



Understanding the role of toggle genes in chronic lymphocytic leukemia proliferation



Olga Sirbu¹, Gunjan Agarwal^{1,2}, Alessandro Giuliani³ & Kumar Selvarajoo^{1,4,5,6} ✉

Cancer cell populations, such as chronic lymphocytic leukemia (CLL), are characterized by aberrant proliferation and plasticity: cells may switch between states so increasing population heterogeneity. Previous works have shown that gene expression noise can impact cell-state transition. To gain better insights into transcriptome-wide expression dynamics and the effect of noise on state transition, here we investigate RNA-Seq data of proliferative (PC) and non-proliferative (NPC) CLL cells. Various data analytics were applied to the whole transcriptome, switch-like toggle (ON/OFF) genes, temporal differentially expressed (DE) genes, and randomly selected genes. Collectively, we identified 2713 temporal DE genes (DESeq2 with 4-fold, $p < 0.05$) and 1704 toggle genes shaping the differentiation process over a period of 96 h, with 604 overlapping genes between them. Despite their lower numbers compared to DE, toggle genes contributed significantly to gene expression noise in both cell types. Toggle gene analyses revealed the enrichment of genes involved in processes such as G-alpha signaling and muscle contraction as proliferation related RHO-GTPase, interleukin and chemokine signaling, and lymphoid cell communication. Thus, toggle genes, although being random ON/OFF genes, show gene expression functional variability. These results suggest that toggle genes play an important role in shaping cell population plasticity.

Chronic lymphocytic leukemia (CLL) is the most common type of leukemia in adults, with a median age of diagnosis and onset of 70 years¹. It is characterized by the uncontrolled proliferation of monoclonal lymphoid cells, specifically transformed mature CD5+ and CD23+ lymphocytes which are impaired in their function^{2,3}. Due to the heterogeneous nature of CLL, current treatment approaches for the disease are complex and suboptimal^{2,4–8}. Previously, it has been observed that tumors can leverage genetic, epigenetic, and stochastic variability to foster the necessary plasticity that leads to resistance and treatment evasion^{9–12}. While CLL is known to exhibit significant clonal and metabolic plasticity, its transcriptomic plasticity remains underexplored. Thus, transcriptome-wide analytics, that are capable of tracking systemic responses in gene expression, is necessary and it offers an important avenue for the study of CLL plasticity.

The construction of gene expression landscapes^{13,14} allows to understand transcriptome-wide expression dynamics, especially in the context of

cancer. This approach implies the conceptualization of living cells as dynamic systems that occupy specific states at any given moment. As cells undergo dynamic processes they move through the landscape, eventually tending towards conditions of stability or equilibrium, known as “attractors” (Fig. 1A)^{8,13–16}. Thus, the gene expression trajectories that cells follow as they move through the expression landscape are important for cell-fate decision making.

For cancer, we can think of a simplified cell-fate landscape with only two attractors: a normal state, and a cancer state. Under normal circumstances, cells are more likely to settle into the normal cell attractor, and very large perturbations are necessary to cause a cell to move to the cancer attractor (Fig. 1B, left). However, cancer cell transcriptomes exhibit a level of plasticity that endows them with unpredictable behaviors and patterns, rarely seen in normal healthy cells^{15,17–19}. In the case of an altered landscape (being this alteration coming from diverse initial causes, such as genetic

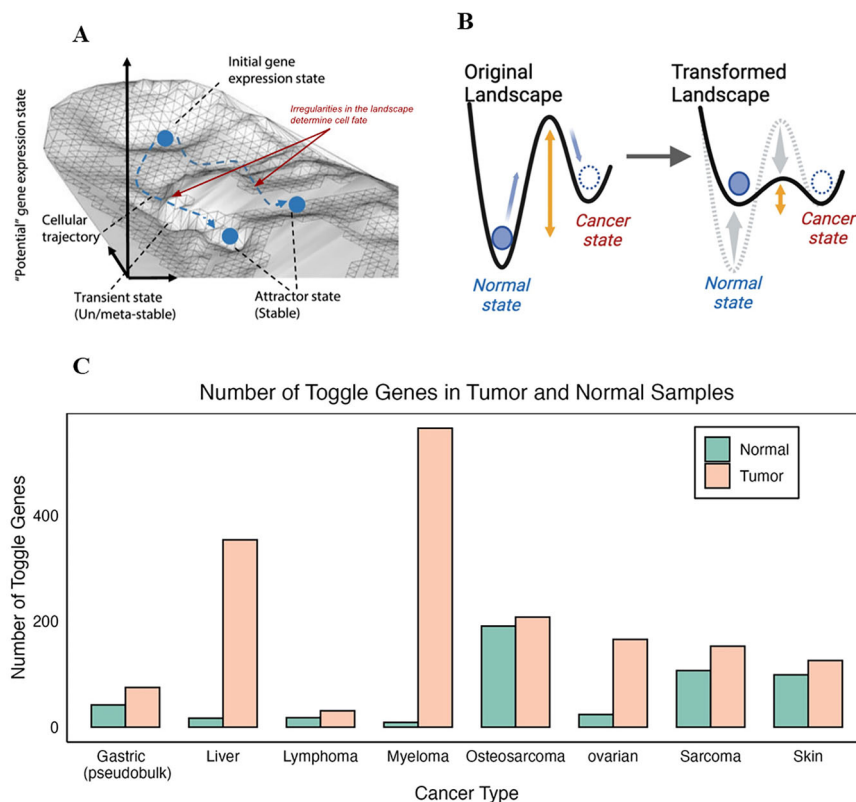
¹Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Singapore, Republic of Singapore. ²Engineering Systems and Design (ESD), Singapore University of Technology and Design (SUTD), Singapore, Republic of Singapore. ³Environment and Health Department, Istituto Superiore di Sanità, Rome, Italy. ⁴Synthetic Biology Translational Research Program, Yong Loo Lin School of Medicine, National University of Singapore (NUS), Singapore, Republic of Singapore. ⁵Synthetic Biology for Clinical and Technological Innovation (SynCTI), National University of Singapore (NUS), Singapore, Republic of Singapore. ⁶School of Biological Sciences, Nanyang Technological University (NTU), Singapore, Republic of Singapore.

✉ e-mail: kumar_selvarajoo@bii.a-star.edu.sg

Fig. 1 | Attractor landscape and toggle genes.

A Transcriptome expression landscape shows how cells follow certain trajectories to settle into different attractor states, figure adapted from ref. 14.

B Normal (left) and transformed (right) simplified cell fate landscape which shows that cells require larger perturbations (yellow arrows) to exit their current normal state and have the potential to fall into the cancer attractor. In a transformed landscape, the energy barrier required changes due to changes in attractor depth and thus state changes are more likely to occur. **C** Breakdown of toggle genes extracted from normal (orange) and tumor (green) samples shows a higher incidence of toggle genes in tumor samples across 8 investigated cancer sets (Table S2). Only datasets containing paired normal and tumor samples were selected to allow for direct comparison, however, the incidence of toggle genes in cancer is ubiquitous. Figure 1A was adapted from [14].



mutations, chromosomal aberrations, or microenvironmental stimuli), the perturbation required to exit the normal attractor and settle into a new cancer state is significantly smaller (Fig. 1B, right). Therefore, perturbations such as gene expression noise, can play major roles in shaping cancer states^{4,17,20–23}. Often the highly variable genes, such as the differentially expressed (DE) genes between the attractor states, can play crucial roles for the state transition. Hence focusing on such genes' expression over the state-transition period is crucial.

In addition to DE genes, gene expression noise plays a significant role in producing diversity and shaping complex biological processes^{20,24,25}. During cell fate decision making, transcriptome-wide noise has been associated with controlling lineage choices in mammalian progenitor cells, allowing for the emergence of outlier cells contributing to population proclivity²⁴. On a smaller scale, noise in the expression of individual genes has also been found to be equally important; in *B. subtilis*, controlling transcriptional and translational noise of *comK* was associated with vegetative- and competent-state transitions²⁶. Likewise, in cancer, noise can play a significant role, as evidenced by the increasing expression diversity observed in late-stage tumors and their association with cancer outcomes^{17,22,23}.

Gene expression noise level affect cell-state transition, in a way similar to the effect of temperature in state transition in inorganic matter. In addition to such 'standard' noise following continuous distribution, a 'discrete' noise coming from toggle genes²⁷ is at play. These genes exhibit a "switch-like" behavior, being "OFF" in one sample (or condition) and "ON" in another, leading to significant weighted noise across samples. This phenomenon has been observed across a wide range of organisms, from unicellular to human mammalian cells, and appears to be consistent regardless of the RNA extraction method employed (Tables S1, S2). Of particular interest, toggle genes show a higher incidence in cancer and cell proliferation data, where they contribute significantly to transcriptome-wide noise²⁷. Moreover, our observations indicate a greater prevalence of toggle genes in tumor samples compared to their healthy counterparts (Fig. 1C). In various cancers, including but not limited to prostate, lung, and breast cancer,

similar molecular switches have been observed that not only contribute to drug resistance but also provide the molecular plasticity required for proliferation, metastasis, and uncontrolled growth^{27–31}. Toggle genes have also been observed in other situation; the alternation between the lytic and lysogenic phases of phage lambda^{32,33}, many endogenous retrovirus (ERV) sequences exhibit a bi-stable (yes/no) activation behavior, inherited from their viral origins³⁴. Furthermore, the frequency of ERVs positively correlates with evolutionary complexity and varies significantly between cell lines^{35,36}.

Thus, the investigation of switch-like or toggle genes, on top of DE genes, in especially cancer during periods of proliferation, is crucial for understanding the role of gene expression variability in cellular plasticity. In this study, we aim to expand the current understanding of CLL proliferation in the context of transcriptomic plasticity by specifically investigating the influence of toggle genes alongside temporal differential gene (DE) expression analyses. We expand the definition of toggle genes to include comparisons between samples of the same condition, capturing variability in gene expression across distinct samples. To achieve this, we made use of CLL transcriptomic data from several studies (Table S3), with an increased focus on temporal transcriptomic data from a recent study conducted by Schleiss et al.³⁷ that investigated the proliferative signature of CLL patient cells by segregating tumor cells into proliferative cells (PC), and non-proliferative cells (NPC)^{37,38}. By leveraging advanced data analytics techniques—ranging from correlation, noise analysis, dimensionality reduction and gene enrichment—our objective is to elucidate the complex interplay, and the role played by toggle genes and differentially expressed genes in CLL proliferation.

Results

CLL transcriptome data

For all considered CLL datasets (Table S3, refs. 38–40), we first performed gene expression filtering using statistical distribution fitting and threshold-based filtering (Fig. S1, Methods)^{14,41}. From the whole transcriptome, this process removed very low and technically noisy genes, leaving only robust gene

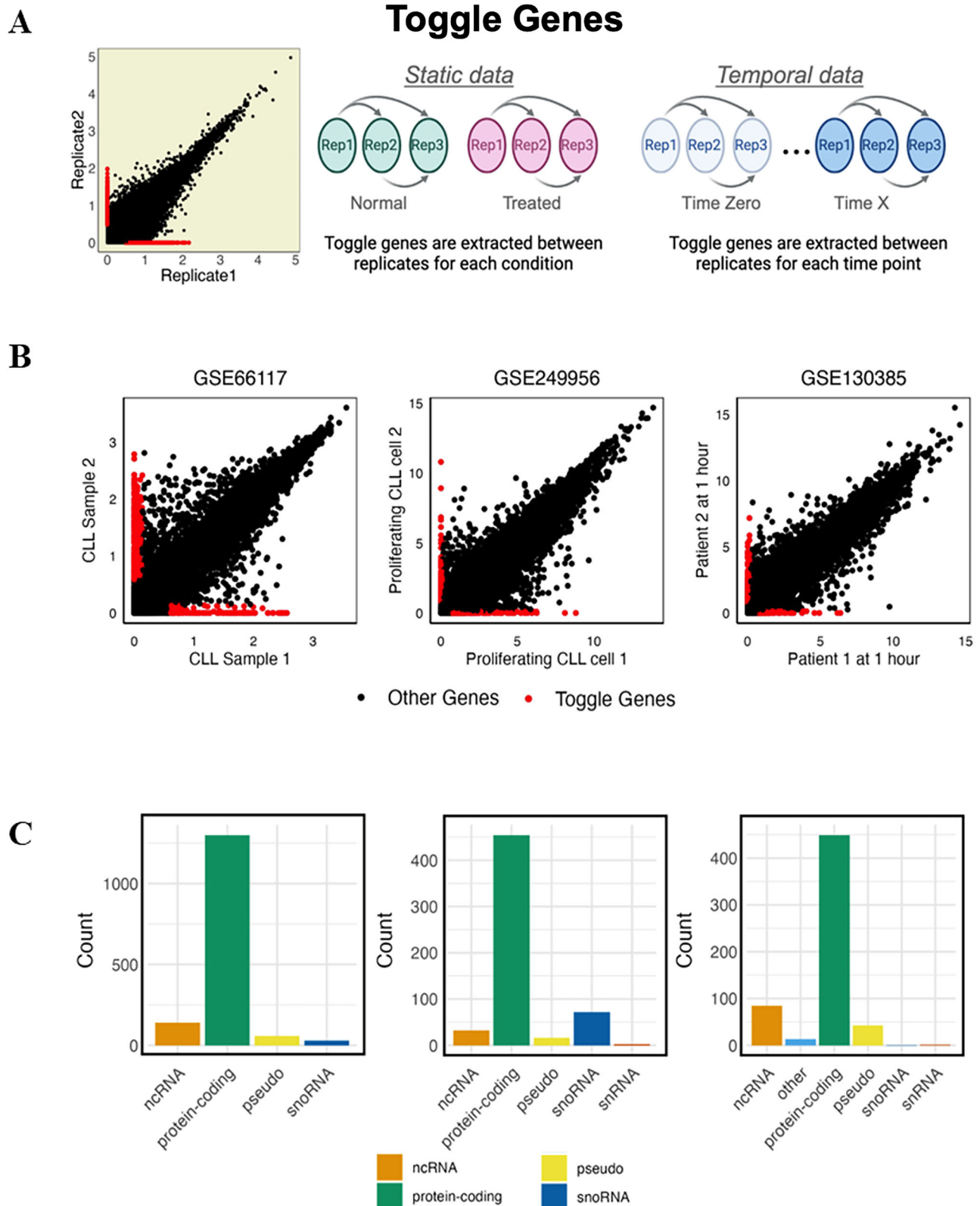


Fig. 2 | Cancer toggle genes and their broad biotypes. **A** Illustrative schematic of the extraction of toggle genes within the larger transcriptome. **B** Presence of toggle genes (red) within the transcriptome-wide scatter in three investigated CLL datasets: GSE66117, GSE249956, GSE130385 using TMM normalization method (C) Biotype distribution of toggle genes in three CLL datasets (Table S3), with the prevalent biotype being protein-coding genes.

Table 1 | Number of extracted toggle genes and DE genes

Gene Set	Number of Toggle Genes									
	Total:	0 h	1 h	1h30min	3h30min	6h30min	12 h	24 h	48 h	96 h
Same- condition Toggle Genes	1704	621	610	554	578	545	301	383	285	309
Temporal DE Genes, FC > 2	9148	NA	689	2580	4547	3986	3656	3643	5007	4600
Temporal DE Genes, FC > 4	2713	NA	179	515	1171	1177	1077	1184	1422	1319

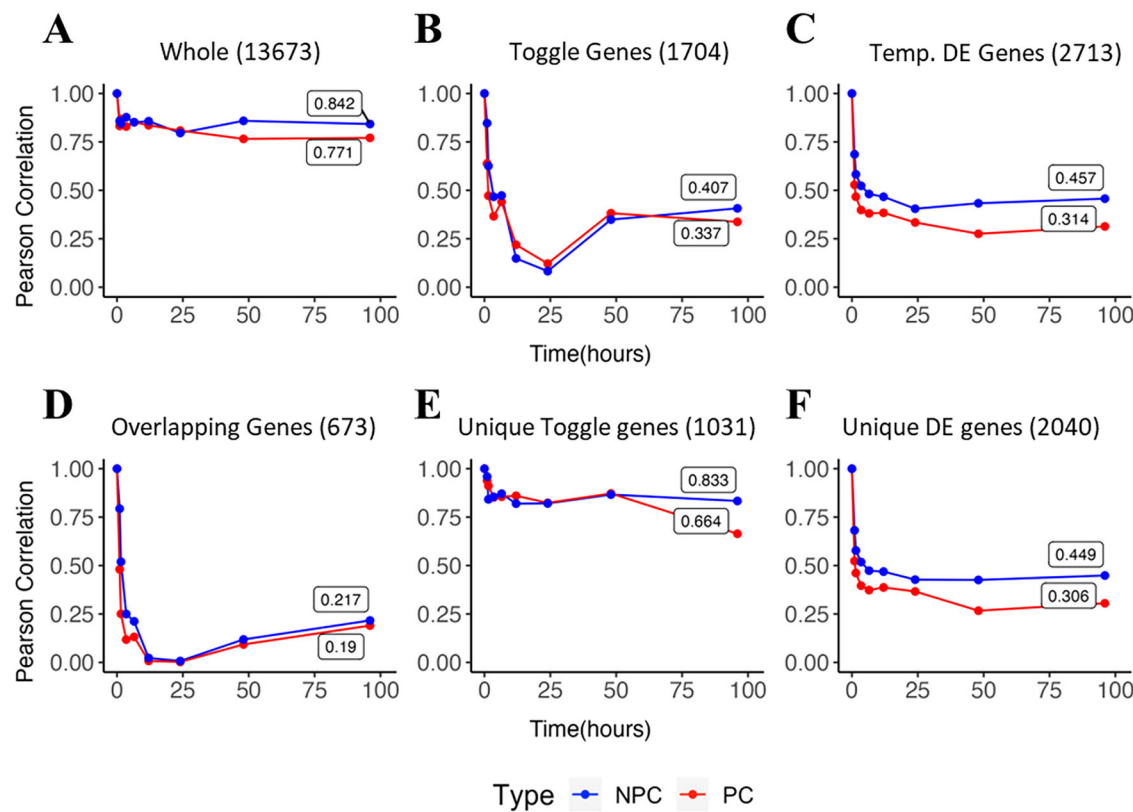


Fig. 3 | Average autocorrelation of PC and NPC cells across time. A Pearson autocorrelation for the whole transcriptome (~13 K genes). B Pearson autocorrelation for extracted toggle genes (~1.7 K genes). C Pearson autocorrelation for high fold-change (4FC) differentially expressed temporal genes (~3 K genes). D Pearson autocorrelation for the overlapping genes between DEG and Toggle genes. E Pearson autocorrelation for unique toggle genes. F Pearson autocorrelation for Unique DEG.

expressions for further analyses (Table S3). The same was done for the CLL proliferating cells (PC) and non-proliferating cells (NPC) at 9 time points after B cell receptor (BCR) stimulation ($n = 0, 1, 1.5, 3.5, 6.5, 12, 24, 48, 96$ h, GSE130385).

The presence of toggle genes in CLL data

Toggle genes were identified in all three CLL datasets by comparing gene expressions between distinct patient samples exposed to the same disease state. These were termed as toggle genes from same-condition samples, that is, genes with zero expression in one sample and positive expression above a noise threshold in another (Fig. 2A). The noise threshold was derived using statistical distribution fitting analysis (Methods), to ensure that the identified toggle genes reflect genuine biological variability rather than technical noise.

In the transcriptome-wide scatterplots (Fig. 2B), toggle genes (red) are distributed along the x- and y-axes in all datasets. Biotype analysis revealed that the majority of these genes are protein-coding, irrespective of the RNA extraction method used (Table S3), while a smaller subset consists of non-coding genes. The consistent identification of toggle genes in all datasets, combined with their predominance as protein-coding genes, highlights the inherent randomness and instability within CLL transcriptomes.

The concept of ‘randomness’ in toggle genes exhibits a unique characteristic. Typically, randomness is associated with statistical distributions such as uniform, normal, or, more commonly in biological systems, log-normal continuous distributions. However, toggle genes introduce a different form of randomness: toggling is inherently a discrete binary process at the single-cell level. When this behavior extends to the cell population level in the form of unbalanced toggles, it leads to a pronounced symmetry breaking within the population, ultimately driving the system in a specific direction⁴². We will explore this concept further in the following discussion.

Tracking the temporal global, DE and toggle genes response

As cell proliferation is a dynamic process, we next investigated the behavior of toggle genes in CLL proliferation using the PC and NPC dataset. DESeq2 analysis identified 9148 temporal DE genes between time points t_0 and t_m , applying a two-fold change and a p -value below 0.05 (Table 1). While not unexpected, this substantial gene set, representing 71% of the filtered transcriptome, suggested extensive involvement of DE genes in cell proliferation. To facilitate comparison with the smaller toggle gene set (1704 genes), the threshold was increased to a four-fold change, reducing the DE gene set to 2713 genes (Table 1). This stricter threshold helps exclude genetic elements that merely follow the system’s general dynamics due to inter-gene

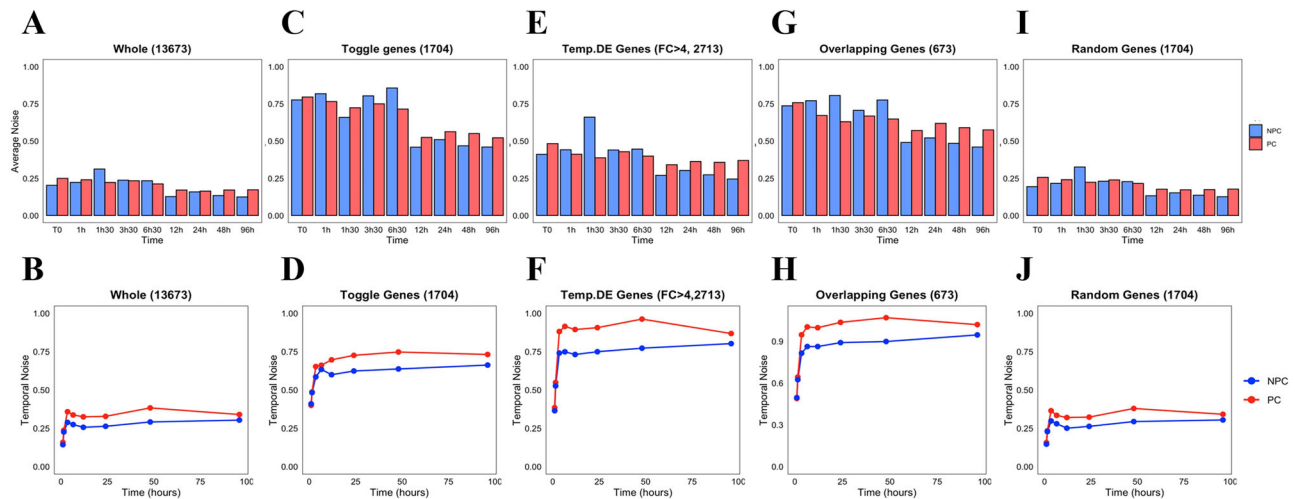


Fig. 4 | Noise changes in time for PC (red) and NPC (blue) samples.

A Transcriptome-wide average noise changes in time between same-condition samples. **B** Average noise changes relative to time zero for the whole transcriptome. **C** Average noise changes in time between samples for toggle genes. **D** Average noise changes relative to time zero for toggle genes. **E** Average noise changes in time between samples for high FC DE genes. **F** Average noise changes relative to time zero

for 4FC DE genes. **G** Average noise changes in time between samples for overlapping DE and toggle genes. **H** Average noise changes relative to time zero for overlapping DE and toggle genes. **I** Average noise changes in time between samples for random subsets of 1704 genes. **J** Average noise changes relative to time zero for random subsets of 1704 genes.

correlations⁴³, without being directly involved in the phenomenon under investigation.

Subsequent temporal Pearson and Spearman correlation analyses of the transcriptome, toggle genes, and temporal DE genes revealed a rapid decline in correlation between 3.5 and 6.5 h, followed by stabilization (Fig. 3A–C, Pearson, Fig. S2, Spearman). Both PC and NPC groups exhibited similar effects, particularly during the critical first four time points. Toggle genes, despite deriving through comparison between same time point and same conditions showed dynamic responses similar to temporal DE genes. Notably, 673 overlapping genes between toggle and temporal DE genes displayed the most significant correlation drop between 12 and 24 h, nearly reaching zero, before partial recovery (Fig. 3D). After removing these overlapping genes, the unique temporal DE genes exhibited a more pronounced response than the unique toggle genes (Fig. 3E–F), suggesting that the strong temporal signal in toggle genes is largely driven by the overlapping subset.

Overall, these results suggest that the pronounced changes in correlation observed for DE genes and especially their intersection with toggle genes reflect the proliferative processes occurring within CLL cells. Both gene sets exhibit significantly larger responses compared to the rest of the transcriptome, with their intersection capturing some of the most dynamically responsive genes in both PC and NPC groups.

Toggle genes possess the highest gene expression noise

Gene expression noise, measured as the squared coefficient of variation (CV^2), was evaluated for the whole transcriptome and for specific gene sets: toggle genes, DE genes, overlapping genes, and random subsets (Methods). Two types of noise were assessed: (1) between-sample noise, capturing variability among samples at the same time point, and (2) temporal noise, capturing changes in gene expression over time relative to the baseline (t_0) (Fig. 4, Fig. S3).

For between-sample noise, toggle genes exhibited the highest variability levels, followed by DE and overlapping genes, in both PC and NPC groups. Noise levels peaked at 6.5 h post-stimulation across all gene sets, suggesting increased variability among same-condition samples at this time point. This heightened variability reflects greater heterogeneity within the population, which stabilized at later time points (Fig. 4a, c, e, g, i, S3).

Temporal noise analysis showed that DE genes exhibited slightly higher levels than toggle genes, with both sets displaying significantly greater

noise compared to the whole transcriptome or random subsets (Fig. 4b, d, f, h, j, S3). Notably, overlapping genes, despite representing only a small fraction of the other gene sets, exhibited the greatest temporal changes, highlighting their substantial contribution to transcriptome-wide noise and their distinct dynamic behavior over time. Although toggle genes were selected based on sample-to-sample variability at a single time point, their temporal noise levels were also elevated, indicating that some of these genes may display dynamic behavior across time as well. Notably, the increased temporal variability of toggle genes is intrinsically linked to their toggling nature, causing them to oscillate between two extremes. This characteristic makes them natural ‘noise amplifiers,’ particularly when an imbalance occurs in their oscillation between ON and OFF states⁴².

To further examine gene expression variability over time, we analysed Shannon entropy across gene sets (Methods). Notably, the whole transcriptome and random subsets exhibited stable or relatively constant entropy (Fig. S4A, G, H), while toggle and DE genes displayed more dynamic behaviors (Fig. S4B–D). Toggle genes showed a steady decline in entropy, reaching a minimum at 24 h, followed by a pronounced increase at 48 and 96 h. DE genes, on the other hand, showed a gradual increase in entropy across all time. This biphasic trend for toggle genes suggests an initial period of transcriptional convergence followed by renewed variability or divergence in expression. The late stage increase in entropy may reflect the stable reactivation of toggle gene expressions or the emergence of distinct subpopulations, during differentiation or proliferation, responding in a coordinated but heterogeneous manner, although further experimental work is required to confirm this.

Lastly, we analyzed temporal toggle genes, defined as genes toggling in expression between time points (t_0 and t_n). A total of 2,561 temporal toggle genes were identified. However, noise and autocorrelation analyses revealed weaker responses for temporal toggle genes compared to same-condition toggle genes, likely due to differences in the size and composition of the gene sets (Fig. 4 and Fig. S3). Despite this, the analysis of temporal toggle genes provides additional insights into transcriptomic variability over time and emphasizes the complexity of gene expression dynamics in CLL.

Gene Enrichment analyses of toggle and DE genes for PC and NPC

Now that we have shown both toggle and DE genes are important for shaping temporal dynamics and variability in CLL, to understand their

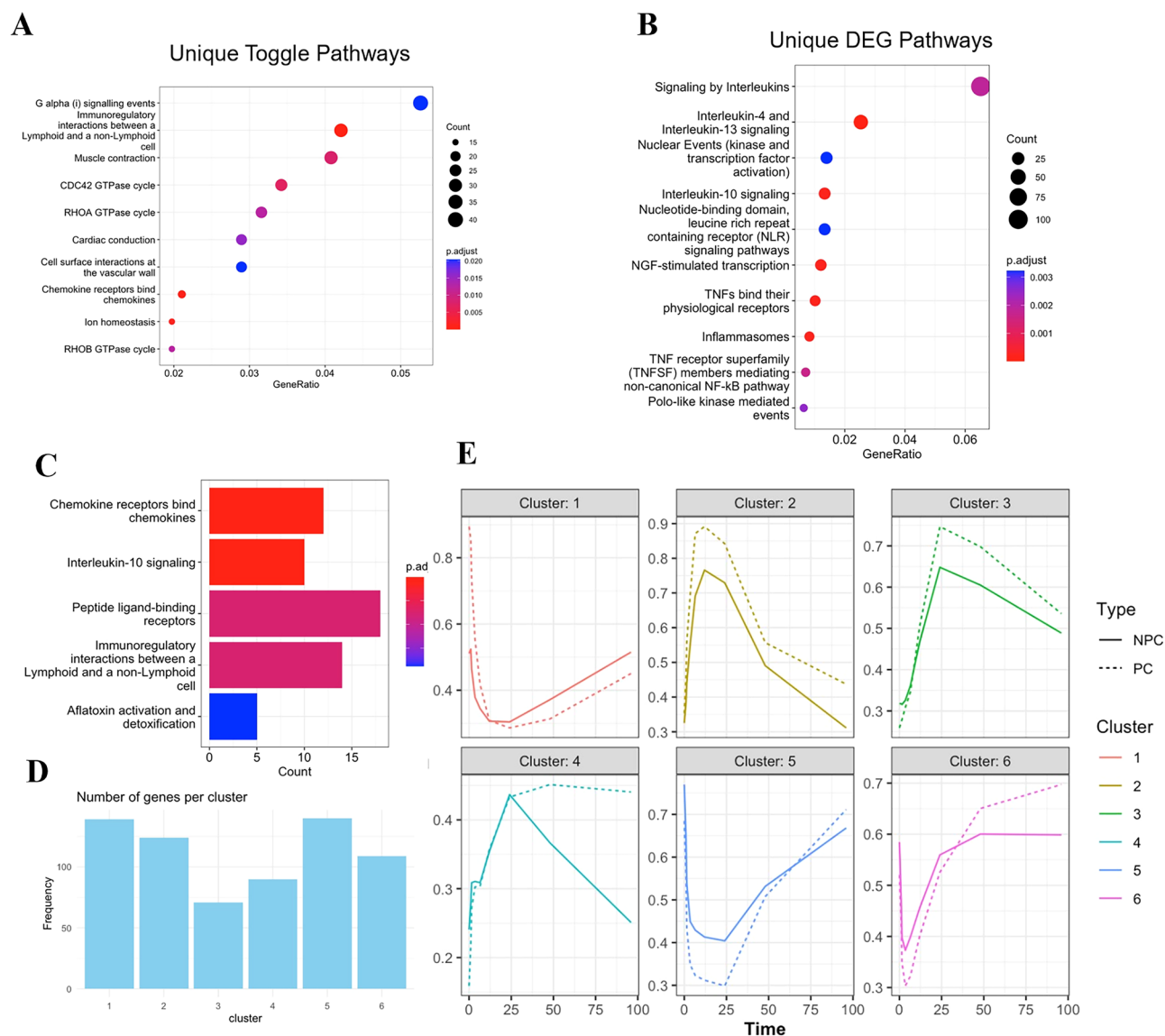


Fig. 5 | Functional enrichment analysis of DE genes and toggle genes and the overlap between the different sets. A Enriched Reactome pathways for unique toggle genes. **B** Enriched Reactome pathways in temporal DE genes. **C** Enriched

Reactome pathways for overlapping DEG and toggle genes. **D** Average temporal signatures of overlapping gene clusters based on hierarchical clustering.

biological functions, the *Reactome* pathway enrichment analysis was conducted. Toggle genes were enriched in key processes such as lymphoid cell communication and RHO GTPases, while DE genes were associated with immune-related pathways, including interleukin signaling and TNF-related processes (Fig. 5A - toggle genes, 5B - DEGs, 6C - overlapping genes). Notably, overlapping genes, which shared characteristics of both toggle and DE genes, were particularly enriched in chemokine receptor processes, interleukin signaling, and lymphoid immunoregulatory interactions.

Given that the experimental setup involved cell treatment with chemokines and interleukins to stimulate survival and proliferation, the enrichment of these processes among toggle and overlapping genes serves as a proof of principle, underscoring their biological significance. This alignment between the observed enrichment and the experimental conditions also reinforces the importance of toggle genes in the cellular responses studied.

Further analysis of overlapping genes identified six clusters ranging from 70 to over 100 genes, each with distinct temporal expression profiles (Fig. 5D). Sharp early responses were observed in interleukin signaling and chemokine-related processes, particularly in clusters 1 and 2 (Fig. S5A, B).

Additionally, cell cycle checkpoint processes exhibited a delayed response, peaking at 24 h before declining, consistent with the major transcriptomic changes noted in earlier analyses (S5).

The following are top 10 toggle genes based on their squared of variation (CV): SOX2, NCS1, ALPP, GPR34, EEPD1, SPNS2, CYP2C18, SIX3, F2RL2, RPRML. Notably, SOX2, NCS1 and SPNS2 stand out for their potential involvement in the proliferation of CLL cells. SOX2, a transcription factor that is necessary for maintaining stem cell properties, has been shown to contribute to the self-renewal and tumorigenic potential of leukemia stem cells⁴⁴. NCS1 (neuronal calcium sensor 1) encodes a protein that regulates calcium signaling, which has been found to be essential for immune cell function and activation, with its dysregulation potentially driving leukemogenesis⁴⁵. Lastly, SPNS2, involved in transporting sphingosine-1-phosphate (S1P), affects leukemic cell migration and survival, which are both essential for CLL cells in the lymph node microenvironment⁴⁶. The genes identified, such as SOX2, NCS1, and SPNS2, influence critical processes like cell signaling, migration, and self-renewal in CLL cells, all of which contribute to CLL progression and provide potential targets for future therapeutic strategies.

In summary, the enrichment analysis demonstrates that toggle genes, especially those overlapping with DE genes, are involved in critical biological processes related to immune function, cell cycle regulation, and differentiation, aligning closely with the experimental conditions designed to activate these pathways.

Discussion

The study of cancer presents significant challenges not only because of the disease's inherent complexity and aggressiveness but also due to its heterogeneous nature, including cellular plasticity, compounded by a limited understanding of transitions between cancer states^{8,10}. Cellular plasticity and state transitions are thought to be influenced by transcriptomic instability, which has been previously linked to tumor progression and treatment resistance^{9,12}. As observed in previous studies, the transcriptomes of cancer cells are often unstable and display unique expression deviations^{43–45}. This underscores the need for approaches that capture transcriptomic variability, including noise, which has been shown to play a role in shaping cell states and tipping cellular trajectories^{13,17,20,24,26}.

Molecular “switch-like” behaviors, characterized by flexibility and plasticity, have been shown to contribute to adaptive and evasive behaviors in cancer cells^{29,31,32}. Toggle genes, which exhibit binary “ON/OFF” expression patterns, represent a specific instance of this phenomenon. Especially since noise, or gene expression variability, is critical for cell- or attractor-state transition¹⁹, studying genes that contribute most to noise may provide clues to controlling unwanted state-transition such as normal cells becoming cancerous cells.

Our findings showed an increased incidence of toggle genes in cancer samples compared to healthy or adjacent tissues from the same individuals. This observation highlights the variability within cancer transcriptomes, which may reflect broader processes like proliferation or immune modulation. On a more general perspective, the higher proportion of toggle genes in cancer is consistent with the ‘noise amplifier’ role allowing cancer cells to explore a wider phase space exploration than healthy cells. This noise amplification has very important consequences in terms of therapy resistance and recurrence of cancer⁸.

By focusing on the temporal transcriptomic dynamics of CLL cells following BCR stimulation—a key driver of proliferation in this disease—we sought to investigate how toggle genes and transcriptomic noise contribute to variability during the proliferative response. Rather than implying causality, we aimed to show that these transcriptomic features align and correlate with the instabilities observed during CLL proliferations.

We identified 1704 toggle genes and 2713 DE genes with a significant temporal response (above 4-fold change). Auto-correlation analysis revealed a sharp decline in transcriptome correlation between 3.5 and 6.5 h post-stimulation, coinciding with early proliferation events. This pattern of variability, particularly in PCs, suggests that transcriptomic instability accompanies the proliferative process. A subset of 673 toggle genes overlapped with DE genes, showing the largest temporal shifts, while unique toggle genes displayed variability across same-condition samples. This distinction was further supported by dimensionality reduction, noise, and entropy analyses, which revealed that overlapping genes exhibit characteristics of both toggle and DE genes. These findings reinforce the idea that transcriptomic instability underlies the dynamic responses observed during CLL proliferation.

The enrichment analysis provided additional insight into the biological relevance of toggle and overlapping genes. The enrichment of toggle-genes involved in G-alpha signaling, muscle contraction, and cardiac conduction could be considered as largely unexpected, while the enrichment of chemokine and interleukin signaling, aligns with the experimental conditions designed to promote survival and proliferation. In this respect, it is worth noting that cytoskeleton remodeling (driven by the same genes linked to muscle contraction) is since long time recognized as a crucial player in cancer⁴⁵ while being at the same time an obliged step in cell division. Similar considerations hold for cardiac conduction genes⁴⁶ and G-alpha signaling⁴⁷. The presence of differentially enriched pathways validates the notion that

toggle-genes observed variability reflects biologically meaningful responses rather than pure random noise. Furthermore, the enrichment of RHO-GTPase signaling suggests potential novel mechanisms underlying cancer proliferation, offering new directions for investigation.

Interestingly, the overlapping genes represent a subset of the transcriptome that bridges temporal responsiveness and variability across samples. This dual role highlights their importance in both proliferation and plasticity. For instance, processes like chemokine signaling, which are well-established in CLL, were also enriched in toggle genes, indicating their potential contribution to both immune modulation and cellular heterogeneity. This supports the hypothesis that toggle genes reflect disturbances within important processes as evidenced by their transcriptomic expression, that can contribute to variations in disease progression.

Finally, our findings on RHO-GTPases underscore their significance in cancer dynamics^{48,49}. Their consistent temporal expression patterns, coupled with differences between PC and NPC groups, suggest they play a regulatory role in tumor initiation and progression. These genes, identified as toggle genes in this study, may serve as key regulators of cellular behaviors essential for cancer development, making them potential therapeutic targets in CLL.

Overall, our study highlights the role of transcriptomic instability as a feature of cancer proliferation. Toggle genes, particularly those overlapping with DE genes, provide evidence of this instability, reflecting both temporal changes and population-level variability. By identifying the dynamic interplay between noise, gene expression dynamics, and cellular behavior, this study deepens our understanding of CLL's proliferative signature and its complex molecular underpinnings from a system dynamics viewpoint. Future work should further explore these transcriptomic features to uncover their impact on disease progression and actual treatment outcomes, with the aim of developing more targeted novel therapeutics.

Methods

Pre-processing

For the time series data (GSE130385³⁷), we first removed genes with constant zero expression in all samples (24,477) and performed trimmed mean of M values (TMM) normalization⁵⁰ on the remaining gene counts. Gene expression distribution fitting was then performed using *fitdistrplus*⁵¹, and *mass*⁵², for several distribution types: log-normal, log-logistic, Pareto, Burr, and Weibull. Lastly, an expression cut-off was identified (TMM = 5) and used to filter for genes with expression above the cut-off in at least one sample, with the final number of genes being 13,673.

TMM normalization

To correct for library size and composition biases between samples, Trimmed Mean of M-values (TMM) normalization was applied using the *calcNormFactors()* function from the edgeR package⁵³. This method calculates scaling factors for each sample by comparing log-fold changes (M-values) of gene expression relative to a reference sample, typically the one with median library size. Extreme M-values and lowly expressed genes are trimmed to avoid distortion from outliers, and the resulting factors are used to compute normalized counts per million (CPM). These normalized expression values were then used for all downstream analyses.

Toggle gene extraction

Same-condition sample toggle genes were identified and extracted as defined by Giuliani, et al.²⁷:

$$X_{toggle} = \left\{ x_{ij} \mid (0 \leq x_{i1} < \varepsilon \text{ and } x_{i2} > \varepsilon) \text{ or } (x_{i1} > \varepsilon \text{ and } 0 \leq x_{i2} < \varepsilon) \right\} \quad (1)$$

where, x_{ij} represents the expression vector of the i -th gene for two samples $j = 1, 2$ of the same condition. The parameter ε denotes the minimum expression threshold determined from statistical distribution fitting step above (Table S3). Similarly, temporal toggle genes were extracted using the same criteria across different time (n) points of the same condition: $j = 0, n^{th}$ time point.

For each condition with three biological samples, toggle genes are identified pairwise, meaning that a gene may toggle between any two samples within the condition without requiring toggling across all sample pairs. Similarly, temporal toggle genes were extracted using the same criteria but applied across different time points of the same condition: $j = t_0t_1, t_0t_2, \dots, t_0t_m$, where t_0t_n represents comparison between t_0 and t_n time points, comparing all time points with initial time t_0 . This approach ensures that toggling behavior is evaluated consistently across both same-condition and temporal contexts.

DE gene extraction

Temporal DE genes were extracted using *DESeq2*⁵⁴, using a fold-change of 2 and 4 as indicated in maintext. DE analysis was performed between the initial time points (t_0) and the n -th time points (t_n), where $n > 0$, for both PC and NPC conditions. Only genes that passed a threshold of p -value < 0.05 were retained.

Correlation

Autocorrelation refers to correlation changes with respect to t_0 and is computed by calculating the correlation between t_0 and t_m , respectively. Two auto-correlation metrics were deployed in this analysis: Pearson correlation and Spearman correlation.

Pearson. Pearson correlation between two vectors can be calculated as:

$$r(X, Y) = \frac{1}{n} \frac{\sum_{i=1}^n (x_i - \mu_X)(y_i - \mu_Y)}{\sigma_X \sigma_Y} \quad (2)$$

where μ_X and μ_Y are the mean values for vectors X and Y , and similarly σ_X and σ_Y represent the standard deviations. In the case of autocorrelation, X always refers to the initial time point, and Y to each subsequent time point.

Spearman. Like Pearson correlation, Spearman rank correlation between X and Y is defined as:

$$\rho(X, Y) = 1 - \frac{6 \sum_{i=1}^n (r_{x,i} - r_{y,i})^2}{n(n^2 - 1)} \quad (3)$$

where $r_{x,i}$ and $r_{y,i}$ represent the ranks of the i -th observation (gene) in the initial time point and the considered time point.

Noise

Noise between any two samples was computed using:

$$\eta_{i(jk)}^2 = \frac{\sigma_{i(jk)}^2}{\mu_{i(jk)}^2} = 2 \frac{(x_{ij} - x_{ik})^2}{(x_{ij} + x_{ik})^2} \quad (4)$$

where x_{ij} and x_{ik} are the values of a gene (i) in j th and k th samples. Average noise is calculated by averaging the summed noise values of all genes between all pairs considered giving a final noise formula for m considered genes:

$$n^2 = \frac{1}{m} \sum_{i=1}^m \eta_i^2 \quad (5)$$

For temporal noise, the calculation was performed for each time point with respect to t_0 , and for sample noise, the calculation was performed between all samples of any given sample condition.

Entropy

Shannon entropy was computed for each bulk RNA-seq sample based on the empirical distribution of binned expression values. The number of bins

was determined using Doane's rule,

$$b = 1 + \log_2 \left(1 + \frac{|g|}{\sigma^2} \right) \quad (6)$$

where n is the number of expressed genes, g is the skewness, and σ^2 is the standard error of skewness. Entropy was calculated as:

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (7)$$

where $p(x_i)$ is the proportion of values in bin i . All computations were performed in *R* using a custom implementation.

Hierarchical clustering

Hierarchical clustering for toggle genes and DEG was performed using the *stats* package in *R*, where first a distance matrix between the samples was computed for each corresponding gene set. Next, Ward clustering⁵⁵ method was applied to group genes with similar temporal expression patterns. For each identified cluster, the mean TMM expression across all timepoints was plotted to visualize temporal expression patterns for both PC and NPC.

GO and network analysis

For GO analysis, several analytic tools were performed. Gene enrichment analysis for Biological Processes was performed using *clusterProfiler*⁵⁶ in *R*, using a threshold of p -value < 0.05 . Next, GO networks were generated in Cytoscape using ClueGO⁵⁷, with specificity chosen as global, and a significance threshold below 0.05. Lastly, *Reactome*⁵⁸ pathway analysis was performed to gain further understanding of the enriched pathways within the temporal proliferative signature with a similar threshold of $p < 0.05$.

Data availability

The codes for the analysis used in this manuscript are available from the corresponding authors upon request. The CLL transcriptomic data used in this manuscript can be found in the GEO database using the accession numbers: GSE66117, GSE249956, and GSE130385.

Received: 26 March 2025; Accepted: 1 August 2025;

Published online: 11 August 2025

References

1. Fresa, A. et al. Treatment options for elderly/unfit patients with chronic lymphocytic leukemia in the era of targeted drugs: a comprehensive review. *J. Clin. Med.* **10**, 5104 (2021).
2. González-Gascón-y-marín, I. et al. From biomarkers to models in the changing landscape of chronic lymphocytic leukemia: Evolve or become extinct. *Cancers (Basel)* **13**, 1782 (2021).
3. Lightfoot, T., Smith, A. & Roman, E. Leukemia. *International Encyclopedia of Public Health* 410–418 <https://doi.org/10.1016/B978-0-12-803678-5.00253-8> (2023).
4. Patel, K. & Pagel, J. M. Current and future treatment strategies in chronic lymphocytic leukemia. *J. Hematol. Oncol.* **14**, 1–20 (2021). 2021 14:1.
5. Herman, S. E. M. et al. Ibrutinib-induced lymphocytosis in patients with chronic lymphocytic leukemia: correlative analyses from a phase II study. *Leukemia* **28**, 2188–2196 (2014).
6. Woyach, J. A., Johnson, A. J. & Byrd, J. C. The B-cell receptor signaling pathway as a therapeutic target in CLL. *Blood* **120**, 1175–1184 (2012).
7. Herman, S. E. M. et al. Ibrutinib inhibits BCR and NF-κB signaling and reduces tumor proliferation in tissue-resident cells of patients with CLL. *Blood* **123**, 3286–3295 (2014).
8. Huang, S. Reconciling non-genetic plasticity with somatic evolution in cancer. *Trends Cancer* **7**, 309–322 (2021).

9. Bhat, G. R. et al. Cancer cell plasticity: from cellular, molecular, and genetic mechanisms to tumor heterogeneity and drug resistance. *Cancer Metastasis Rev.* 1–32 <https://doi.org/10.1007/S10555-024-10172-Z> (2024).
10. Boumahdi, S. & de Sauvage, F. J. The great escape: tumour cell plasticity in resistance to targeted therapy. *Nat. Rev. Drug Discov.* **19**, 39–56 (2019). *2019* 19:1.
11. Shi, Z. D. et al. Tumor cell plasticity in targeted therapy-induced resistance: mechanisms and new strategies. *Signal Transduct. Target Ther.* **8**, 113 (2023).
12. Qin, S. et al. Emerging role of tumor cell plasticity in modifying therapeutic response. *Signal Transduct. Target Ther.* **5**, 1–36 (2020). *2020* 5:1.
13. Huang, S. Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. *Cancer Metastasis Rev.* **32**, 423–448 (2013).
14. Bui, T. T. & Selvarajoo, K. Attractor concepts to evaluate the transcriptome-wide dynamics guiding anaerobic to aerobic state transition in *Escherichia coli*. *Sci. Rep.* **10**, 1–14 (2020).
15. Huang, S. & Kauffman, S. How to escape the cancer attractor: rationale and limitations of multi-target drugs. *Semin Cancer Biol.* **23**, 270–278 (2013).
16. Huang, S., Ernberg, I. & Kauffman, S. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin Cell Dev. Biol.* **20**, 869 (2009).
17. Li, Q. et al. Dynamics inside the cancer cell attractor reveal cell heterogeneity, limits of stability, and escape. *Proc. Natl. Acad. Sci. USA* **113**, 2672–2677 (2016).
18. Groves, S. M. & Quaranta, V. Quantifying cancer cell plasticity with gene regulatory networks and single-cell dynamics. *Front. Netw. Physiol.* **3**, 1225736 (2023).
19. Selvarajoo, K. Understanding multimodal biological decisions from single cell and population dynamics. *Wiley Interdiscip. Rev. Syst. Biol. Med* **4**, 385–399 (2012).
20. Kellogg, R. A. & Tay, S. Noise facilitates transcriptional control under dynamic inputs. *Cell* **160**, 381–392 (2015).
21. Dou, Z. et al. HJURP promotes malignant progression and mediates sensitivity to cisplatin and WEE1-inhibitor in serous ovarian cancer. *Int. J. Biol. Sci.* **18**, 1188 (2022).
22. Han, R. et al. Increased gene expression noise in human cancers is correlated with low p53 and immune activities as well as late stage cancer. *Oncotarget* **7**, 72011 (2016).
23. Pina, C. Contributions of transcriptional noise to leukaemia evolution: KAT2A as a case-study. *Philos. Trans. R. Soc. B Biol. Sci.* **379**, 20230052 (2024).
24. Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544–547 (2008).
25. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA* **99**, 12795–12800 (2002).
26. Maamar, H., Raj, A. & Dubnau, D. Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science* (1979) **317**, 526–529 (2007).
27. Giuliani, A., Bui, T. T., Helmy, M. & Selvarajoo, K. Identifying toggle genes from transcriptome-wide scatter: a new perspective for biological regulation. *Genomics* **114**, 215–228 (2022).
28. Dai, X., Healy, S., Yli-Harja, O. & Ribeiro, A. S. Tuning cell differentiation patterns and single cell dynamics by regulating proteins' functionalities in a toggle switch. *J. Theor. Biol.* **261**, 441–448 (2009).
29. Xu, Y. et al. ZNF397 deficiency triggers TET2-driven lineage plasticity and AR-targeted therapy resistance in prostate cancer. *Cancer Discov.* **14**, OF1–OF26 (2024).
30. Kwon, J. et al. USP13 drives lung squamous cell carcinoma by switching lung club cell lineage plasticity. *Mol. Cancer* **22**, 1–24 (2023).
31. Klebe, M. et al. Frequent molecular subtype switching and gene expression alterations in lung and pleural metastasis from luminal A-type breast cancer. *JCO Precis Oncol.* **4**, 848–859 (2020).
32. Cao, Y., Lu, H. M. & Liang, J. Probability landscape of heritable and robust epigenetic state of lysogeny in phage lambda. *Proc. Natl. Acad. Sci. USA* **107**, 18445–18450 (2010).
33. Arkin, A., Ross, J. & McAdams, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected *Escherichia coli* cells. *Genetics* **149**, 1633–1648 (1998).
34. Bermejo, A. V., Ragonnaud, E., Daradoumis, J. & Holst, P. Cancer associated endogenous retroviruses: ideal immune targets for adenovirus-based immunotherapy. *Int. J. Mol. Sci.* **21**, 1–21 (2020).
35. Xu, X., Zhao, H., Gong, Z. & Han, G. Z. Endogenous retroviruses of non-avian/mammalian vertebrates illuminate diversity and deep history of retroviruses. *PLoS Pathog.* **14**, e1007072 (2018).
36. Tokuyama, M. et al. ERVmap analysis reveals genome-wide transcription of human endogenous retroviruses. *Proc. Natl. Acad. Sci. USA* **115**, 12565–12572 (2018).
37. Schleiss, C. et al. Temporal multiomic modeling reveals a B-cell receptor proliferative program in chronic lymphocytic leukemia. *Leukemia* **35**, 1463 (2021).
38. Schleiss, C. et al. BCR-associated factors driving chronic lymphocytic leukemia cells proliferation ex vivo. *Sci. Rep.* **9**, 701 (2019).
39. Kushwaha, G. et al. Hypomethylation coordinates antagonistically with hypermethylation in cancer development: a case study of leukemia. *Hum. Genom.* **10**, Suppl 2, 18 (2016).
40. Pozzo, F. et al. Early reappearance of intracлонаl proliferative subpopulations in ibrutinib-resistant chronic lymphocytic leukemia. *Leukemia* **38**, 1712–1721 (2024).
41. Tien, B. T., Giuliani, A. & Selvarajoo, K. Statistical distribution as a way for lower gene expressions threshold cutoff. *Org. J. Biol. Sci.* **2**, 55–58 (2018).
42. Yong, C. & Gyorgy, A. Stability and robustness of unbalanced genetic toggle switches in the presence of scarce resources. *Life (Basel)* **11**, 271 (2021).
43. Tsuchiya, M. et al. Gene expression waves. Cell cycle independent collective dynamics in cultured cells. *FEBS J.* **274**, 2878–2886 (2007).
44. Liu, K. et al. The multiple roles for Sox2 in stem cell maintenance and tumorigenesis. *Cell Signal* **25**, 1264–1271 (2013).
45. Boeckel, G. R. & Ehrlich, B. E. NCS-1 is a regulator of calcium signaling in health and disease. *Biochim Biophys. Acta Mol. Cell Res.* **1865**, 1660–1667 (2018).
46. Chi, H. Sphingosine 1-phosphate and immune regulation: trafficking and beyond. *Trends Pharm. Sci.* **32**, 16 (2010).
47. Chaudhary, P. K. & Kim, S. An insight into GPCR and G-proteins as cancer drivers. *Cells* **10**, 3288 (2021).
48. Aga, R. B. & Ridley, A. J. Rho GTPases: regulation and roles in cancer cell biology. *Small GTPases* **7**, 207–221 (2016).
49. Sveen, A., Johannessen, B., Teixeira, M. R., Lothe, R. A. & Skotheim, R. I. Transcriptome instability as a molecular pan-cancer characteristic of carcinomas. *BMC Genom.* **15**, 1–13 (2014).
50. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, 1–9 (2010).
51. Delignette-Muller, M. L. & Dutang, C. fitdistrplus: an R package for fitting distributions. *J. Stat. Softw.* **64**, 1–34 (2015).
52. Venables, W. N., Ripley, B. D. *Modern Applied Statistics with S*, 4th edn. (Springer, New York, 2002). <https://www.stats.ox.ac.uk/pub/MASS4/>.
53. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139 (2009).

54. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
55. Ward, J. H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).
56. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284 (2012).
57. Bindea, G. et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091 (2009).
58. Fabregat, A. et al. Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinforma.* **18**, 1–9 (2017).

Acknowledgements

The work was supported by ARIA research scholarship to G.A., and the core budget of Bioinformatics Institute, ASTAR. Figures 1B and 2A were created with “BioRender.com”. The authors thank Dr Prakash Arumugam for critical comments.

Author contributions

Olga Sirbu: conceptualization (supporting), writing original draft, formal analysis (lead), reviewing and editing; Gunjan Agarwal: writing, formal analysis (supporting), reviewing and editing; Alessandro Giuliani: writing, reviewing and editing; Kumar Selvarajoo: conceptualization (lead), writing, reviewing, editing and supervision.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41540-025-00575-1>.

Correspondence and requests for materials should be addressed to Kumar Selvarajoo.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025