

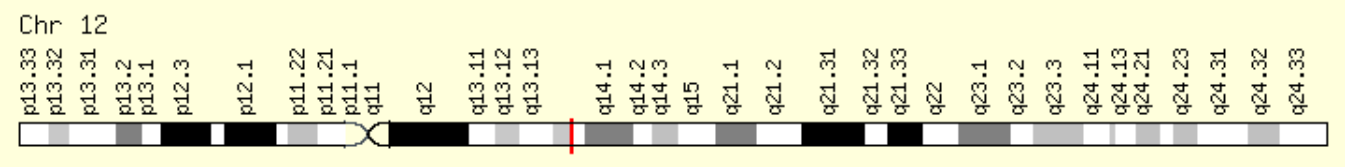
Genomic Locations for STAT2 Gene

Genomic Locations for STAT2 Gene
chr12:56,341,597-56,360,253
(GRCh38/hg38)
Size:
18,657 bases
Orientation:
Minus strand

chr12:56,735,381-56,754,037
(GRCh37/hg19)
Size:
18,657 bases
Orientation:
Minus strand

Genomic View for STAT2 Gene

Genes around STAT2 on UCSC Golden Path with GeneCards custom track



STAT3 is on chromosome 17

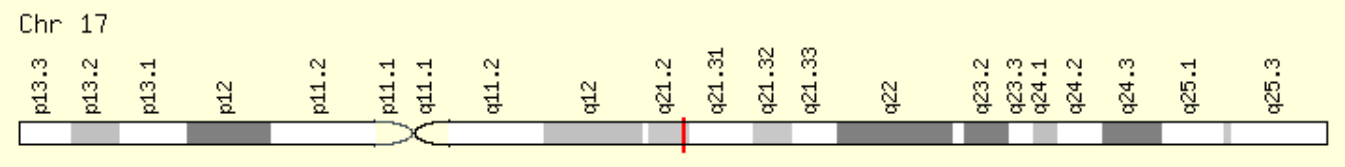
Genomic Locations for STAT3 Gene

Genomic Locations for STAT3 Gene
chr17:42,313,324-42,388,568
(GRCh38/hg38)
Size:
75,245 bases
Orientation:
Minus strand

chr17:40,465,342-40,540,586
(GRCh37/hg19)
Size:
75,245 bases
Orientation:
Minus strand

Genomic View for STAT3 Gene

Genes around STAT3 on UCSC Golden Path with GeneCards custom track



NOTCH1 is on chromosome 9

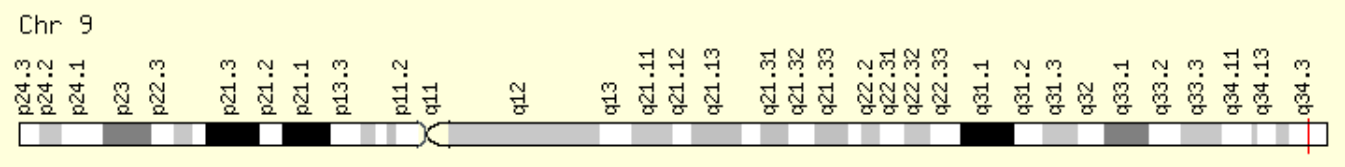
Genomic Locations for NOTCH1 Gene

Genomic Locations for NOTCH1 Gene
chr9:136,494,433-136,546,048
(GRCh38/hg38)
Size:
51,616 bases
Orientation:
Minus strand

chr9:139,388,896-139,440,314
(GRCh37/hg19)
Size:
51,419 bases
Orientation:
Minus strand

Genomic View for NOTCH1 Gene

Genes around NOTCH1 on UCSC Golden Path with GeneCards custom track



BCL2 is on chromosome 18

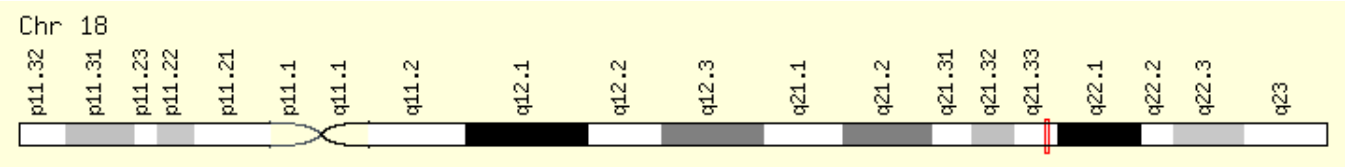
Genomic Locations for BCL2 Gene

Genomic Locations for BCL2 Gene
chr18:63,123,346-63,320,128
(GRCh38/hg38)
Size:
196,783 bases
Orientation:
Minus strand

chr18:60,790,579-60,987,361
(GRCh37/hg19)
Size:
196,783 bases
Orientation:
Minus strand

Genomic View for BCL2 Gene

Genes around BCL2 on UCSC Golden Path with GeneCards custom track



EEF2 is on chromosome 19

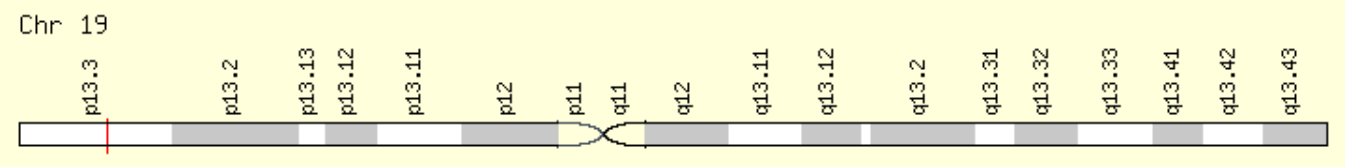
Genomic Locations for EEF2 Gene

Genomic Locations for EEF2 Gene
chr19:3,976,056-3,985,463
(GRCh38/hg38)
Size:
9,408 bases
Orientation:
Minus strand

chr19:3,976,054-3,985,467
(GRCh37/hg19)
Size:
9,414 bases
Orientation:
Minus strand

Genomic View for EEF2 Gene

Genes around EEF2 on UCSC Golden Path with GeneCards custom track



ABL1 is in chromosome 9

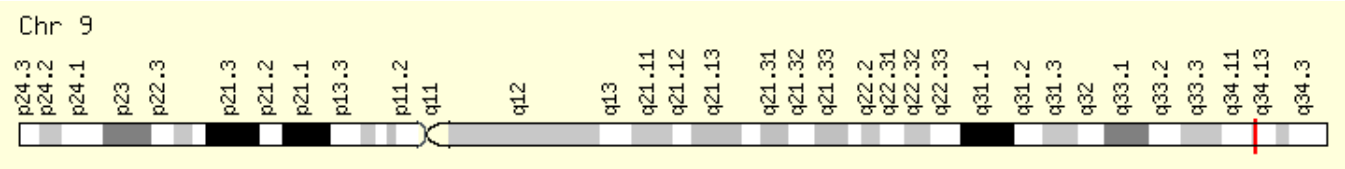
Genomic Locations for ABL1 Gene

Genomic Locations for ABL1 Gene
chr9:130,713,016-130,887,675
(GRCh38/hg38)
Size:
174,660 bases
Orientation:
Plus strand

chr9:133,589,268-133,763,062
(GRCh37/hg19)
Size:
173,795 bases
Orientation:
Plus strand

Genomic View for ABL1 Gene

Genes around ABL1 on UCSC Golden Path with GeneCards custom track



BCR3 (BCRP3) is in chromosome 22

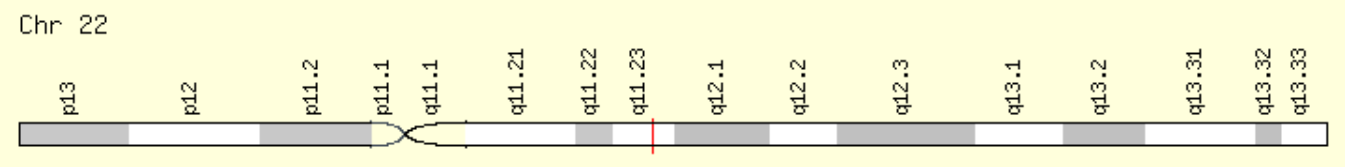
Genomic Locations for BCRP3 Gene

Genomic Locations for BCRP3 Gene
chr22:24,632,915-24,653,360
(GRCh38/hg38)
Size:
20,446 bases
Orientation:
Plus strand

chr22:25,028,882-25,049,327
(GRCh37/hg19)
Size:
20,446 bases
Orientation:
Plus strand

Genomic View for BCRP3 Gene

Genes around BCRP3 on UCSC Golden Path with GeneCards custom track



SF3B1 is in chromosome 2

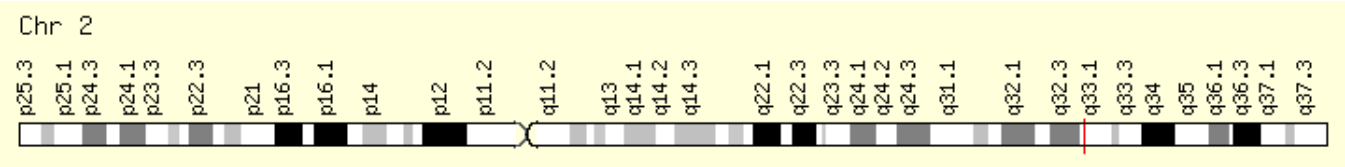
Genomic Locations for SF3B1 Gene

Genomic Locations for SF3B1 Gene
chr2:197,388,515-197,435,091
(GRCh38/hg38)
Size:
46,577 bases
Orientation:
Minus strand

chr2:198,254,508-198,299,815
(GRCh37/hg19)
Size:
45,308 bases
Orientation:
Minus strand

Genomic View for SF3B1 Gene

Genes around SF3B1 on UCSC Golden Path with GeneCards custom track



ZAP70 is on chromosome 2

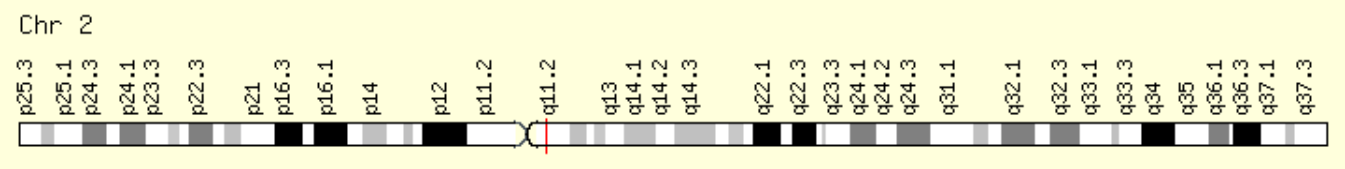
Genomic Locations for ZAP70 Gene

Genomic Locations for ZAP70 Gene
chr2:97,713,560-97,744,327
(GRCh38/hg38)
Size:
30,768 bases
Orientation:
Plus strand

chr2:98,330,023-98,356,325
(GRCh37/hg19)
Size:
26,303 bases
Orientation:
Plus strand

Genomic View for ZAP70 Gene

Genes around ZAP70 on UCSC Golden Path with GeneCards custom track



MYC is on chromosome 8

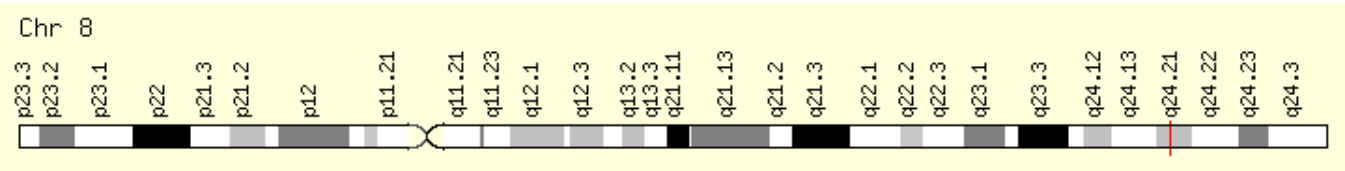
Genomic Locations for MYC Gene

Genomic Locations for MYC Gene
chr8:127,735,434-127,742,951
(GRCh38/hg38)
Size:
7,518 bases
Orientation:
Plus strand

chr8:128,747,680-128,753,680
(GRCh37/hg19)
Size:
6,001 bases
Orientation:
Plus strand

Genomic View for MYC Gene

Genes around MYC on UCSC Golden Path with GeneCards custom track



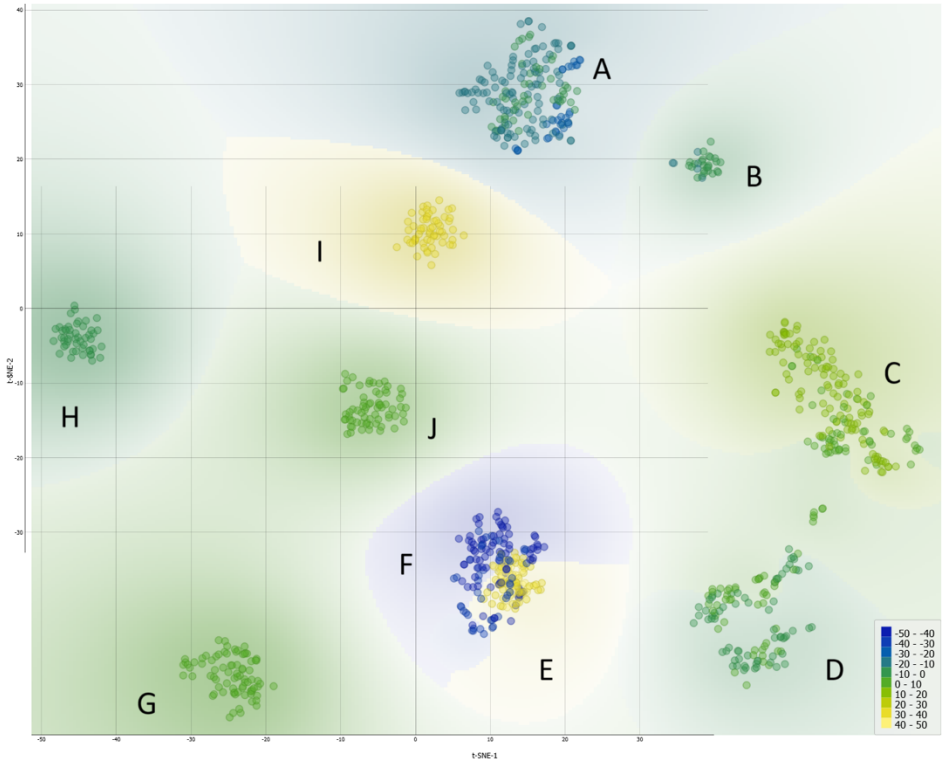
Appendix 2. Union of the Published Gene List Reviewed in this Study

ABCA2	CDC26	FAM174B	KCNK1	NPAS1	REPS1	TCF3
ABCA7	CDS1	FAM24A	KIAA0984	NPR1	RFC5	TCF4
ABCE1	CDT1	FARP1	KIAA1009	NR2C2	RHOD	TCF7
ABI3	CEBP8	FBN1	KIAA1529	NR3C2	RHOQ	TCF7L1
ABRA	CEBPA	FBXW11	KIAA1545	NRIP1	RHOU	TCF7L2
ACVR1	CEP68	FBXW2	KLF6	NSD1	RIMKLB	TGFB1I1
ADAM17	CER1	FBXW4	KLHL23	NUDT1	RIPK5	TGFBR3
ADAM19	CHAF1A	FGF4	KREMEN1	NUP107	RNFT2	TLE1
ADAM29	CHD8	FGL2	KREMEN2	NUP210	ROR1	TLE2
ADAMTS7	CHPT1	FKBP4	KRT18	NUP62CL	ROR2	TLE2
AES	CLDN3	FLJ10357	KRT7	ONECUT1	RPP25	TLE3
AGPAT2	CLDN7	FLJ20125	KRT8	OR2F1	RPRC1	TMC5
AGPAT4	CLEC2B	FLJ20160	L3MBTL4	OR7E19P	RRAGC	TMEM126B
AKR1C2	CNBD1	FLJ20489	LAMA5	P2RX1	RSPO4	TMEM132A
AKT1	CNOT1	FLJ20674	LDOC1	PAICS	RUVBL1	TNFRSF7
ALDH1A2	COL7A1	FLJ23556	LEF1	PCDHGA1	RUVBL2	TNFSF11
ALLC	CPM	FLJ40759	LILRB3	PCDHGA11	RYK	TNFSF13
ANAPC5	CPNE7	FRAT1	LLGL2	PCDHGA3	SAP30	TNS3
ANKRD57	CPT1A	FRAT2	LPL	PCDHGC3	SEC11A	TP73L
ANXA2	CPZ	FRMD4A	LRCH4	PCP4	SEL1L	TPCN1
ANXA3	CREB3	FRMPD1	LRP12	PDE11A	SEMA4B	TPST2
APC	CREBBP	FRZB	LRP5	PDE4DIP	SEMA4D	TRAK2
APC2	CRHR1	FSCB	LRP5L	PEA15	SEMA5A	TRAPPC6A
APOB	CRY1	FUT8	LRP6	PEX5	SENP2	TRERF1
APP	CRYBB2P1	FZD1	LRRFIP2	PFTK1	SEP10	TRIM2
APRT	CSF1	FZD2	LSM4	PGLS	SEPP1	TRIM43
ARHGAP8	CSNK1A1	FZD3	LTK	PHEX	SERPINB2	TRIM9

ARHGEF17	CSNK1D	FZD4	LUM	PIGC	SERPINF1	TRSPAP1
ARSD	CSNK1E	FZD5	MAGED2	PIGF	SEZ6L	UCHL1
ASPA	CSNK1G1	FZD6	MAK	PIN1	SFN	UGCG
ASPSCR1	CSNK2A	FZD7	MAN2B1	PITX2	SFRP1	UHRF1
ATOX1	CSNK2A1	FZD8	MANEA	PKP2	SFRP2	UNC93B1
ATP10B	CSPG6	FZD9	MAP3K5	PKP3	SFRP4	USP7
ATP13A1	CTAG2	GALNT2	MAP3K7IP1	PLAA	SFRP5	VAMP5
ATP2A3	CTBP1	GAST	MAP4K1	PLCXD2	SIGLEC8	VASH1
ATP6V0E	CTBP2	GBF1	MAP7	PLD1	SIRPG	VASP
AVP	CTCF	GEMIN4	MARCKS	PLEKHG4	SIX6	VPS37B
AXIN1	CTNNB1	GGTLA4	MBOAT1	PNCK	SKP2	WDFY4
AXIN2	CTNNBIP1	GHRHR	MCM2	POLE2	SLA2	WDR62
B3GNT4	CXXC4	GINS2	MCM3AP	POLR2G	SLC12A9	WIF1
B4GALT2	CYB5R1	GJB3	MCOLN3	PORCN	SLC25A20	WIPI1
B4GALT6	DAAM1	GNG8	MCP	PPL	SLC25A23	WISP1
BANF2	DCI	GNGT2	MDFIC	PPM1A	SLC27A2	WNT1
BCAT1	DCLK2	GNPTAB	ME2	PPM1L	SLC9A3R1	WNT10A
BCL9	DDX23	GOT2	MED9	PPP2CA	SMAD2	WNT11
BRD7	DEGS1	GPLD1	MGC9913	PPP2R1A	SMAD3	WNT16
BTN3A3	DENND1C	GPR177	MICAL1	PPP2R3A	SMAD4	WNT2
BTRC	DIP2C	GSK3A	MINK1	PPP2R5B	SMARCA4	WNT2B
C10orf75	DIXDC1	GSK3B	MKLN1	PPP4R2	SMPD1	WNT3
C10orf86	DKK1	Gene	MMP14	PRELP	SNAI2	WNT4
C12orf10	DKK2	HBP1	MMP2	PRICKLE1	SOSTDC1	WNT5A
C16orf24	DKK3	HDAC1	MRPL41	PRKCA	SOX17	WNT5B
C16orf33	DKK4	HEG1	MRPS33	PRKCB	SOX4	WNT6
C16orf45	DLEC1	HGF	MTM1	PRKRIR	SPARC	WNT7A
C19orf21	DMD	HIST1H1C	MUC5B	PRPF40B	SPATA1	WNT7B

C1orf77	DOCK9	HLA-DQA1	MYC	PRR4	SPG20	WNT8A
C20orf42	DSC1	HNF1A	MYO15B	PSD3	SPHK1	WNT9A
C20orf67	DSP	HNRPLL	MYO5C	PSEN1	SPINT2	WSB2
C20orf74	DUS2L	HOM-TES-103	MZF1	PSMD8	SRPK2	YOEL1
CAL-COCO1	DVL1	HOMER3	NAPB	PTCH1	SRPX	ZAP70
CALD1	DVL2	HOXD3	NAV3	PTGER3	ST5	ZBTB33
CALU	DVL3	HSP90B1	NBEAL2	PTGES3	SUPT3H	ZCCHC11
CAPG6	DYNLL1	ICA1L	NBPF8	PTPLAD2	SUSD3	ZFHX1B
CARS2	EDG5	IGF1R	NDP	PTPN6	SYCP1	ZFP64
CASKIN2	EGR3	IGF2R	NDUFC2	PTPRG	SYDE1	ZNF135
CBWD2	EHMT1	IL13RA2	NDUFV1	PYGO1	SYNC1	ZNF451
CBY1	ELF1	INVS	NEBL	PYGO2	SYT5	ZNF578
CCDC24	ENG	IRAK2	NIP-SNAP3B	RAB26	TACSTD1	ZNF667
CCNB2	EP300	JAG2	NKD1	RAB7L1	TBC1D1	ZNF787
CCT4	EPHA3	JTV1	NKD2	RARA	TBC1D22A	
CCT7	EPS15L1	JUP	NLK	RBM35B	TBC1D2B	
CCT8	EXOSC1	KALRN	NLRC3	RCC1	TBP	
CD58	EXPH5	KCNJ16	NOP17	RCN1	TCEB3	

Appendix 3. Integrative Analysis of CLL by t-SNE Visualization



t-Distributed Stochastic Neighbor Embedding (t-SNE) Analysis of CLL Samples and Subgroups. The visualization displays a non-linear dimensionality reduction of the complex gene expression data, with each point representing individual samples. The layout highlights the nuanced relationships and 10 clusters labeled from A to J within the CLL dataset consisted of the 1001 patients, uncovering uncaptured subtleties through the network analysis. Samples included in the dataset are either diagnosed CLL patients (sick) or wild-type patients (normal) without CLL.

- Cluster A encapsulates a significant cohort focusing on the Wnt signaling pathway, a key player in CLL pathogenesis, with a total of 179 patients. Within this cluster, 21 are wild type, while 158 are CLL patient samples, all derived from the GSE31048 dataset. This dataset offers a unique look at both normal and CLL-affected B cells, allowing for a direct comparison of Wnt pathway gene expression and Wnt-regulated gene expression. The marked difference in numbers between the normal (12 and 9, respectively) and sick (149 and 9, respectively) groups underscores the aberrant expression within CLL-affected B cells, highlighting the pathway's prominence in these cellular states. The study's in-depth focus on the Wnt pathway is well-founded, as it is pivotal in cellular processes that are often disrupted in CLL, thus potentially illuminating new therapeutic avenues.
- Cluster B includes a smaller, yet focused subset of 42 patients, of which 12 are wild type and 30 are CLL patients. This cluster continues the examination of the Wnt pathway's role in CLL as part of the GSE31048 study, indicative of a unique or divergent role of Wnt signaling in this subset. It is particularly noteworthy that this dataset includes expression data from CLL B cells with and without Wnt3a treatment, providing insights into the pathway's functionality and potential for targeted therapies. The comparative analysis of Wnt gene expression between the normal and CLL B cells offers additional evidence for the pathway's critical role in the disease process.

- Cluster C, with 136 patients, 24 normal and 112 sick, merges data from two distinct studies, GSE10139 and GSE50006, providing a broader scope by incorporating a genomic approach from GSE10139 to improve prognosis and therapeutic response predictions, and juxtaposes this with expression data from CLL tumors and healthy donor B cells from GSE50006. The inclusion of CLL-blood samples enriches the dataset, illustrating the heterogeneity within CLL and potentially reflecting different disease phases or subtypes. The blend of these datasets furnishes a more comprehensive understanding of the disease, highlighting the heterogeneity of CLL and reinforcing the necessity of personalized medicine approaches.
- Cluster D presents a cohort of 152 patients, predominantly sick (144) with a small representation of normal B cells (8), combining data from GSE10139 and GSE50006. This distribution, primarily composed of CLL and CLL-blood samples, continues to emphasize the genetic and expression-level diversity found in CLL, supporting the need for in-depth analysis to discern the nuances of the disease's progression and the potential response to treatments.
- Cluster E is a homogeneous group consisting entirely of 100 sick patients from the GSE49896 dataset. This study spotlights the microRNA-150's influence on B-Cell Receptor signaling by modulating GAB1 and FOXP1 gene expressions, which are implicated in CLL. MicroRNAs are crucial post-transcriptional regulators, and their role in CLL adds an additional layer to understanding the disease's complexity and potential intervention points.
- In Cluster F, 130 CLL patients from the GSE39671 dataset are studied, all of whom have undergone treatment. The data represent a temporal progression, with sampling times to first treatment recorded, allowing for an exploration of the disease's evolution over time. The dataset's analysis provides prognostic subnetworks which can help predict disease progression and highlight the converging pathways in CLL, opening new avenues for tailored treatments.
- Cluster G, comprising 75 CLL patients from the GSE69034 study, delves into the gene expression profiles linked with the MYD88 L265P mutations in conjunction with IGHV mutation status. The presence of the MYD88 L265P mutation, a notable variant found within the MYD88 gene that encodes a key adaptor protein in the Toll-like receptor and IL-1 receptor pathways, has been tied to specific prognostic outcomes in CLL. This mutation is known to activate downstream signaling pathways aberrantly, which can contribute to the uncontrolled proliferation of B cells characteristic of CLL. The dataset's inclusion in the study facilitates a detailed investigation into the mutation's role and its pathway associations in CLL, offering a potential explanation for the varying responses to treatment observed in patient populations. By analyzing the gene expression patterns influenced by the MYD88 L265P mutation alongside the IGHV mutation status, a well-established prognostic marker in CLL, it unravels the complex interplay between genetic aberrations and their impact on the disease's clinical course. The correlation between MYD88 L265P mutations and factors such as treatment resistance, disease progression, and overall survival can be assessed. This is particularly crucial, as the mutation's impact on signaling pathways may suggest new therapeutic targets or strategies for intervention. Groundbreaking biomarkers are likely to be identified for early detection and prognosis by understanding the biological context in which these mutations operate, while also highlighting the therapeutic relevance of targeting the MYD88 pathway in certain subsets of CLL patients, which implies the importance of precision medicine in the management of CLL. Based on the insights into the specific mutations driving the disease in individual patients, therapies can be customized to target these genetic abnormalities more effectively. In the case of MYD88 L265P, its presence could signify a need for

targeted inhibitors that can mitigate its downstream effects, thereby introducing a new dimension to personalized CLL treatment paradigms.

- Cluster H is a cohort of 84 CLL patients from GSE28654, all carrying the IgVHMT mutation and exhibiting negative ZAP-70 expression. The absence of ZAP-70 expression, a kinase linked to CLL, together with the mutational profile, provides a critical connection for investigation. This relationship implicates the substantial impact of the mutation on CLL's clinical progression and pinpoints the need for detailed genetic analysis in crafting specialized treatments.
- In Cluster I, 28 sick patients from GSE28654 are categorized by the presence of the IgVHUM mutation and positive ZAP-70 expression, helping to understand the disease's heterogeneity since ZAP-70 positivity is often linked with a more aggressive CLL form. The combination of mutational status and ZAP-70 expression levels provides valuable prognostic information.
    - The expression of ZAP-70 in CLL and its relevance as a molecular marker is particularly illuminating. For Cluster H, the collective profile of CLL patients characterized by the IgVHMT mutation yet displaying an absence of ZAP-70 expression represents a subset where traditional prognostic markers may predict a more favorable clinical course. In the broader landscape of our findings, this cluster could suggest that ZAP-70's negativity may reflect a less aggressive form of CLL, where the malignant B cells might not engage in the same signaling pathways that are characteristic of more virulent variants. Consequently, these insights bolster the argument for personalized therapeutic approaches, enabling clinicians to tailor treatments to the specific molecular makeup presented by individual CLL cases. Conversely, patients in Cluster I, characterized by the IgVHUM mutation concomitant with positive ZAP-70 expression, suggest a more aggressive manifestation of the disease. This association aligns with the understanding that ZAP-70 positivity mirrors the behavior of unmutated IgVH status, commonly linked to a robust disease progression and a less favorable response to conventional therapies. Here, ZAP-70 serves not just as a prognostic marker but potentially as a therapeutic target, whereby modulation of its expression or function could impact CLL cell survival. This reiterates the substantial role that ZAP-70 plays in CLL. It acts as a bifurcation point in the disease's prognostic roadmap, where its expression could either denote a need for more aggressive treatment or suggest a less intensive therapeutic course. The interplay of ZAP-70 with IgVH mutation status, as demonstrated in our clusters, provides a clearer understanding of disease heterogeneity and patient stratification. The overall results of the study thus advocate for the integration of ZAP-70 status into prognostic models and therapeutic decision-making algorithms, emphasizing its contribution not only to prognostication but potentially to the development of targeted CLL therapies.
- Cluster J, mirroring Cluster G, includes another set of 75 CLL patients from the GSE69034 dataset, indicating the significant role of MYD88 L265P mutations in CLL, providing a robust dataset for the exploration of mutation-associated gene expression patterns and their prognostic significance.

The heterogeneity of gene mutations across CLL patients underscores the intricate complexity of this malignancy, accentuating the necessity for individualized therapeutic strategies. The disparities unearthed by t-SNE analysis manifest in the distinct molecular signatures differentiating normal B cells from CLL-B cells, which reflect divergent evolutionary trajectories within the disease's progression. Notably, the aberrant expression of Wnt pathway genes in CLL cells, as revealed by our cluster analysis, pinpoints this pathway's pivotal role in CLL pathobiology. The presence of specific gene expressions within clusters, particularly those highlighted by the t-SNE method (clusters A and B), points to the pathway's disrupted regulation, which is suggestive of patient-specific disease

mechanisms that contribute to CLL's heterogeneity. Simultaneously, the funding emphasizes that alterations in the Wnt signaling pathway are not universally present but vary among patients, reinforcing the pathway's contribution to the disease complexity. The therapeutic potential of targeting Wnt pathway proteins is corroborated by their identified roles in vital cellular functions, with their significance accentuated by Betti number estimates, which propose these proteins as central players in CLL's pathogenesis rather than inconsequential elements. Such insights solidify the imperative for a more comprehensive, multi-scalar study from cellular to genomic dimensions to forge ahead with personalized treatments for CLL.

Author Contributions: GLK suggested the study. EAR and GLK selected the datasets. EAR conducted the computational study and wrote the first draft with input from GLK. HTS supported EAR and provided consulting on the project. GP contributed the study on drugs and rewrote the paper with the assistance of JZ. HH performed the clustering analysis. JAT contributed to the cancer drug analysis and wrote the paper with GP and EAR.

Funding: JZ, HH, EAR and HTS acknowledge no external funding for this work.

Data Availability Statement: The data for the CLL study is from the GEO Website and more specifically are the following datasets: GSE10139, GSE28654, GSE31048, GSE39671, GSE49896, GSE50006, and GSE69034.

Acknowledgments: JAT acknowledges the funding support received from NSERC (Canada) for this project.

Conflicts of Interest: We declare no conflicts of interest.

References

1. NHS. "Overview - Chronic lymphocytic leukaemia". Feb. 07, 2022. Available from: https://www.nhs.uk/conditions/chronic-lymphocytic-leukaemia/ [Accessed Jul. 30, 2022].

2. Zhang S, Kipps TJ. The pathogenesis of chronic lymphocytic leukemia. Annu Rev Pathol. 2014;9:103-18. doi: 10.1146/annurev-pathol-020712-163955. Epub 2013 Aug 26. PMID: 23987584; PMCID: PMC4144790.

3. Trojani A, Di Camillo B, Tedeschi A, Lodola M, Montesano S, Ricci F, Vismara E, Greco A, Veronese S, Orlacchio A, Martino S, Colombo C, Mura M, Nichelatti M, Colosimo A, Scarpati B, Montillo M, Morra E. Gene expression profiling identifies ARSD as a new marker of disease progression and the sphingolipid metabolism as a potential novel metabolism in chronic lymphocytic leukemia. Cancer Biomark. 2011-2012;11(1):15-28. doi: 10.3233/CBM-2012-0259. PMID: 22820137.

4. Hallek M, Shanafelt TD, Eichhorst B. Chronic lymphocytic leukaemia. Lancet. 2018 Apr 14;391(10129):1524-1537. doi: 10.1016/S0140-6736(18)30422-7. Epub 2018 Feb 21. PMID: 29477250.

5. M. Kamdar, "CLL Society," Prognostic Factors in CLL, Sep. 27, 2017. https://cllsociety.org/2017/09/prognostic-factors-cll/ [accessed Jul. 31, 2022].

6. Mukkamalla SKR, Taneja A, Malipeddi D, Master SR. Chronic Lymphocytic Leukemia. [Updated 2023 Jan 15]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK470433/

7. Cohen JA, Bomben R, Pozzo F, Tissino E, Härzschel A, Hartmann TN, Zucchetto A, Gattei V. An Updated Perspective on Current Prognostic and Predictive Biomarkers in Chronic Lymphocytic Leukemia in the Context

of Chemoimmunotherapy and Novel Targeted Therapy. Cancers (Basel). 2020 Apr 7;12(4):894. doi: 10.3390/cancers12040894. PMID: 32272636; PMCID: PMC7226446.

8. Redaelli A, Laskin BL, Stephens JM, Botteman MF, Pashos CL. The clinical and epidemiological burden of chronic lymphocytic leukaemia. Eur J Cancer Care (Engl). 2004 Jul;13(3):279-87. doi: 10.1111/j.1365-2354.2004.00489.x. PMID: 15196232.

9. Janovská P, Bryja V. Wnt signalling pathways in chronic lymphocytic leukaemia and B-cell lymphomas. Br J Pharmacol. 2017 Dec;174(24):4701-4715. doi: 10.1111/bph.13949. Epub 2017 Aug 30. PMID: 28703283; PMCID: PMC5727250.

10. Bewarder M, Stilgenbauer S, Thurner L, Kaddu-Mulindwa D. Current Treatment Options in CLL. Cancers (Basel). 2021 May 19;13(10):2468. doi: 10.3390/cancers13102468. PMID: 34069354; PMCID: PMC8158749.

11. Yang J, Nie J, Ma X, Wei Y, Peng Y, Wei X. Targeting PI3K in cancer: mechanisms and advances in clinical trials. Mol Cancer. 2019 Feb 19;18(1):26. doi: 10.1186/s12943-019-0954-x. PMID: 30782187; PMCID: PMC6379961.

12. Brown JR. The PI3K pathway: clinical inhibition in chronic lymphocytic leukemia. Seminars in Oncology. 2016 Apr;43(2):260-264. doi: 10.1053/j.seminoncol.2016.02.004. Epub 2016 Feb 8. PMID: 27040704.

13. Gentile M, Petrungaro A, Uccello G, Vigna E, Recchia AG, Caruso N, Bossio S, De Stefano L, Palummo A, Storino F, Martino M, Morabito F. Venetoclax for the treatment of chronic lymphocytic leukemia. Expert Opin Investig Drugs. 2017 Nov;26(11):1307-1316. doi: 10.1080/13543784.2017.1386173. Epub 2017 Oct 9. PMID: 28972395.

14. Nagasaka M, Potugari B, Nguyen A, Sukari A, Azmi AS, Ou SI. KRAS Inhibitors- yes but what next? Direct targeting of KRAS- vaccines, adoptive T cell therapy and beyond. Cancer Treat Rev. 2021 Dec;101:102309. doi: 10.1016/j.ctrv.2021.102309. Epub 2021 Oct 21. PMID: 34715449.

15. Martorana F, Motta G, Pavone G, Motta L, Stella S, Vitale SR, Manzella L, Vigneri P. AKT Inhibitors: New Weapons in the Fight Against Breast Cancer? Front Pharmacol. 2021 Apr 29;12:662232. doi: 10.3389/fphar.2021.662232. PMID: 33995085; PMCID: PMC8118639.

16. Bay1125976, Akt1 Inhibitor - Selleckchem.com. Available from: https://www.selleckchem.com/products/bay1125976.html [Accessed Jul. 30, 2022].

17. Martínez-Martín S, Soucek L. MYC inhibitors in multiple myeloma. Cancer Drug Resist. 2021 Aug 13;4(4):842-865. doi: 10.20517/cdr.2021.55. PMID: 35582389; PMCID: PMC8992455.

18. "MYCI361 (NUCC-0196361)", MedchemExpress.Com.

19. "MICi975 (NUCC-0200975)", MedchemExpress.Com.

20. Whitfield JR, Soucek L. The long journey to bring a Myc inhibitor to the clinic. J Cell Biol. 2021 Aug 2;220(8):e202103090. doi: 10.1083/jcb.202103090. Epub 2021 Jun 23. PMID: 34160558; PMCID: PMC8240852.

21. Rietman EA, Scott JG, Tuszynski JA, Klement GL. Personalized anticancer therapy selection using molecular landscape topology and thermodynamics. Oncotarget. 2017 Mar 21;8(12):18735-18745. doi: 10.18632/oncotarget.12932. PMID: 27793055; PMCID: PMC5386643.

22. Rietman EA, Platig J, Tuszynski JA, Lakka Klement G. Thermodynamic measures of cancer: Gibbs free energy and entropy of protein-protein interactions. J Biol Phys. 2016 Jun;42(3):339-50. doi: 10.1007/s10867-016-9410-y. Epub 2016 Mar 24. PMID: 27012959; PMCID: PMC4942417.

23. Benzekry S, Tuszynski JA, Rietman EA, Lakka Klement G. Design principles for cancer therapy guided by changes in complexity of protein-protein interaction networks. Biol Direct. 2015 May 28;10:32. doi: 10.1186/s13062-015-0058-5. PMID: 26018239; PMCID: PMC4445818.

24. Hinow P, Rietman EA, Omar SI, Tuszyński JA. Algebraic and topological indices of molecular pathway networks in human cancers. Math Biosci Eng. 2015 Dec;12(6):1289-302. doi: 10.3934/mbe.2015.12.1289. PMID: 26775864.

25. Breitkreutz D, Hlatky L, Rietman E, Tuszynski JA. Molecular signaling network complexity is correlated with cancer patient survivability. Proc Natl Acad Sci U S A. 2012 Jun 5;109(23):9209-12. doi: 10.1073/pnas.1201416109. Epub 2012 May 21. PMID: 22615392; PMCID: PMC3384193.

26. Brant EJ, Rietman EA, Klement GL, Cavaglia M, Tuszynski JA. Personalized therapy design for systemic lupus erythematosus based on the analysis of protein-protein interaction networks. PLoS One. 2020 Mar 19;15(3):e0226883. doi: 10.1371/journal.pone.0226883. PMID: 32191711; PMCID: PMC7081981.

27. Huang S. On the intrinsic inevitability of cancer: from foetal to fatal attraction. Semin Cancer Biol. 2011 Jun;21(3):183-99. doi: 10.1016/j.semcancer.2011.05.003. Epub 2011 May 26. PMID: 21640825.

28. Moris N, Pina C, Arias AM. Transition states and cell fate decisions in epigenetic landscapes. Nat Rev Genet. 2016 Nov;17(11):693-703. doi: 10.1038/nrg.2016.98. Epub 2016 Sep 12. PMID: 27616569.

29. Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabuddhe NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LD, Patil AH, Nanjappa V, Radhakrishnan A, Prasad S, Subbannayya T, Raju R, Kumar M, Sreenivasamurthy SK, Marimuthu A, Sathe GJ, Chavan S, Datta KK, Subbannayya Y, Sahu A, Yelamanchi SD, Jayaram S, Rajagopalan P, Sharma J, Murthy KR, Syed N, Goel R, Khan AA, Ahmad S, Dey G, Mudgal K, Chatterjee A, Huang TC, Zhong J, Wu X, Shaw PG, Freed D, Zahari MS, Mukherjee KK, Shankar S, Mahadevan A, Lam H, Mitchell CJ, Shankar SK, Satishchandra P, Schroeder JT, Sirdeshmukh R, Maitra A, Leach SD, Drake CG, Halushka MK, Prasad TS, Hruban RH, Kerr CL, Bader GD, Iacobuzio-Donahue CA, Gowda H, Pandey A. A draft map of the human proteome. Nature. 2014 May 29;509(7502):575-81. doi: 10.1038/nature13302. PMID: 24870542; PMCID: PMC4403737.

30. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, Mathieson T, Lemeer S, Schnatbaum K, Reimer U, Wenschuh H, Mollenhauer M, Slotta-Huspenina J, Boese JH, Bantscheff M, Gerstmair A, Faerber F, Kuster B. Mass-spectrometry-based draft of the human proteome. Nature. 2014 May 29;509(7502):582-7. doi: 10.1038/nature13319. PMID: 24870543.

31. Guo Y, Sheng Q, Li J, Ye F, Samuels DC, Shyr Y. Large scale comparison of gene expression levels by microarrays and RNAseq using TCGA data. PLoS One. 2013 Aug 20;8(8):e71462. doi: 10.1371/journal.pone.0071462. PMID: 23977046; PMCID: PMC3748065.

32. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn). 2015;19(1A):A68-77. doi: 10.5114/wo.2014.47136. PMID: 25691825; PMCID: PMC4322527. Url: https://cancergenome.nih.gov/ [Accessed Aug. 27, 2023].

33. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006 Jan 1;34(Database issue):D535-9. doi: 10.1093/nar/gkj109. PMID: 16381927; PMCID: PMC1347471. Url: https://thebiogrid.org/ [Accessed Aug. 27, 2023].

34. Liberles, Arno. Introduction to Theoretical Organic Chemistry. Collier Macmillan, 1968.

35. Beard DA, Qian H. Chemical Biophysics: Quantitative Analysis of Cellular Systems. Cambridge Texts in Biomedical Engineering. Cambridge University Press; 2008:296-306. doi:10.1017/CBO9780511803345.015.

36. Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R. NCBI GEO: mining millions of expression profiles--database and tools. Nucleic Acids Res. 2005 Jan

1;33(Database issue):D562-6. doi: 10.1093/nar/gki022. PMID: 15608262; PMCID: PMC539976. Url: https://www.ncbi.nlm.nih.gov/geo/ [Accessed Aug. 27, 2023].

37. Friedman DR, Weinberg JB, Barry WT, Goodman BK, Volkheimer AD, Bond KM, Chen Y, Jiang N, Moore JO, Gockerman JP, Diehl LF, Decastro CM, Potti A, Nevins JR. A genomic approach to improve prognosis and predict therapeutic response in chronic lymphocytic leukemia. Clin Cancer Res. 2009 Nov 15;15(22):6947-55. doi: 10.1158/1078-0432.CCR-09-1132. Epub 2009 Oct 27. PMID: 19861443; PMCID: PMC2783430.

38. Boulesteix AL, Slawski M. Stability and aggregation of ranked gene lists. Brief Bioinform. 2009 Sep;10(5):556-68. doi: 10.1093/bib/bbp034. PMID: 19679825.

39. Ein-Dor L, Kela I, Getz G, Givol D, Domany E. Outcome signature genes in breast cancer: is there a unique set? Bioinformatics. 2005 Jan 15;21(2):171-8. doi: 10.1093/bioinformatics/bth469. Epub 2004 Aug 12. PMID: 15308542.

40. Wang L, Shalek AK, Lawrence M, Ding R, Gaublomme JT, Pochet N, Stojanov P, Sougnez C, Shukla SA, Stevenson KE, Zhang W, Wong J, Sievers QL, MacDonald BT, Vartanov AR, Goldstein NR, Neuberg D, He X, Lander E, Hacohen N, Regev A, Getz G, Brown JR, Park H, Wu CJ. Somatic mutation as a mechanism of Wnt/β-catenin pathway activation in CLL. Blood. 2014 Aug 14;124(7):1089-98. doi: 10.1182/blood-2014-01-552067. Epub 2014 Apr 28. PMID: 24778153; PMCID: PMC4133483.

41. Chuang HY, Rassenti L, Salcedo M, Licon K, Kohlmann A, Haferlach T, Foà R, Ideker T, Kipps TJ. Sub-network-based analysis of chronic lymphocytic leukemia identifies pathways that associate with disease progression. Blood. 2012 Sep 27;120(13):2639-49. doi: 10.1182/blood-2012-03-416461. Epub 2012 Jul 26. PMID: 22837534; PMCID: PMC3460686.

42. Rosenwald A, Alizadeh AA, Widhopf G, Simon R, Davis RE, Yu X, Yang L, Pickeral OK, Rassenti LZ, Powell J, Botstein D, Byrd JC, Grever MR, Cheson BD, Chiorazzi N, Wilson WH, Kipps TJ, Brown PO, Staudt LM. Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. J Exp Med. 2001 Dec 3;194(11):1639-47. doi: 10.1084/jem.194.11.1639. PMID: 11733578; PMCID: PMC2193523.

43. Klein U, Tu Y, Stolovitzky GA, Mattioli M, Cattoretti G, Husson H, Freedman A, Inghirami G, Cro L, Baldini L, Neri A, Califano A, Dalla-Favera R. Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. J Exp Med. 2001 Dec 3;194(11):1625-38. doi: 10.1084/jem.194.11.1625. PMID: 11733577; PMCID: PMC2193527.

44. Mraz M, Chen L, Rassenti LZ, Ghia EM, Li H, Jepsen K, Smith EN, Messer K, Frazer KA, Kipps TJ. miR-150 influences B-cell receptor signaling in chronic lymphocytic leukemia by regulating expression of GAB1 and FOXP1. Blood. 2014 Jul 3;124(1):84-95. doi: 10.1182/blood-2013-09-527234. Epub 2014 May 1. PMID: 24787006; PMCID: PMC4125356.

45. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022 Jul 5;50(W1):W216-W221. doi: 10.1093/nar/gkac194. PMID: 35325185; PMCID: PMC9252805. Url: https://david.ncifcrf.gov/ [Accessed Aug. 27, 2023]

46. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27-30. doi: 10.1093/nar/28.1.27. PMID: 10592173; PMCID: PMC102409. Url: https://www.genome.jp/kegg/ [Accessed Aug. 27, 2023]

47. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 2005 Jan 1;33(Database

issue):D514-7. doi: 10.1093/nar/gki033. PMID: 15608251; PMCID: PMC539987. Url: https://www.omim.org/ [Accessed Aug. 27, 2023]

48. Liu Y, Patel L, Mills GB, Lu KH, Sood AK, Ding L, Kucherlapati R, Mardis ER, Levine DA, Shmulevich I, Broaddus RR, Zhang W. Clinical significance of CTNNB1 mutation and Wnt pathway activation in endometrioid endometrial carcinoma. J Natl Cancer Inst. 2014 Sep 10;106(9):dju245. doi: 10.1093/jnci/dju245. PMID: 25214561; PMCID: PMC4200060.

49. Zhang JD, Wiemann S. "Bioconductor," KEGGgraph: a graph approach to KEGG PATHWAY in R and Bioconductor. Bioinformatics. 2009;25(11):1470-1471. Url: https://www.bioconductor.org/packages/release/bioc/html/KEGGgraph.html [Accessed Aug. 07, 2023]

50. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498-504. doi: 10.1101/gr.1239303. PMID: 14597658; PMCID: PMC403769. Url: https://cytoscape.org/ [Accessed Aug. 07, 2023].

51. van Riggelen J, Yetil A, Felsher DW. MYC as a regulator of ribosome biogenesis and protein synthesis. Nat Rev Cancer. 2010 Apr;10(4):301-9. doi: 10.1038/nrc2819. PMID: 20332779.

52. Wu CH, Sahoo D, Arvanitis C, Bradon N, Dill DL, Felsher DW. Combined analysis of murine and human microarrays and ChIP analysis reveals genes associated with the ability of MYC to maintain tumorigenesis. PLoS Genet. 2008 Jun 6;4(6):e1000090. doi: 10.1371/journal.pgen.1000090. Erratum in: PLoS Genet. 2013 Oct;9(10). doi:10.1371/annotation/a0e06cef-a7e4-4ec9-9f35-9df5e50bf7a2. PMID: 18535662; PMCID: PMC2390767.

53. Ljungström V, Cortese D, Young E, Pandzic T, Mansouri L, Plevova K, Ntoufa S, Baliakas P, Clifford R, Sutton LA, Blakemore SJ, Stavroyianni N, Agathangelidis A, Rossi D, Höglund M, Kotaskova J, Juliusson G, Belessi C, Chiorazzi N, Panagiotidis P, Langerak AW, Smedby KE, Oscier D, Gaidano G, Schuh A, Davi F, Pott C, Strefford JC, Trentin L, Pospisilova S, Ghia P, Stamatopoulos K, Sjöblom T, Rosenquist R. Whole-exome sequencing in relapsing chronic lymphocytic leukemia: clinical impact of recurrent RPS15 mutations. Blood. 2016 Feb 25;127(8):1007-16. doi: 10.1182/blood-2015-10-674572. Epub 2015 Dec 16. PMID: 26675346; PMCID: PMC4768426.

54. Braun R, Ronquist S, Wangsa D, Chen H, Anthuber L, Gemoll T, Wangsa D, Koparde V, Hunn C, Habermann JK, Heselmeyer-Haddad K, Rajapakse I, Ried T. Single Chromosome Aneuploidy Induces Genome-Wide Perturbation of Nuclear Organization and Gene Expression. Neoplasia. 2019 Apr;21(4):401-412. doi: 10.1016/j.neo.2019.02.003. Epub 2019 Mar 22. PMID: 30909073; PMCID: PMC6434407.

55. Chen H, Chen J, Muir LA, Ronquist S, Meixner W, Ljungman M, Ried T, Smale S, Rajapakse I. Functional organization of the human 4D Nucleome. Proc Natl Acad Sci U S A. 2015 Jun 30;112(26):8002-7. doi: 10.1073/pnas.1505822112. Epub 2015 Jun 15. PMID: 26080430; PMCID: PMC4491792.

56. Shi Y, Su XB, He KY, Wu BH, Zhang BY, Han ZG. Chromatin accessibility contributes to simultaneous mutations of cancer genes. Sci Rep. 2016 Oct 20;6:35270. doi: 10.1038/srep35270. PMID: 27762310; PMCID: PMC5071887.

57. Sidiropoulos N, Mardin BR, Rodríguez-González FG, Bochkov ID, Garg S, Stütz AM, Korbel JO, Aiden EL, Weischenfeldt J. Somatic structural variant formation is guided by and influences genome architecture. Genome Res. 2022 Apr;32(4):643-655. doi: 10.1101/gr.275790.121. Epub 2022 Feb 17. PMID: 35177558; PMCID: PMC8997353.

58. Melo US, Schöpflin R, Acuna-Hidalgo R, Mensah MA, Fischer-Zirnsak B, Holtgrewe M, Klever MK, Türkmen S, Heinrich V, Pluym ID, Matoso E, Bernardo de Sousa S, Louro P, Hülsemann W, Cohen M, Dufke A,

Latos-Bieleńska A, Vingron M, Kalscheuer V, Quintero-Rivera F, Spielmann M, Mundlos S. Hi-C Identifies Complex Genomic Rearrangements and TAD-Shuffling in Developmental Diseases. Am J Hum Genet. 2020 Jun 4;106(6):872-884. doi: 10.1016/j.ajhg.2020.04.016. Epub 2020 May 28. PMID: 32470376; PMCID: PMC7273525.

59. Pal K, Forcato M, Ferrari F. Hi-C analysis: from data generation to integration. Biophys Rev. 2019 Feb;11(1):67-78. doi: 10.1007/s12551-018-0489-1. Epub 2018 Dec 20. PMID: 30570701; PMCID: PMC6381366.

60. Jabbari K, Bernardi G. An Isochore Framework Underlies Chromatin Architecture. PLoS One. 2017 Jan 6;12(1):e0168023. doi: 10.1371/journal.pone.0168023. PMID: 28060840; PMCID: PMC5218411.

61. Upender MB, Habermann JK, McShane LM, Korn EL, Barrett JC, Difilippantonio MJ, Ried T. Chromosome transfer induced aneuploidy results in complex dysregulation of the cellular transcriptome in immortalized and cancer cells. Cancer Res. 2004 Oct 1;64(19):6941-9. doi: 10.1158/0008-5472.CAN-04-0474. PMID: 15466185; PMCID: PMC4772432.

62. Grade M, Ghadimi BM, Varma S, Simon R, Wangsa D, Barenboim-Stapleton L, Liersch T, Becker H, Ried T, Difilippantonio MJ. Aneuploidy-dependent massive deregulation of the cellular transcriptome and apparent divergence of the Wnt/beta-catenin signaling pathway in human rectal carcinomas. Cancer Res. 2006 Jan 1;66(1):267-82. doi: 10.1158/0008-5472.CAN-05-2533. PMID: 16397240; PMCID: PMC4737482.

63. Tsafrir D, Bacolod M, Selvanayagam Z, Tsafrir I, Shia J, Zeng Z, Liu H, Krier C, Stengel RF, Barany F, Gerald WL, Paty PB, Domany E, Notterman DA. Relationship of gene expression and chromosomal abnormalities in colorectal cancer. Cancer Res. 2006 Feb 15;66(4):2129-37. doi: 10.1158/0008-5472.CAN-05-2569. PMID: 16489013.

64. Speedy HE, Beekman R, Chapaprieta V, Orlando G, Law PJ, Martín-García D, Gutiérrez-Abril J, Catovsky D, Beà S, Clot G, Puiggròs M, Torrents D, Puente XS, Allan JM, López-Otín C, Campo E, Houlston RS, Martín-Subero JI. Insight into genetic predisposition to chronic lymphocytic leukemia from integrative epigenomics. Nat Commun. 2019 Aug 9;10(1):3615. doi: 10.1038/s41467-019-11582-2. PMID: 31399598; PMCID: PMC6689100.

65. Beekman R, Chapaprieta V, Russiñol N, Vilarrasa-Blasi R, Verdaguer-Dot N, Martens JHA, Duran-Ferrer M, Kulis M, Serra F, Javierre BM, Wingett SW, Clot G, Queirós AC, Castellano G, Blanc J, Gut M, Merkel A, Heath S, Vlasova A, Ullrich S, Palumbo E, Enjuanes A, Martín-García D, Beà S, Pinyol M, Aymerich M, Royo R, Puiggros M, Torrents D, Datta A, Lowy E, Kostadima M, Roller M, Clarke L, Flicek P, Agirre X, Prosper F, Baumann T, Delgado J, López-Guillermo A, Fraser P, Yaspo ML, Guigó R, Siebert R, Martí-Renom MA, Puente XS, López-Otín C, Gut I, Stunnenberg HG, Campo E, Martin-Subero JI. The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. Nat Med. 2018 Jun;24(6):868-880. doi: 10.1038/s41591-018-0028-4. Epub 2018 May 21. PMID: 29785028; PMCID: PMC6363101.

66. Lin YM, Wang CM, Jeng JC, Leprince D, Shih HM. HIC1 interacts with and modulates the activity of STAT3. Cell Cycle. 2013 Jul 15;12(14):2266-76. doi: 10.4161/cc.25365. PMID: 24067369; PMCID: PMC3755077.

67. Kiefer Y, Schulte C, Tiemann M, Bullerdiek J. Chronic lymphocytic leukemia-associated chromosomal abnormalities and miRNA deregulation. Appl Clin Genet. 2012 Mar 12;5:21-8. doi: 10.2147/TACG.S18669. PMID: 23776377; PMCID: PMC3681189.

68. Rietman EA, McDonald JT, Hlatky L. Organism mutation stability and cancer. Hypothesis in Clinical Medicine, editors M.M. Shoja, P.D. Agutter, et al. Nova Science Publishers, 2013, pp135-145.

69. Dekker J, Belmont AS, Guttman M, Leshyk VO, Lis JT, Lomvardas S, Mirny LA, O'Shea CC, Park PJ, Ren B, Politz JCR, Shendure J, Zhong S; 4D Nucleome Network. The 4D nucleome project. Nature. 2017 Sep

13;549(7671):219-226. doi: 10.1038/nature23884. Erratum in: Nature. 2017 Nov 22;: PMID: 28905911; PMCID: PMC5617335.

70. Lee BH, Wu Z, Rhie SK. Characterizing chromatin interactions of regulatory elements and nucleosome positions, using Hi-C, Micro-C, and promoter capture Micro-C. Epigenetics Chromatin. 2022 Dec 21;15(1):41. doi: 10.1186/s13072-022-00473-4. Erratum in: Epigenetics Chromatin. 2023 Jun 7;16(1):22. PMID: 36544209; PMCID: PMC9768916.

71. Sawh AN, Mango SE. Chromosome organization in 4D: insights from C. elegans development. Curr Opin Genet Dev. 2022 Aug;75:101939. doi: 10.1016/j.gde.2022.101939. Epub 2022 Jun 24. PMID: 35759905.

72. Ouillette P, Collins R, Shakhan S, Li J, Li C, Shedden K, Malek SN. The prognostic significance of various 13q14 deletions in chronic lymphocytic leukemia. Clin Cancer Res. 2011 Nov 1;17(21):6778-90. doi: 10.1158/1078-0432.CCR-11-0785. Epub 2011 Sep 2. PMID: 21890456; PMCID: PMC3207001.

73. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, Costello CE, Cravatt BF, Fenselau C, Garcia BA, Ge Y, Gunawardena J, Hendrickson RC, Hergenrother PJ, Huber CG, Ivanov AR, Jensen ON, Jewett MC, Kelleher NL, Kiessling LL, Krogan NJ, Larsen MR, Loo JA, Ogorzalek Loo RR, Lundberg E, MacCoss MJ, Mallick P, Mootha VK, Mrksich M, Muir TW, Patrie SM, Pesavento JJ, Pitteri SJ, Rodriguez H, Saghatelian A, Sandoval W, Schlüter H, Sechi S, Slavoff SA, Smith LM, Snyder MP, Thomas PM, Uhlén M, Van Eyk JE, Vidal M, Walt DR, White FM, Williams ER, Wohlschlager T, Wysocki VH, Yates NA, Young NL, Zhang B. How many human proteoforms are there? Nat Chem Biol. 2018 Feb 14;14(3):206-214. doi: 10.1038/nchembio.2576. PMID: 29443976; PMCID: PMC5837046.

74. Valencia-Serna J, Aliabadi HM, Manfrin A, Mohseni M, Jiang X, Uludag H. siRNA/lipopolymer nanoparticles to arrest growth of chronic myeloid leukemia cells in vitro and in vivo. Eur J Pharm Biopharm. 2018 Sep;130:66-70. doi: 10.1016/j.ejpb.2018.06.018. Epub 2018 Jun 18. PMID: 29913272.