Leukemia (2021) 35:1463–1474
https://doi.org/10.1038/s41375-021-01221-5

ARTICLE

Transcriptional control and signal transduction, cell cycle

Temporal multiomic modeling reveals a B-cell receptor proliferative program in chronic lymphocytic leukemia

Cedric Schleiss[1,2] · Raphael Carapito[1,2,3] · Luc-Matthieu Fornecker[4,5] · Leslie Muller[6] · Nicodème Paul[1,2] · Ouria Tahar[1,2,3] · Angelique Pichot[1,2] · Manuela Tavian[4] · Alina Nicolae[4] · Laurent Miguet[4,7] · Laurent Mauvieux[4,7] · Raoul Herbrecht[4,5] · Sarah Cianferani[6] · Jean-Noel Freund[4] · Christine Carapito[6] · Myriam Maumy-Bertrand[2,8] · Seiamak Bahram[1,2,3] · Frederic Bertrand[8,9,10] · Laurent Vallat[1,2,3,11,12]

Received: 1 October 2020 / Revised: 21 February 2021 / Accepted: 9 March 2021 / Published online: 8 April 2021

Abstract

B-cell receptor (BCR) signaling is crucial for the pathophysiology of most mature B-cell lymphomas/leukemias and has emerged as a therapeutic target whose effectiveness remains limited by the occurrence of mutations. Therefore, deciphering the cellular program activated downstream this pathway has become of paramount importance for the development of innovative therapies. Using an original ex vivo model of BCR-induced proliferation of chronic lymphocytic leukemia cells, we generated 108 temporal transcriptional and proteomic profiles from 1 h up to 4 days after BCR activation. This dataset revealed a structured temporal response composed of 13,065 transcripts and 4027 proteins, comprising a leukemic proliferative signature consisting of 430 genes and 374 proteins. Mathematical modeling of this complex cellular response further highlighted a transcriptional network driven by 14 early genes linked to proteins involved in cell proliferation. This group includes expected genes (EGR1/2, NF-kB) and genes involved in NF-kB signaling modulation (TANK, ROHF) and immune evasion (KMO, IL4I1) that have not yet been associated with leukemic cells proliferation. Our study unveils the BCR-activated proliferative genetic program in primary leukemic cells. This approach combining temporal measurements with modeling allows identifying new putative targets for innovative therapy of lymphoid malignancies and also cancers dependent on ligand–receptor interactions.

These authors contributed equally: Frederic Bertrand, Laurent Vallat

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41375-021-01221-5.

✉ Frederic Bertrand
fbertran@unistra.fr

✉ Laurent Vallat
vallat@unistra.fr

[1] Laboratoire d'ImmunoRhumatologie Moléculaire, INSERM UMR-S1109, LabEx Transplantex, Plateforme Genomax, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Université de Strasbourg, Strasbourg, France

[2] Fédération Hospitalo-Universitaire (FHU) Omicare, Université de Strasbourg, Strasbourg, France

[3] Laboratoire d'Immunologie, Plateau Technique de Biologie, Pôle de Biologie, Nouvel Hôpital Civil, Strasbourg, France

[4] Université de Strasbourg, INSERM, IRFAC UMR-S1113, Strasbourg, France

[5] Service d'Hématologie, Institut de Cancérologie Strasbourg Europe (ICANS), Strasbourg, France

[6] Laboratoire de Spectrométrie de Masse BioOrganique, Université de Strasbourg, CNRS, IPHC, UMR 7178, Strasbourg, France

[7] Laboratoire d'Hématologie, Pôle de Biologie, Hôpital de Hautepierre, Hôpitaux Universitaires de Strasbourg, Strasbourg, France

[8] Institut de Recherche Mathématique Avancée, CNRS UMR 7501, LabEx IRMIA, Université de Strasbourg, Strasbourg, France

[9] Present address: Fédération Hospitalo-Universitaire (FHU) Omicare, Université de Strasbourg, Strasbourg, France

[10] Present address: Institut Charles Delaunay, ROSAS, M2S, Université de Technologie de Troyes, Troyes, France

[11] Present address: Université de Strasbourg, INSERM, IRFAC UMR-S1113, Strasbourg, France

[12] Present address: Laboratoire d'Hématologie, Pôle de Biologie, Hôpital de Hautepierre, Hôpitaux Universitaires de Strasbourg, Strasbourg, France

Introduction

In B-cell lymphomas and leukemias such as marginal zone lymphoma and chronic lymphocytic leukemia (CLL), chronic antigenic activation of the B-cell antigen receptor (BCR) sustains aberrant lymphocyte behavior and uncontrolled monoclonal proliferation [1]. In physiological condition, BCR-mediated cell activation is crucial for proliferation and differentiation of lymphoid progenitors [2], as well as for recognition of pathogen-derived antigens by mature B lymphocytes. By contrast in CLL, BCR-mediated lymphocyte activation by various antigens [3], reinforced by microenvironmental factors [4], triggers aberrant cell proliferation in CLL's proliferation centers [5]. The exploration of pathways that transfer information from BCR engagement to the nucleus revealed signaling aberrations that are reinforced by ectopic expression of the protein kinase ZAP70 in the most aggressive forms of CLL [6]. The use of specific inhibitors targeting the tumor cell survival dependency on key signaling proteins (BTK, Pi3K) [7] has proven its efficacy in clinics, however mutations in elements of these pathways ultimately lead to tumor resistance and escape. Thus, the need for alternative therapeutic approaches requires identifying novel targets. The genetic program downstream the BCR signaling cascades, namely the resulting sequential and concerted expression of multiple genes and proteins, is a promising candidate, but it still remains poorly understood.

Ex vivo models of BCR stimulation in patients' primary CLL cells [5, 8–10] have been developed to decipher these downstream genetic programs. However, transcriptional responses analyzed at sparse time points after BCR engagement have yielded only limited insight into the complexity of the cellular response. Moreover, it is worth noting that the experimental conditions using isolated BCR activation in these previous studies led to CLL lymphocyte apoptosis instead of proliferation [11], which therefore missed the malignant proliferative output of BCR engagement observed in patients. Recently, we designed a novel ex vivo experimental setting in which BCR engagement coupled to minimal mandatory costimulating agents (CD40L, IL-4 and IL-21) recapitulate the proliferation of primary CLL cells [12].

In the present study, we used this improved ex vivo culture model to generate a unique set of 108 combined transcriptional and proteomic profiles over time after activation of human primary CLL cells. As existing analysis methods of such high dimensional datasets are limited in terms of precision to select the relevant actors of the genetic program, we have developed a mathematical approach, validated on synthetic datasets and supported by extensive simulations, allowing the selection of critical actors of the CLL's cell response [13]. Furthermore, in order to characterize the underlying structure of the concerted temporal interactions between these actors, we refined our previously developed regression method based on linear equations [14] which has proved its capacity to handle high dimensional dataset while taking into account the inherent sparsity of biological processes [15, 16]. By using this mathematical modeling approach applied to the 108 omics points of measurement, we characterized the temporal cellular response of CLL cells to BCR activation and we identified within this response a nested and structured core proliferative program that could sustain CLL cell leukemogenesis.

Materials and methods

Subjects, B-cell isolation, and culture conditions

B cells from peripheral blood were obtained from six untreated CLL patients whose biological characterization, performed at the University Hospitals of Strasbourg, included flow cytometry analysis, cytogenetic with FISH, IGHV status, and TP53 mutational profile. All selected patients had a Matutes score of 5/5, unmutated IGHV, and wild-type TP53 (Table 1). This study was approved by the ethic committee (CPP Est IV) of

Table 1 Clinical and biological characteristics of CLL patients.

Sample	Sex	Age at diagnosis	IGHV status[a]	VH identity (%)	ZAP70 status[b]	CD38[c]	cytogenetic	Binet stage	Lymphocytes (G/L)	BCR response[d]
CLL-P1	F	70	UM	100	Pos	Neg	tri12	A	18	P
CLL-P2	M	67	UM	99	Pos	Neg	tri12	A	22	P
CLL-P3	F	72	UM	100	Pos	Pos	del13q	A	11	P
CLL-NP1	F	58	UM	100	Pos	Neg	0	A	28	NP
CLL-NP2	M	61	UM	100	Pos	Pos	0	A	56	NP
CLL-NP3	F	55	UM	99	Pos	Neg	del13q	A	15	NP

[a]≥ 98% of IGHV identity for defining unmutated (UM) CLL cells [52].

[b]<7 threshold of T cells/CLL cells ratio of ZAP70 mean fluorescence intensity expression for defining ZAP70-positive CLL cells [53].

[c]≥30% threshold for defining CD38-positive CLL.

[d]>25% of cell division-dependent decrease in CFSE staining intensity measured by flow cytometry at day 6 after initial B-cell receptor activation for defining proliferative (P) samples, and <20% for defining nonproliferative (NP) samples.

Strasbourg University Hospitals and all patients gave written informed consent. CLL B cells were negatively selected using the RosetteSep B-cell enrichment cocktail (STEMCELL Technologies, Vancouver, Canada) and density gradient centrifugation. CLL cells (>96% CD19+/CD5+) were stained with 0.5 µM carboxyfluorescein succinimidyl ester (CFSE) (CellTrace, Thermo Fisher, Waltham, MA, USA) and were stimulated with F(ab′)2 anti-human IgM (Jackson ImmunoResearch, West Grove, PA, USA), CD40L (Enzo Life Sciences, Villeurbanne, France), IL-4 (R&D Systems-Bio-Techne, Lille, France), and IL-21 (Invitrogen, Maryland, USA) in soluble medium as described previously [12]. Cell apoptosis, evaluated using FITC-Annexin V and propidium iodide (apoptosis detection kit, BD Biosciences, San Jose, CA, USA), 4 days after BCR engagement by flow cytometry [12] (Cytomics FC500, Beckman-Coulter, Fullerton, CA, USA) showed 86–98% live cells (median: 89%) in all samples. B-cell proliferation, defined by cell division-dependent decrease in CFSE staining intensity, was monitored 6 days after BCR engagement by flow cytometry as previously described [12].

Transcriptomic analysis

Before BCR engagement (T0) and at eight time points after activation (1 h, 1 h 30 min, 3 h 30 min, 6 h 30 min, 12 h, 24 h, 48 h, 96 h), 4.10^6 cells were resuspended in 1 mL TRIzol (Sigma-Aldrich, Saint-Louis, MO, USA). Total RNA was purified using the RNeasy Mini kit (Qiagen, Hilden, Germany). After ribosomal RNA depletion, the sequencing library was prepared with the Ion Total RNA-seq kit v2 (Thermo Scientific) and the sequencing was performed on an Ion Proton sequencer with the Ion PI Hi-Q Sequencing 200 Kit (Thermo Scientific). Reads was estimated with the package HTSeq [17] and the edgeR package [18] was used to derive the reads per kilobase per million values. The transcriptomic dataset is available in GEO (GSE130385).

Proteomic analysis

Before BCR stimulation (T0) and at eight time points after stimulation (1 h, 2 h, 4 h, 7 h, 12 h, 24 h, 48 h, 96 h), 8.10^6 cells were resuspended in lysis buffer. Proteins were acetone precipitated and 10 µg of each sample were concentrated in a stacking gel band, in-gel reduced, alkylated, and trypsin digested. NanoLC-MS/MS analyses of extracted trypsic peptides were performed on a nanoAcquity UPLC device (Waters Corporation) coupled to a Q-Exactive Plus mass spectrometer (Thermo Scientific) operated in data dependent acquisition mode. Label-free extracted ion chromatogram-based quantification was performed using MaxQuant software (version 1.5.5.1) [19]. The proteomics dataset was deposited to the ProteomeXchange Consortium via the PRIDE partner repository (PXD013573).

Gene expression and protein abundance analysis

Quality-based filtering of low expressed genes was performed with the HTSFilter package [20]. The selection of differentially expressed (DE) genes was made with the glmLRT and the glmTreat functions of the edgeR package [18]. Identification of temporal clusters of gene expression was performed with the HTSCluster package [21]. After quantile normalization, differential analysis of protein abundancies was made using the peptide-level robust ridge regression implemented in the MSqRob package [22].

Clustering and network reverse engineering

After selection according to their differential expression and temporal profile, genes and proteins were divided into temporal clusters for network reverse engineering. We had to model N gene or protein actors for the reverse engineering across $T = 8$ time points and for a number of $P = 3$ individuals (3 proliferative and 3 nonproliferative samples); we denote by x_{npt} the observed value (gene expression or protein abundancy) of the actor n for an individual p at a time point t. For any actor of the network n among the total N, the mathematical model was written

$$\widetilde{x}_{np.} = \sum_{n'=1}^{N} \omega_{n'n} \boldsymbol{F}_{m(n')m(n)} \widetilde{x}_{n'p.} + \boldsymbol{\varepsilon}_{np}, \quad 1 \le p \le P.$$

In this model, N is the total number of actors, $k \mapsto m(k)$ is the function that maps an actor to its time cluster, $\boldsymbol{F_{ij}}$ is a T square matrix that describes the action of the actors belonging to cluster i on an actor that belongs to cluster j, ω_{kl} is the strength of the connection from actor k toward actor l and $\boldsymbol{\varepsilon_{np}}$, and $1 \le p \le P$ is a T dimensional random vector with zero mean and unit variance I_T.

The code written for selection of actors and reverse engineering the temporal cellular program in this study is available as an R-package (https://fbertran.github.io/Patterns/) [16].

Experimental procedure is summarized in Fig. S1 and methods are detailed in Supplementary Information.

Results

Identification of a structured proliferative signature in BCR stimulated CLL cells

Experimental design and multiomic dataset

To investigate the proliferative response of primary human CLL cells after ex vivo BCR engagement, BCR stimulation was performed on six untreated CLL samples of the most

aggressive subgroup (IGHV unmutated), including three samples that proliferate and three samples that do not proliferate ex vivo within our culture conditions, the latter being used as controls (Table 1). Transcriptional (RNA-seq) and proteomic (mass spectrometry) responses of these CLL samples were determined at T0 (before stimulation) and at eight time points between 1 and 96 h after BCR-mediated cell activation, generating a total of 108 points of measurement (Fig. S1). A total of 23,348 transcripts and 4664 unique proteins were identified and quantified in the whole dataset. After quality-based filtering, 13,065 transcripts and 4027 proteins expressed at least in one of the 108 samples were retained for further analysis.

Unsupervised analysis identified a structured BCR response

The temporal transcriptional response was explored by unsupervised multidimensional scaling which summarizes within one dot on a two-dimensional graph the 500 most expressed genes of each sample (Fig. 1A). This representation revealed the temporality of the response on the *X* axis and the proliferative status of the samples on the *Y* axis. The dots corresponding to the proliferative and nonproliferative samples were separated at T0 along the *Y* axis and they all followed a structured evolution from T1 to T8 along the *X* axis after BCR engagement. This analysis emphasized the structured nature of the transcriptional BCR response. In addition, hierarchical clustering analysis strengthened this structured nature since it identified four clusters, each made of consecutive time point of measurement (Fig. 1B). Moreover, unsupervised temporal gene expression analysis revealed clusters of genes exhibiting structured temporal patterns of expression (Fig. S2A), characterizing the transcriptional response to an exogenous stress [23, 24].

In comparison, the unsupervised proteomic analysis appeared less structured after BCR stimulation than the transcriptional one, mainly at early time points (T1–T6). However, a tendency for samples' segregation with respect of their proliferative response was observed at later time

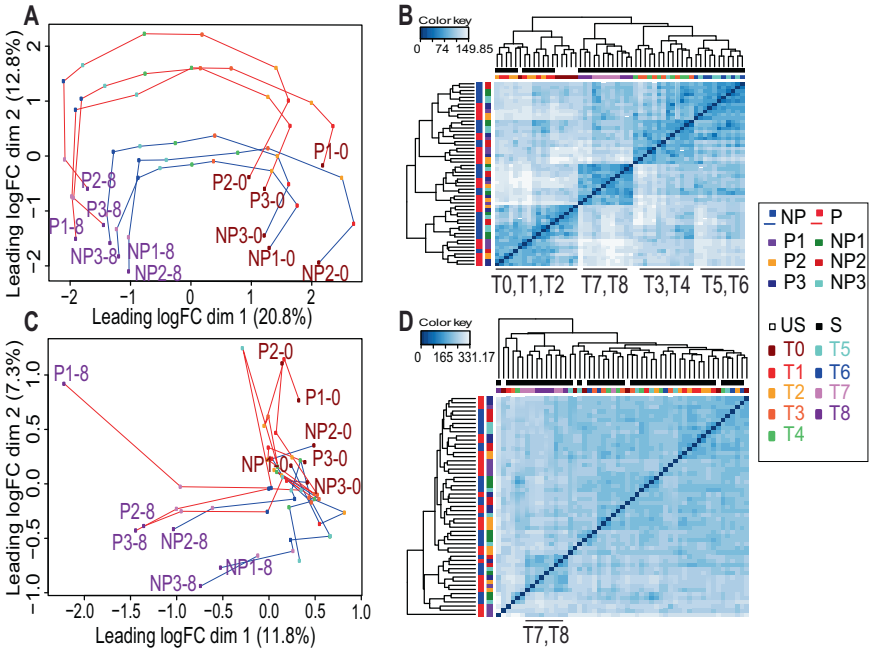


Fig. 1 Unsupervised statistical analysis of genes and proteins expression. A Multidimensional scaling plot (MDS) analysis based on the expression of the 500 most expressed genes for each pairwise comparisons between the samples (among a total of 13,065 normalized gene expressions), analyzed before (T0) and at eight time points (T1–T8) after ex vivo B-cell antigen receptor activation for six chronic lymphocytic leukemia (CLL) patients (three proliferative samples (P1-3) and three control nonproliferative samples (NP1–3)). The MDS graphs were constructed from the LogFC of the expressions/abundances at different time points (T1–T8 versus T0). Each dot represents the transcriptional profile of one CLL cell sample at a specific time point. A color code represents the different time points. Successive time points of a same cell sample are linked in the graph (red line for proliferative samples and blue line for nonproliferative samples). **B** Hierarchical clustering of all samples and all time points, based on the expression of the 500 most expressed genes. Dendrograms from clustering are added to the left side and to the top of the image. The abbreviations of the times (T0–T8) represented in the different time clusters observed on the hierarchical clustering are shown at the bottom. **C** MDS analysis based on the expression of the 500 most abundant proteins for each pairwise comparisons between the samples, analyzed before and after ex vivo cell activation for the six CLL patients. Each dot represents the proteome of one CLL cell sample at a specific time point before (T0) and at eight time points (T1–T8) after cell stimulation. A color code represents the different time points. Successive time points of a same cell sample are linked in the graph. **D** Hierarchical clustering of all samples and all time points, based on the expression of the 500 most abundant proteins.

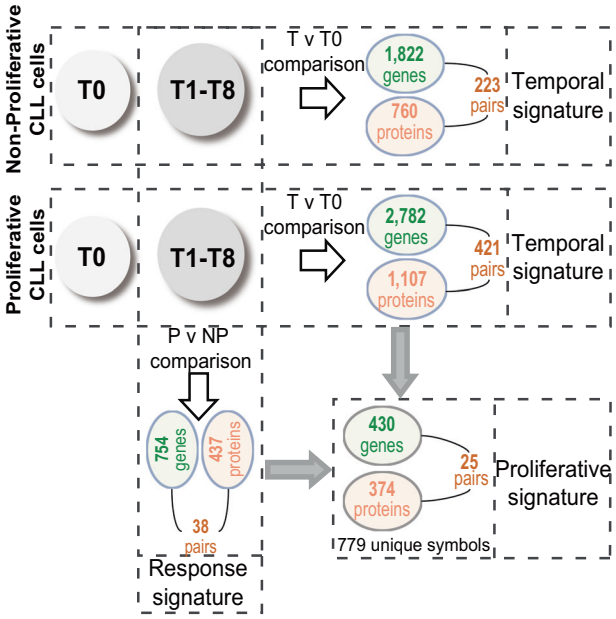


Fig. 2 Supervised statistical analysis of genes and proteins temporal expression. Temporal signature (T versus T0 comparison, horizontally). Number of genes differentially expressed (DE) and proteins differentially abundant (DA) in proliferative ($n = 3$) and control nonproliferative CLL samples ($n = 3$) across time after cell stimulation (T1–T8), compared to initial (T0) expression/abundance (FDR < 1%). Response signature (proliferative versus nonproliferative comparison, vertically). Number of DE genes and DA proteins in proliferative CLL cells, compared to nonproliferative cells (FDR < 5%). Proliferative signature (combination of T versus T0 and proliferative versus nonproliferative comparisons). The intersection of the list of DE genes and DA proteins expressed in proliferative samples after BCR engagement compared to T0 (T versus T0), and the list of DE genes and DA proteins in the proliferative samples compared to the nonproliferative samples (proliferative versus nonproliferative) identifies the "proliferative signature" of genes and proteins specifically DE/DA after stimulation in proliferative samples.

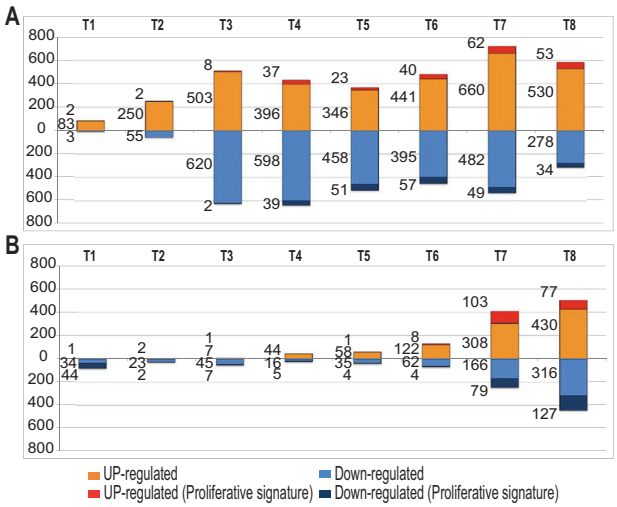


Fig. 3 Supervised statistical analysis of genes and proteins temporal expression in proliferating CLL cells. A Number of DE genes and **B** number of DA proteins at each time point after cell stimulation in proliferating CLL cells. At each time point (T1–T8), the number of genes and proteins up- or downregulated (T versus T0 Log2FC) are shown in orange or blue, respectively. The number of genes and proteins specifically up- or downregulated in proliferative compared to nonproliferative cells are shown in dark orange or dark blue respectively in the graph.

points (T7 and T8) (Fig. 1C). This was further confirmed by hierarchical clustering (Fig. 1D). In addition, the unsupervised temporal protein expression analysis allowed identification of clusters of proteins with structured patterns of abundance modulation after stimulation (Fig. S2B).

The supervised analysis revealed a proliferative signature

Having evidenced the structured nature of the global CLL cell response to BCR activation, we next characterized the proliferative signature within this response, defined as the genes and proteins DE in the proliferative samples compared to the nonproliferative samples after BCR-mediated CLL cells activation. To determine this signature, we first established the list of genes and proteins significantly (FDR: <1%) up- or downregulated in the stimulated (T1–T8) versus unstimulated (T0) samples, defining a temporal signature (Fig. 2, Table S1). In the proliferative samples, 2782 DE genes and 1107 differentially abundant (DA) proteins were assigned to this temporal signature, from which 421 were pairs of common symbols (gene and corresponding protein). The nonproliferative CLL samples showed a less important temporal signature with 1822 DE genes and 760 DA proteins, from which 223 were common symbols. Secondly, we analyzed the list of genes and proteins significantly up- or downregulated (FDR: <5%) in the proliferative versus nonproliferative samples over the T1–T8 timecourse (Fig. 2, Table S1). This response signature comprised 754 DE genes and 437 DA proteins from which 38 were common symbols. The intersection of the temporal signature with the response signature of the proliferative samples showed 430 DE genes and 374 DA proteins, corresponding to 779 unique symbols, characterizing the proliferative signature after BCR engagement (Fig. 2, Table S1).

Strong gene-to-protein correlation within the temporal signature of the proliferative samples

As expected, we observed a delay between the transcriptional and the translational response after cell stimulation. Indeed, as many as 1133 genes (511 upregulated and 622 downregulated) were already DE in proliferative samples in the 3 h after BCR stimulation (Fig. 3A), whereas the proteomic modulation became obvious only after 24–48 h (Fig. 3B). Moreover, analyzing the correlation rate between gene expression and protein abundance in the set of 421

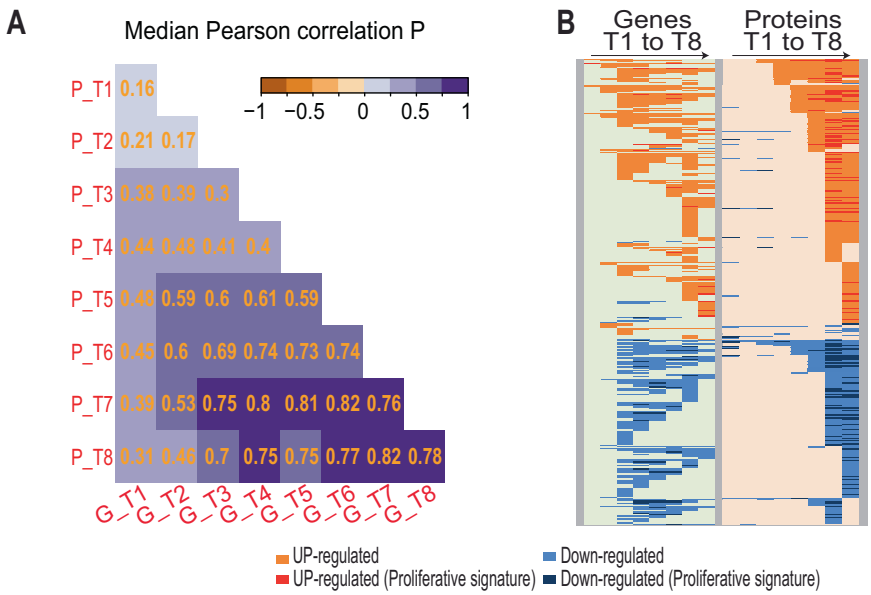


Fig. 4 Correlation of gene expressions and protein abundancies. A Correlation between gene (G) and corresponding protein (P) at each time point after BCR engagement. The median value of the individual Pearson gene/protein correlation is indicated and represented with a color scale. **B** Heat map of the temporal expression/abundance of the 421 gene–protein pairs in the proliferating cells. Each line represents the temporal expression of a gene and its corresponding protein. At each time point, upregulated (T versus T0-positive Log2FC) or downregulated (T versus T0-negative Log2FC) genes and proteins are shown in red or blue, respectively.

genes/proteins pairs of the temporal signature revealed a low median Pearson correlation at the initial time points (38% at 6 h) which strikingly increased up to 82% at 48–96h (Fig. 4A). This was further confirmed by the heat map of temporal expression of these 421 gene/protein pairs (Fig. 4B) showing 90% of concordance between DE genes and DA proteins expression, with a median translation delay of 6 h.

Functions of proteins involved in the temporal signature

We next analyzed the function of the 1107 up- or down-regulated proteins of the temporal signature in the proliferative CLL cells. Biological process annotations (GO BP terms) were collected in order to calculate the number of proteins involved in each process during the timecourse after cell activation. Albeit no significant proteomic enrichment was noticed at early time point, an increase in the number of upregulated proteins related to signaling, transcriptional activity, and cell activation was observed between 7 and 24 h (T4–T6), and a second upsurge of upregulated proteins related to signaling, metabolism, transcriptional processes, cell cycle regulation, DNA replication, nuclear division, and proliferation occurred from days 2 to 4 (T7 and T8) (Fig. S3). The most represented functions during this last period were related to cell cycle regulation, DNA replication, nuclear division, and proliferation, consistent with the onset of proliferation observed in these cells after 4 days post BCR stimulation. Also, the number of proteins participating in antigen processing and peptide presentation was increased, consistent with BCR stimulation in lymphoid cells.

Looking at GO BP for downregulated proteins revealed a transiently decreased number of proteins related to signaling, metabolism, and differentiation within 1 h (T1) after cell activation, potentially reflecting a catabolism phenomenon (Fig. S3). The number of downregulated proteins remained low until 24 h (T2–T6) and no particular BP enrichment could be evidenced. However, a specific enrichment in downregulated proteins related to signaling, transcription, and cell cycle was observed at days 2–4 (T7 and T8), suggesting a negative control of these pathways at later time points.

Altogether, this multiomic approach highlights the structured nature of the temporal response to BCR stimulation in primary CLL cells, characterized by an early transcriptional component progressively relayed by a proteomic component including elements related to the onset of cell proliferation.

Mathematical modeling of the CLL proliferative program

BCR response program inference in proliferative CLL cells

To model the cellular response displayed in the multiomic dataset, we used a mathematical unsupervised reverse engineering approach based on regression and system of

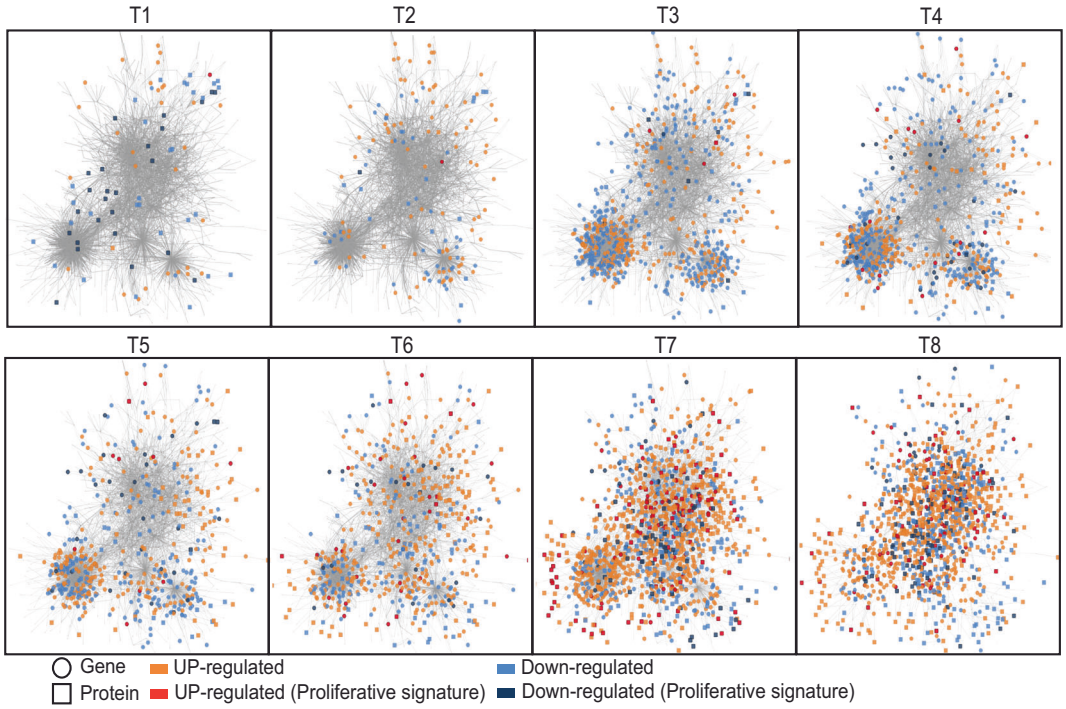


Fig. 5 Temporal propagation in the transcriptional and proteomic network of 2167 genes and 1074 proteins induced after BCR stimulation in proliferating cells. Temporal graphical representation of statistical interactions (arrows) between genes (circle) and/or proteins (square) across time in the proliferating CLL cells after B-cell receptor stimulation. A color code represents genes and proteins differentially (DE/DA T versus T0) upregulated (orange) or downregulated (blue) at each time point after cell activation. Genes or proteins specifically up- or downregulated in proliferating cells from the proliferative signature (DE/DA T versus T0 and DE/DA P versus NP) are represented in dark orange or dark blue, respectively. Graphical representation made with Cytoscape software.

equations that already proved its efficiency in our previous transcriptional study [15]. Temporal matrix of interactions between genes and proteins (Fig. S4A–C) was estimated by mean of penalized regression using a weighted variant of stability selection algorithm [25] in order to retain the best potential regulators of each gene or protein in the network and to determine the timing of these interactions after cell stimulation. To enhance the reverse engineering relevance [26], we imposed biological constraints by favoring links based on known transcriptional or protein–protein relationships implemented in the RegNetwork database [27]. The robustness of this inference has been ascertained by cross validation [28] and the best result was retained by linear regression estimation. Performances of the resulting model, including sensitivity, precision, predictive positive value, and *F*-score, were validated with simulated data and compared to the performance of other algorithms (Fig. S4D–F).

Inferring the formalized model with the temporal dataset of proliferating CLL cells identified a regulatory network of 2167 genes and 1074 proteins representing 2846 unique symbols (Table S2), among which 395 gene–protein pairs, connected by 53,131 oriented links (Fig. S5). This network exhibited a scale-free topology, where a limited number of hub genes and proteins with multiple links (12 genes and 52 proteins exhibit ≥ 10 statistical links) drove the structure of this cellular program. A detailed graphical representation allowed reconstructing the timeline of this program (Fig. 5). This temporal representation showed the DE genes and DA proteins at each time point after cell stimulation and revealed the dynamic propagation of the transcriptional and proteomic expression waves after BCR stimulation in proliferative CLL cells.

Deciphering a CLL proliferative program within the BCR response

Considering the ability of CLL cells to generate a proliferative response after BCR activation, we investigated whether a subnetwork sustaining cell proliferation can be identified within the above response network. To address this, we identified within the proteomic dataset of proliferating CLL cells the 267 proteins associated with the BP terms “cell cycle regulation” and “proliferation,” designed as “seeding proteins,” and their 243 connected neighbors (gene or protein) in the network (Fig. S6A). Analyzing the level of interconnection within this subgroup of proteins and genes revealed a nested subnetwork comprising 388 elements including 173 of the seeding proteins linked to 215 neighbors (Fig. S6B, Table S2). Of note, among these

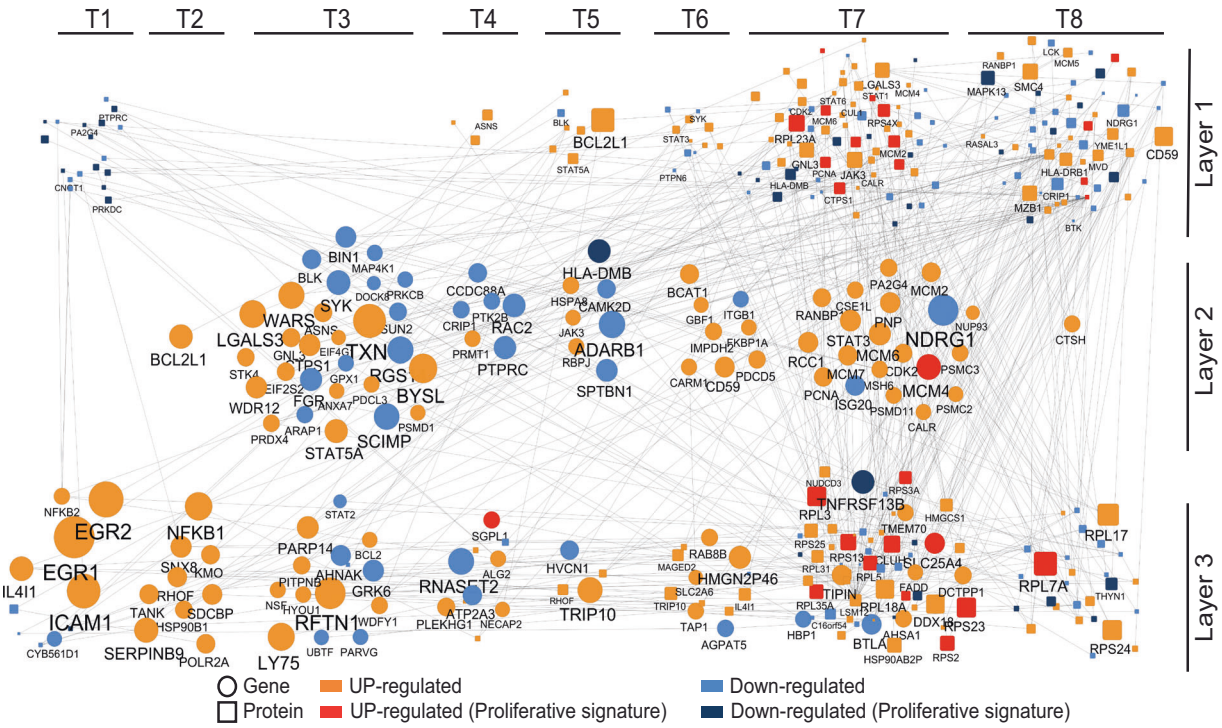


Fig. 6 Nested temporal proliferative program induced after BCR stimulation in proliferating CLL cells. The proliferative temporal subnetwork is represented in a time ordered graph, with genes (circle) and proteins (square) represented at their first time point of differential expression after cell activation (first time DE/DA T versus T0). Genes and proteins up- or downregulated are represented in orange or blue, respectively, and size of circles and squares are proportional to fold changes (Log2FC T versus T0). The 173 seeding proteins involved in "cell cycle" or "proliferation" are grouped in the upper part of the graph (layer#1). The 71 genes coding some of these 173 proteins are represented in the middle (layer#2). The 50 genes and 94 proteins also included in this proliferative subnetwork are grouped in the lower part of the graph (layer#3).

388 genes and proteins elements, 31% belonged to the "proliferative signature" defined above in the supervised analysis (see also Fig. 2). This nested subnetwork could be stratified into three layers of actors (Fig. S6C).

The first layer corresponds to the 173 seeding proteins associated with cell cycle regulation or proliferation processes. These proteins segregate within two groups according to their chronological involvement during the temporal response (Fig. 6). The first temporal group comprises 18 proteins, downregulated at T1 after cell activation. These proteins are mainly involved in transcriptional repression (CNOT1, PA2G4), negative regulation of BCR signaling (INPP5D), or apoptotic process (PRKDC). The second group is made of 155 proteins whose changes occurred from T4 onward, just preceding the initiation of cell proliferation. Within this group, 85 proteins are upregulated and 70 are downregulated proteins. The upregulated proteins comprised factors involved in G1/S transition or DNA replication (e.g., PCNA, CDK2, CUL1, RANBP1, MCM), whereas downregulated ones show elements involved in signaling downstream of the BCR (BLK, BTK, LCK, SYK), potentially reflecting negative regulation mechanisms after BCR engagement.

The second layer corresponds to the 71 genes present in the subnetwork that encode proteins of the first layer (Fig. 6). Remarkably, expression modulation of these transcripts is highly correlated with the modulation of abundance of their corresponding proteins with an offset of 3–6 h.

The third layer of this subnetwork comprises 144 elements made of 50 genes and 94 proteins. At early time points (T1 and T2), only 19 elements are present of which 14 genes exhibit a strong upregulation corresponding to the very early transcriptional response to CLL cell activation. These genes encode major transcription factors (TFs) involved in the G0/G1 transition (EGR1, EGR2), in the regulation of B-cell proliferation, and differentiation after BCR activation (NFKB1, NFKB2) which are also involved in lymphomagenesis [1]. Another genes within this group encode molecules involved in signaling and NF-kB modulation (TANK, RAS homolog family member F (RHOF), syndecan-binding protein (SDCBP)), immune evasion (KMO, IL4I1, SERPINB9), and cell adhesion (ICAM1).

Remarkably, the mathematical modeling of the temporal multiomic data allowed to trace back the sequential organization of the proliferative CLL response from the protein

effectors of cell proliferation at day 4 to the early molecular events induced after BCR activation.

Important actors of the proliferative subnetwork are missing in nonproliferative cells

We next investigated if this proliferative subnetwork composed of TFs and hub proteins is specific to the CLL proliferative program. For this purpose, we inferred with the same approach the formalized model with the temporal datasets of nonproliferative samples. This led to the identification of a regulatory network of 1399 genes and 750 proteins representing 1933 unique symbols, which were connected by 17,332 oriented links (Table S2). Functional analysis (BP terms) revealed 193 proteins associated with "cell cycle regulation" or "proliferation" process. Among them, 89% were shared with the seeding proteins displayed in the proliferative samples and corresponds to the core response program.

The search for neighbors of these 193 proteins revealed 131 elements with a nested subnetwork comprising 114 actors (37 genes and 77 proteins) (Fig. S7). Looking at this subnetwork, we identified a lower number of links and hubs in contrast to the proliferative samples (only 3 elements show ≥5 statistical links in the nonproliferative subnetwork, compared to 15 elements in the proliferative subnetwork). In addition, the nonproliferative samples exhibited a remarkable difference with the proliferative samples at the early time points (T1 and T2) where only 2 of the 14 early BCR-responsive genes were identified and those did not include the TFs EGR1, EGR2, NFKB1, and NFKB2. This observation emphasizes the critical relevance of these 14 early responsive genes in sustaining a BCR-mediated cell proliferation program in CLL cells.

Discussion

The characterization of the cellular program sustaining CLL cells proliferation after BCR engagement is a major step to understand mature B-cell leukemogenesis with the ultimate goal of developing innovative therapies targeting the nuclear response to BCR activation instead of the cytoplasmic pathways that can be bypassed in resistant cancer cells. However, studying the proliferative program in primary CLL cells is challenging because of the difficulty to experimentally recapitulate cell proliferation ex vivo. Here, we used the recently developed culture model based on BCR engagement to induce CLL cell proliferation [12] which allowed, by a temporal multiomic approach, deciphering the dynamic and structured nature of the proliferative program triggered by BCR activation coupled to costimulating agents. For this study, special attention was paid to the similarity of the pathological characteristics among patients (untreated, IGHV unmutated, Binet A/Rai 0/1 stage), which was retrospectively attested by the absence of DE genes and DA proteins with an FDR > 5% at T0 before BCR engagement.

Our large multiomic study highlights in human primary cancer cells the coordination between the dynamic gene and protein responses after exogenous cell stimulation. So far, only few studies have addressed this relationship showing a relatively weak gene-to-protein correlation of 30–60% in yeast [29–31], murine fibroblasts, or human cancer cell lines [32, 33]. Although we showed only a 38% correlation in the first 6 h after BCR engagement, the ratio strikingly increased up to 82% at the later time points in the proliferative cells. This witnesses the progressive emergence and propagation of the organization of transcription subsequently translated into a functional protein pattern triggered by cell activation. In addition, we observed a similar median delay of 3–6 h between gene and protein expression as in yeast [29, 31]. This delay could explain the lack of temporal structure revealed by the unsupervised multidimensional scale analysis of protein abundance early after cell stimulation in contrast to the structured response displayed at the transcriptional level.

Remarkably, the proliferative subnetwork identified by modeling the response of aggressive CLL lymphocytes to BCR activation with costimulating agents comprised several genes encoding important TFs downstream the BCR signaling pathway, including genes previously identified in common in the lymphocytic response to BCR alone in healthy donors and patients with indolent or aggressive CLL [15] (Table S2). This indicates that aggressive CLL lymphocytes still retain similarities with healthy lymphocytes for their response to BCR, and is consistent with the ability of the temporal multiomic modeling approach used here to reconstruct the temporal and functional relationships from the first TFs committed 1 h after BCR engagement to proteins sustaining proliferation days after stimulation. Moreover, if we retain the 374 proteins of the proliferative signature (Fig. 2), instead of retaining the 267 proteins with a GO term of proliferation as seeding proteins, the modeling also highlights a subnetwork comprising 13 of these 14 overexpressed genes, which shows the robustness of this approach.

Among the TFs identified here in the response of aggressive CLL lymphocytes, EGR1 and EGR2 are zinc-finger TFs downstream of the Ras/Raf/MAP kinase pathway that is constitutively activated in various cancers and blood malignancies [34]. EGR1 induces survival and a proliferative response in quiescent cells and is a major driver of mature B-cell lymphomas [35]. Other upregulated genes belonging to the proliferative subnetwork include two members of NF-kB family, NFKB1 (p50) responsible for

transient response after cell stimulation through the canonical NF-kB pathway and NFKB2 (p52) which is crucial for cell differentiation through the noncanonical NF-kB pathway. Via the transcriptional activation of several anti-apoptotic genes, NF-kB members promote survival and proliferation of various cell types. This pathway is also crucial in B-cell leukemogenesis [1] and constitutive NF-kB activation has been described in several B-cell neoplasms [36]. Strikingly, the subnetwork also highlights genes encoding signaling proteins potentially modulating NF-kB activation, but whose implication in leukemia or lymphoma leukemogenesis has not been explored yet. For example, the TRAF family member-associated NF-KB activator (TANK) modulates NF-kB activation through binding with TRAF and TBK1 proteins [37, 38]. The RHOF, representative of the Rho GTPase family implicated in tumorigenesis by regulating cytoskeleton's dynamic [39], has a potential role in germinal center formation [40] and has recently been involved in NF-kB regulation [41]. The SDCBP gene encodes a PDZ domain-containing protein, involved in exosome biogenesis [42] and Rho GTPase family regulation [43], and participates in NF-KB activation in melanoma [44]. The sorting nexin 8 gene, involved in endocytosis and endosomal sorting, interacts with JAK1 and IKKβ and also regulates NF-kB [45].

Furthermore, this subnetwork suggests a prominent activation of immune-evasion mechanisms of CLL cells after BCR and associated factors mediated cell proliferation activation. The genes KMO and IL4I1, respectively, involved in tryptophan catabolism [46] and germinative center formation [47] have T-cell proliferation inhibition abilities [46, 48]. This subnetwork also highlights the role of agents associated and acting in synergy with BCR activation in the ex vivo stimulation model. Among the 388 actors of this subnetwork, ten genes (BCL2L1, EGR2, FGR, ICAM1, PCNA, PRDX4, SERPINB9, STAT3, TAP1, TXN) have also been reported in the transcriptional signature of CD40L [49]. SERPINB9, a serine protease, protects cells from granzyme B associated apoptosis induced by cytotoxic T cells [50] and its expression correlates with clinical outcome of several lymphomas [51].

Noteworthy, the nonproliferative cells response did not exhibit most of the above genes, validating the composition of the core subnetwork of the BCR-mediated cell proliferation. However, comparison of the 388 actors (genes or proteins) of the proliferative subnetwork (Fig. 6) and the 114 actors of the nonproliferating subnetwork (Fig. S7) shows 60 common actors (representing 52% of the total nonproliferating actors and 15% of the proliferating subnetwork actors), constituting the core of the common response of this group of lymphocytes of the aggressive form of CLL.

In conclusion, using a large dataset of temporal transcriptional and proteomic measurements coupled with mathematical modeling, this study unveils the genetic program downstream the signaling cascade activated by the BCR engagement and triggering primary CLL cell proliferation ex vivo. Noteworthy, this program organizes around a limited number of genes and proteins whose sequential commitment drives the cellular response leading to proliferation days after cell activation. These hubs represent potential targets for the development of novel therapeutic strategies for the treatment of aggressive CLL. Beyond CLL, such an approach could be explored in other mature B and T antigen-driven malignancies and could also be extended to other cancer types dependent on ligand–receptor interactions, as for instance the hormone-dependent cancers.

Data availability

RNA-seq data: GEO accession GSE130385. Proteomic data: ProteomeXchange Consortium accession PXD013696. Packages availability: The Cascade package is available at https://CRAN.R-project.org/package=Cascade and https://fbertran.github.io/Cascade/. The SelectBoost package is available at https://fbertran.github.io/SelectBoost/. The Patterns package is available at https://fbertran.github.io/Patterns/.

Acknowledgements The authors are grateful to Antoine Hanauer (Inserm UMR S-1109) and Nathalie Perrusson (Inserm UMR S-1113) for technical help. The authors thank Dr. Anne-Cécile Galoisy and Dr. Caroline Mayeur-Rousse (Laboratoire d'Hématologie, Hôpitaux Universitaires de Strasbourg, France) for CLL cell characterization, Beatrice Uring-Lambert (Laboratoire d'Immunologie, Hôpitaux Universitaires de Strasbourg, France) for help with flow cytometry, Elise Toussaint and Blandine Guffroy (Service d'Hématologie Adulte, Hôpitaux Universitaires de Strasbourg, France) for providing CLL samples, and Jozo Delic (Commissariat à l'Energie Atomique, Fontenay-aux-roses, France) for helpful suggestions.

Author contributions CS and OT performed biological experiments and analyzed data. L-MF, LMu, CC, and SC performed proteomic analysis and analyzed data. AP, RC, NP, and SB performed RNA-seq analysis and analyzed data. LMa performed CLL cells biological characterization. RH provided CLL samples and clinical information. FB and MM-B designed and supervised the mathematical/statistical study and analyzed data. LV designed and supervised the biological study and analyzed data. LV, FB, CS, MT, AN, LMi, and J-NF wrote the manuscript. All authors read and amended the manuscript and concurred with the manuscript and its contents.

Funding This work was supported by grants from the Institut Thématique Multi-Organism cancer initiative within the framework plan cancer 2009-2013 "GenPred project" to LV and MM-B, Initiative d'Excellence-CNRS to LV and FB, the Association pour la Recherche contre le Cancer (ARC), Alsace Cancer association, Ligue Contre le Cancer (Bas-Rhin), and Cancéropole Grand-Est to LV. The Strasbourg School of Medicine Next-Generation Sequencing Center, the

Institut Universitaire de France (IUF), the Ligue contre le Cancer, the LABEX Transplantex [ANR-11-LABX-0070_Transplantex] (French National Research Agency; ANR), and Inserm UMR S-1109 to SB and the CNRS UMR 7501, the LABEX IRMIA [ANR-11-LABX-0055] to FB and MM-B. This work was also supported by the French Proteomic Infrastructure (ProFI ANR-10-INBS-08-03).

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Ethics approval All subjects gave written informed consent for this study, which was approved by the institutional review board of the Strasbourg University Hospitals "Comité de protection des personnes Est IV" on July 6, 2017 (approval file number #09/40).

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

References

1. Shaffer AL 3rd, Young RM, Staudt LM. Pathogenesis of human B cell lymphomas. Annu Rev Immunol. 2012;30:565–610.
2. Niiro H, Clark EA. Regulation of B-cell fate by antigen-receptor signals. Nat Rev Immunol. 2002;2:945–56.
3. Messmer BT, Albesiano E, Efremov DG, Ghiotto F, Allen SL, Kolitz J, et al. Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. J Exp Med. 2004;200:519–25.
4. Caligaris-Cappio F. Role of the microenvironment in chronic lymphocytic leukaemia. Br J Haematol. 2003;123:380–8.
5. Herishanu Y, Perez-Galan P, Liu D, Biancotto A, Pittaluga S, Vire B, et al. The lymph node microenvironment promotes B-cell receptor signaling, NF-kappaB activation, and tumor proliferation in chronic lymphocytic leukemia. Blood. 2011;117:563–74.
6. Gobessi S, Laurenti L, Longo PG, Sica S, Leone G, Efremov DG. ZAP-70 enhances B-cell-receptor signaling despite absent or inefficient tyrosine kinase activation in chronic lymphocytic leukemia and lymphoma B cells. Blood. 2007;109:2032–9.
7. Stevenson FK, Krysov S, Davies AJ, Steele AJ, Packham G. B-cell receptor signaling in chronic lymphocytic leukemia. Blood. 2011;118:4313–20.
8. Guarini A, Chiaretti S, Tavolaro S, Maggio R, Peragine N, Citarella F, et al. BCR ligation induced by IgM stimulation results in gene expression and functional changes only in IgV H unmutated chronic lymphocytic leukemia (CLL) cells. Blood. 2008;112:782–92.
9. Pede V, Rombout A, Vermeire J, Naessens E, Mestdagh P, Robberecht N, et al. CLL cells respond to B-cell receptor stimulation with a microRNA/mRNA signature associated with MYC activation and cell cycle progression. PLoS ONE. 2013;8:e60275.
10. Vallat LD, Park Y, Li C, Gribben JG. Temporal genetic program following B-cell receptor cross-linking: altered balance between proliferation and death in healthy and malignant B cells. Blood. 2007;109:3989–97.
11. Yoshida T, Higuchi T, Hagiyama H, Strasser A, Nishioka K, Tsubata T. Rapid B cell apoptosis induced by antigen receptor ligation does not require Fas (CD95/APO-1), the adaptor protein FADD/MORT1 or CrmA-sensitive caspases but is defective in both MRL-+/+ and MRL-lpr/lpr mice. Int Immunol. 2000;12:517–26.
12. Schleiss C, Ilias W, Tahar O, Guler Y, Miguet L, Mayeur-Rousse C, et al. BCR-associated factors driving chronic lymphocytic leukemia cells proliferation ex vivo. Sci Rep. 2019;9:701.
13. Bertrand F, Aouadi I, Jung N, Carapito R, Vallat L, Bahram S, et al. SelectBoost: a general algorithm to enhance the performance of variable selection methods. Bioinformatics. 2020. https://doi.org/10.1093/bioinformatics/btaa855 [Epub ahead of print].
14. Jung N, Bertrand F, Bahram S, Vallat L, Maumy-Bertrand M. Cascade: a R package to study, predict and simulate the diffusion of a signal through a temporal gene network. Bioinformatics. 2014;30:571–3.
15. Vallat L, Kemper CA, Jung N, Maumy-Bertrand M, Bertrand F, Meyer N, et al. Reverse-engineering the genetic circuitry of a cancer cell with predicted intervention in chronic lymphocytic leukemia. Proc Natl Acad Sci USA. 2013;110:459–64.
16. Bertrand F, Maumy-Bertrand M. Patterns: a modeling tool dedicated to biological network modeling. R package. 2019. https://cran.r-project.org/web/packages/Patterns/index.html.
17. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–9.
18. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012;40:4288–97.
19. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteom. 2014;13:2513–26.
20. Rau A, Gallopin M, Celeux G, Jaffrezic F. Data-based filtering for replicated high-throughput transcriptome sequencing experiments. Bioinformatics. 2013;29:2146–52.
21. Rau A, Maugis-Rabusseau C, Martin-Magniette ML, Celeux G. Co-expression analysis of high-throughput transcriptome sequencing data with Poisson mixture models. Bioinformatics. 2015;31:1420–7.
22. Goeminne LJ, Gevaert K, Clement L. Peptide-level robust ridge regression improves estimation, sensitivity, and specificity in data-dependent quantitative label-free shotgun proteomics. Mol Cell Proteom. 2016;15:657–68.
23. Hao S, Baltimore D. The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. Nat Immunol. 2009;10:281–8.
24. Yosef N, Regev A. Impulse control: temporal dynamics in gene transcription. Cell. 2011;144:886–96.
25. Meinshausen N, Bühlmann P. Stability selection. J R Stat Soc: Ser B. 2010;72:417–73.
26. Ideker T, Dutkowski J, Hood L. Boosting signal-to-noise in complex biology: prior knowledge is power. Cell. 2011;144:860–3.
27. Liu ZP, Wu C, Miao H, Wu H. RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. Database. 2015;2015:1–12.

28. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer; 2009. p. 745.
29. Fournier ML, Paulson A, Pavelka N, Mosley AL, Gaudenz K, Bradford WD, et al. Delayed correlation of mRNA and protein expression in rapamycin-treated cells and a role for Ggc1 in cellular sensitivity to rapamycin. Mol Cell Proteom. 2010;9:271–84.
30. Jayapal KP, Philp RJ, Kok YJ, Yap MG, Sherman DH, Griffin TJ, et al. Uncovering genes with divergent mRNA-protein dynamics in *Streptomyces coelicolor*. PLoS ONE. 2008;3:e2097.
31. Waldbauer JR, Rodrigue S, Coleman ML, Chisholm SW. Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. PLoS ONE. 2012;7:e43432.
32. Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature. 2011;473:337–42.
33. Waters KM, Liu T, Quesenberry RD, Willse AR, Bandyopadhyay S, Kathmann LE, et al. Network analysis of epidermal growth factor signaling using integrated genomic, proteomic and phosphorylation data. PLoS ONE. 2012;7:e34515.
34. Steelman LS, Franklin RA, Abrams SL, Chappell W, Kempf CR, Basecke J, et al. Roles of the Ras/Raf/MEK/ERK pathway in leukemia therapy. Leukemia. 2011;25:1080–94.
35. Gururajan M, Chui R, Karuppannan AK, Ke J, Jennings CD, Bondada S. c-Jun N-terminal kinase (JNK) is required for survival and proliferation of B-lymphoma cells. Blood. 2005;106:1382–91.
36. Bargou RC, Emmerich F, Krappmann D, Bommert K, Mapara MY, Arnold W, et al. Constitutive nuclear factor-kappaB-RelA activation is required for proliferation and survival of Hodgkin's disease tumor cells. J Clin Investig. 1997;100:2961–9.
37. Kawagoe T, Takeuchi O, Takabatake Y, Kato H, Isaka Y, Tsujimura T, et al. TANK is a negative regulator of Toll-like receptor signaling and is critical for the prevention of autoimmune nephritis. Nat Immunol. 2009;10:965–72.
38. Pomerantz JL, Baltimore D. NF-kappaB activation by a signaling complex containing TRAF2, TANK and TBK1, a novel IKK-related kinase. EMBO J. 1999;18:6694–704.
39. Voena C, Chiarle R. RHO family GTPases in the biology of lymphoma. Cells. 2019;8:646.
40. Gouw LG, Reading NS, Jenson SD, Lim MS, Elenitoba-Johnson KS. Expression of the Rho-family GTPase gene RHOF in lymphocyte subsets and malignant lymphomas. Br J Haematol. 2005;129:531–3.
41. Shaverdashvili K, Padlo J, Weinblatt D, Jia Y, Jiang W, Rao D, et al. KLF4 activates NFkappaB signaling and esophageal epithelial inflammation via the Rho-related GTP-binding protein RHOF. PLoS ONE. 2019;14:e0215746.
42. Baietti MF, Zhang Z, Mortier E, Melchior A, Degeest G, Geeraerts A, et al. Syndecan-syntenin-ALIX regulates the biogenesis of exosomes. Nat Cell Biol. 2012;14:677–85.
43. Menezes ME, Shen XN, Das SK, Emdad L, Sarkar D, Fisher PB. MDA-9/syntenin (SDCBP) modulates small GTPases RhoA and Cdc42 via transforming growth factor beta1 to enhance epithelial-mesenchymal transition in breast cancer. Oncotarget. 2016;7:80175–89.
44. Boukerche H, Su ZZ, Emdad L, Sarkar D, Fisher PB. mda-9/Syntenin regulates the metastatic phenotype in human melanoma cells by activating nuclear factor-kappaB. Cancer Res. 2007;67:1812–22.
45. Wei J, Guo W, Lian H, Yang Q, Lin H, Li S, et al. SNX8 mediates IFNgamma-triggered noncanonical signaling pathway and host defense against *Listeria monocytogenes*. Proc Natl Acad Sci USA. 2017;114:13000–5.
46. Moffett JR, Namboodiri MA. Tryptophan and the immune response. Immunol Cell Biol. 2003;81:247–65.
47. Bod L, Douguet L, Auffray C, Lengagne R, Bekkat F, Rondeau E, et al. IL-4-induced gene 1: a negative immune checkpoint controlling B cell differentiation and activation. J Immunol. 2018;200:1027–38.
48. Boulland ML, Marquet J, Molinier-Frenkel V, Moller P, Guiter C, Lasoudris F, et al. Human IL4I1 is a secreted L-phenylalanine oxidase expressed by mature dendritic cells that inhibits T-lymphocyte proliferation. Blood. 2007;110:220–7.
49. Pascutti MF, Jak M, Tromp JM, Derks IA, Remmerswaal EB, Thijssen R, et al. IL-21 and CD40L signals from autologous T cells can induce antigen-independent proliferation of CLL cells. Blood. 2013;122:3010–9.
50. Fritsch K, Finke J, Grullich C. Suppression of granzyme B activity and caspase-3 activation in leukaemia cells constitutively expressing the protease inhibitor 9. Ann Hematol. 2013;92:1603–9.
51. Bladergroen BA, Meijer CJ, ten Berge RL, Hack CE, Muris JJ, Dukers DF, et al. Expression of the granzyme B inhibitor, protease inhibitor 9, by tumor cells in patients with non-Hodgkin and Hodgkin lymphoma: a novel protective mechanism for tumor cells to circumvent the immune system? Blood. 2002;99:232–7.
52. Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood. 1999;94:1840–7.
53. Letestu R, Rawstron A, Ghia P, Villamor N, Boeckx N, Boettcher S, et al. Evaluation of ZAP-70 expression by flow cytometry in chronic lymphocytic leukemia: a multicentric international harmonization process. Cytom B Clin Cytom. 2006;70:309–14.

Supplemental methods (Schleiss et al.)

Source of reagents:

Product	Catalogue number	Supplier
CellTrace CFSE	C34554	Invitrogen (Thermofisher)
Unconjugated Goat F(AB')2 Anti Human IgM FC	109-006-129	Jackson ImmunoResearch
MegaCD40L Protein	ALX-522-110-C010	Enzo Life Sciences
Recombinant Human IL-4 Protein	204-IL-010	R&D systems
IL21 Recombinant Human Protein	10447063	Gibco (Fischer Scientific)
CD19-PC7 (J3-119)	IM3628	Beckman Coulter
Mouse IgG1 -PC7 isotype control	737662	Beckman Coulter
CD5 FITC (BL1a)	A08932	Beckman Coulter
IgG2a-FITC isotype control	A12689	Beckman Coulter
Annexin V FITC	556420	BD Biosciences
propidium iodide	556463	BD Biosciences

mRNA samples preparation

Before BCR-mediated cell activation (T0), and at eight time points after activation (1h, 1h30, 3h30, 6h30, 12h 24h, 48h, 96h), 4.10^6 cells were collected, washed and resuspended in 1mL TRIzol (Sigma-Aldrich). Total RNA was isolated using chloroform and Phase Lock Gel tube (5 Prime) and was purified using the RNeasy Mini kit (Qiagen) according to the manufacturer's recommendation. RNA integrity was assessed with the Agilent total RNA Pico Kit on a 2100 Bioanalyzer instrument (Agilent Technologies). Ribosomal RNA was depleted with the Low Input RiboMinus™ Eukaryote System v2 kit (Thermo Fisher Scientific) following manufacturer's instructions.

RNAseq analysis and data preprocessing

The sequencing library was prepared with the Ion Total RNA-seq kit v2 (Thermo Fisher Scientific) according to the manufacturer's instructions. The libraries were loaded two by two at a concentration of 20 pM on an Ion PI Chip using the Ion Chef Instrument (Thermo Fisher Scientific). The sequencing was performed on an Ion Proton sequencer with the Ion PI Hi-Q Sequencing 200 Kit (Thermo Fisher Scientific). Raw sequencing data were preprocessed according to the manufacturer's recommendations. Adapter sequences were removed and reads shorter than 18bp were excluded. Reads were then mapped against the human genome (Hg19) and a virtual reference transcriptome using TopHat2 [1]. Unmapped reads were mapped to the reference Hg19 using Bowtie2 in local mode [2].

The total reads (SAM format) was used to count reads with the default union-counting mode of the tool htseq-count of the Python package HTSeq [3] (GEO accession GSE130385). We used the edgeR package [4], to derive the reads per kilobase per million (rpkm) values from read counts and gene lengths. These rpkm values were used for differential analysis study and their Voom transformed values for model inference [5]. Small Nucleolar RNA (snor, 370 entries) and the microRNAs (MIR*, 1 651 entries) were filtered out.

Proteomic samples preparation

Before cell stimulation (T0), and at eight time points after stimulation (1h, 2h, 4h, 7h, 12h, 24h, 48h, 96h), 8.10^6 cells were collected, washed and resuspended in 100µl of lysis buffer (Urea 8M, Thiourea 2M, Chaps 4%, Dithiothreitol (DTT) 1%, Triton 10%, TLCK 0,05%, protease inhibitors). After centrifugation (5min, 8000g), 6 volumes of glacial acetone were added to the supernatant and samples were incubated overnight at -20°C. The proteins were pelleted and 10 µg of each sample were resuspended in loading buffer (2% SDS, 0,1M DTT, 10% glycerol, 62.5 mM Tris pH 6.8) and concentrated in one stacking-gel band using a 4% SDS-PAGE gel. The gels were fixed with 45% methanol/3% acetic acid and stained with colloidal Silver Blue. Each band was excised and cut in four pieces prior to in-gel digestion. The gel pieces were washed four times with 100 µL of 75% acetonitrile (ACN) and 25% ammonium bicarbonate buffer (NH4HCO3) at 25 mM and dehydrated with 50 µL of ACN. The cysteine residues were reduced by adding 10 mM DTT for 30 min at 60°C and 30 min at room temperature, and alkylated by adding 55 mM iodoacetamide (IAA) for 20 min in the dark. The bands were then washed three times by adding 50 µL of 25 mM NH4HCO3 and 50 µL of ACN. Gel pieces were dehydrated twice with 50 µL ACN prior to enzymatic digestion. Proteins were cleaved in an adequate volume to cover the gel pieces with a modified porcine trypsin (Promega) solution at a 1:50 w/w enzyme:protein ratio. Digestion was performed overnight at 37°C. Tryptic peptides were extracted twice under agitation, first with 40 µL of 60% ACN in 0.1% formic acid (FA) for 1 h and then with 30 µL of 100% ACN for 40 min. The collected extracts were pooled, the excess of ACN was vacuum dried, and the samples were resolubilized with 25 µL of H2O/ACN/FA (98/2/0.1 v/v/v) and sonicated in ice during 10 min prior to nanoLC-MS/MS analysis.

Mass spectrometry analysis and data preprocessing

NanoLC-MS/MS analyses were performed on a nanoAcquity UPLC device (Waters Corporation) coupled to a Q-Exactive Plus mass spectrometer (Thermo Scientific). Peptide separation was

performed on an ACQUITY UPLC BEH130 C18 column (250 mm × 75 µm with 1.7 µm diameter particles) and a Symmetry C18 precolumn (20 mm × 180 µm with 5 µm diameter particles, Waters). The solvent system consisted of 0.1% FA in water (solvent A) and 0.1% FA in ACN (solvent B). The samples were loaded into the enrichment column over 3 min at 5 µL/min with 99% of solvent A and 1% of solvent B. Peptides were eluted at 450 nL/min with the following gradient of solvent B: from 1 to 35% over 120 min and 35 to 90% over 1 min. The MS capillary voltage was set to 1.8 kV at 250°C. The system was operated in Data-Dependent Acquisition (DDA) mode with automatic switching between MS (50 ms/scan over a 300–1800 m/z range with R = 70 000) and MS/MS (100 ms/scan over a 200–2000 m/z range with R = 17 500) modes. The ten most abundant ions (intensity threshold 2 x 105) were selected on each MS spectrum for further isolation and higher energy collision dissociation (HCD) fragmentation, excluding unassigned and monocharged ions. The dynamic exclusion time was set to 60 s. Raw data collected during nanoLC-MS/MS were processed using MaxQuant (version 1.5.5.1) [6]. Peaks were assigned with the Andromeda search engine with trypsin/P specificity. The database used for the searches contained all human entries extracted from UniProtKB-SwissProt including canonical sequences and isoforms (2016, sept. ; 42 145 entries). The minimum peptide length required was seven amino acids and a maximum of one trypsin missed cleavage was allowed. The precursor mass tolerance was set to 20 ppm for the first search and 4.5 ppm for the main search. The fragment ion mass tolerance was set to 20 ppm. Methionine oxidation was set as a variable modification and carbamidomethylation of cysteines as a fixed modification. The maximum false discovery rate was 1% at peptide and protein levels with the use of a decoy strategy. The “match between runs” option was used. The dataset was made of 53 measurements of 50 503 peptides (6 subjects x 9 time points, outlying sample CLL-NP3_24h was excluded) and we used the data preprocessing pipeline recommended by the MSqRob package [7]. The data were log2 transformed and we applied quantile normalization. We filtered peptides that could be assessed as contaminant and we removed proteins that were only identified by peptides carrying one or more modification sites. Gene symbols were retrieved for all protein groups, whenever possible, from the UniProt database (The UniProt Consortium 2017 [8], using the proteins API [9]. The complete proteomics dataset was deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD013573.

Gene expression Analysis

Selection of differentially expressed (DE) genes. The TREAT test conducts genewise statistical tests for a given coefficient or contrast relative to a specified fold-change threshold. First, separate group analysis was performed to search for genes significantly DE, with glmLRT or glmTreat, at the 5% or 1%

FDR level at a given time (Ti, i>0) versus control (T0=0) in each of the proliferative (P) and non-proliferative (NP) groups (temporal signature). Second, we searched for DE genes between each groups (response signature). glmLRT and glmTreat (whenever possible) detect genes that respond differently in the NP and P groups at a given time or over all time points (with respect to control time values).

Cluster selections of genes. We performed a single run of HTSCluster looking up to K = 1, …, 35 clusters, using the Trimmed Means of M-values (TMM) normalization [10], and the splitting small-EM strategy, an approach similar to that described by Papastamoulis [11]. In HTSCluster, model selection may be performed using the DDSE calibration for slope heuristics, Djump calibration for slope heuristics, Bayesian Information Criterion, and Integrated Completed Likelihood criterion. When the slope heuristics approach may be applied, we used the capushe package [12], to provide diagnostic plots for this slope heuristics in order to ensure that sufficiently complex models have been considered.

Proteomic Data Analysis

Choosing the FDR procedure. According to calibration plots, p-values distributions were not standard and hence the use of the cp4p package [13] to perform more involved FDR correction was required. Almost all the choices of the FDR control method provided by the cp4p package led to the same numbers of DA proteins. We selected the two-stage Benjamini and Hochberg procedure [14] because it can limit the FDR to a given level and was set to 0.05 for the DA analysis. The differential analysis was made using the test.contrast_adjust function of the MSqRob package [7]. MSqRob implements a peptide-level robust ridge regression that improves estimation, sensitivity, and specificity in data-dependent quantitative label-free shotgun proteomics.

Gene and protein selection for network reverse engineering

To merge the gene and protein datasets, we matched the gene and protein groups names. Protein groups that featured multiple gene names were manually curated. The selection of the actors of the networks was made in three steps:

- Base selection = Treat 1% FDR DE genes for any of the 24 tests and 1% FDR DA protein groups for any of the 24 tests. Those 24 tests are the 8 Ti vs T0 tests for each of the two groups P, NP and the between group NP v P analysis.
- Enrichment 1 = NP v P 5% DE genes or DA protein groups for any of the 24 tests (without FDR correction for the 24 tests)

- Enrichment 2 = Based on temporal profiles expected after an exogeneous cell activation.

More precisely, a gene could be selected by the profile based enrichment if Its DE ranking is less than 200 and it features a single peak at a single time, or a single peak for two consecutive times, or a single peak for three consecutive times.

The DE ranking of gene, is its ranking among the other genes with respect to its p-value at a given test. A gene with a DE ranking equal to 1 features the smallest p-value for the test whereas a gene with the highest DE ranking possible, features the largest p-value for the test. We chose 200, a rather high value, as the max DE ranking value in order to perform a large profile based enrichment of the selection. The number of unique gene ID in the final selection is 5 733. Obviously, genes expressions or proteins abundancies must not be constant (within a given group either NP or P) to be included in the reverse engineering algorithm since we need to assess their predictive value on other gene expressions or protein abundancies. As a consequence, we had to filter out 125 protein abundancies. Any of the proteins abundancies that were removed were constant in both of the groups.

Clustering of genes and proteins

We first adjusted the range of the gene expressions and of the protein abundancies in order to improve the similarities between the genes and proteins groups values. To rescale abundancies' distributions to the ranges of the expressions' distributions, we used a multiplying factor: the 50% Trimmed Mean of the ratios of ranges of the abundancies divided by the range of the expressions for every Gene Name for which those two values were available.

To cluster genes and protein we applied Fuzzy c−means clustering (FCM) 15, 16 using the Euclidean distance.

1. Methodology of the clustering.

We used two runs of the FCM methodology, detailed in section 2 below, to get the 62 final clusters of the genes (41 clusters) and proteins (21 clusters) datasets.

First, using FCM, we clustered the subset of genes and proteins, for which both gene expressions and proteins abundances were measured, according to those genes expressions and those protein abundancies, which resulted in 21 clusters. We directly used those 21 clusters to cluster the proteins, since gene expressions were missing for only a few proteins.

Yet, there were hundreds of genes expressions without protein abundance data. To tackle this issue, we performed an additional FCM run and clustered all the genes expressions into 20 clusters using only the gene expression data.

For the genes that may be clustered in the first run (genes and proteins data) and in the second run (genes without proteins data), we retained the clustered found using both gene and protein data since it is likely to be more accurate. Hence, the genes were split into 21+20 clusters and the proteins into the 21 very same clusters as the 21 genes clusters. In addition, this type of clustering was likely to help to simplify the upcoming network inference since it required to model how genes and proteins clusters interact.

2. Methodology for an FCM.

Even if Mfuzz package, that implements FCM, was first developed to cluster microarray data, the algorithms of the Mfuzz package, as claimed by its authors on their webpage (http://mfuzz.sysbiolab.eu), can also be applied to other types of data such as proteins abundances or RNAseq datasets.

We chose to perform the FCM cluster analysis using the Euclidean distance. As result, the expression values of genes were standardized to have a mean value of zero and a standard deviation of one. This ensures that vectors of genes with similar changes in expression are close in Euclidean space, as recommended by Futschik and colleagues 15. To perform FCM, two parameters have to be specified: the number of clusters c and the FCM parameter m.

The FCM parameter m was estimated using the m estimate function of the Mfuzz package that implements the algorithm from Schwammle et al. 17. To determine the number of clusters, we used the minimum centroid distance as a cluster validity index since for an optimal cluster number, we may see a 'drop' of minimum centroid distance when plotted versus a range of cluster number and a slower decrease of the minimum centroid distance for higher cluster number. Again, more information and some examples can be found in the study of Schwammle and Jensen (2010).

Network reverse engineering

Then we designed groups for both the NP and P network inferences in the following way:

1. For the 3 707 gene expression data of the gene ID with gene expression data only, we used the previously derived 20 clusters, numbered from 1 to 20. The minimum size of a group is 68.

2. For the 2 015 gene expression data of the gene ID with both gene expression and protein abundancies data, we used the previously derived 21 clusters, numbered from 21 to 41. The minimum size of a group is 44.

3. For the 2 025 protein abundancies data of the gene ID with protein abundancies data (and for 2,015 of them gene expression data), we used the previously derived 21 clusters, numbered from 42 to 62.

We had to infer, using random forest imputation and the mice R package [18] the group membership values for the 10 proteins that only featured protein abundancies. The minimum size of a group is 44. We will denote by G2P the action of a cluster of genes expressions on a cluster of protein abundancies, G2G the action of a cluster of genes expressions on a cluster of genes expressions, P2G the action of a cluster of protein abundancies on a cluster of genes expressions and P2P the action of a cluster of protein abundancies on a cluster of protein abundancies (Supplementary Fig. 4A).

For any given cluster i among those 62 clusters, we set, for all of them and by analyzing the time course profiles of the cluster members, the first timepoint, denoted by O_i, for which it might begin to have an effect on the other actors of the networks.

In order to take into account relevant biological knowledge, we designed a weighted inference. The algorithm aims to reveal the links between the actors of the network (genes or proteins). A weight can favor or disfavor such a link. It ranges from 0, always in, to $+\infty$, always exclude; a unit weight being neutral.

Information on regulators, targets, as well as confidence (High, Low, Medium) and evidence (Experimental, Predicted), are provided by RegNetwork (341 207 links) [19]. The weight values were modulated according to the actual confidence and evidence values. We chose several thresholds for the weights and matched them with the uncertainty of the biological knowledge of a given action of an actor of the network on another:

always=0<+++<++<+<1<-<+Inf=exclude

The weights also take several biological or modelling assumptions into account:

- If both gene expressions and protein abundancies were measured for a given gene ID, only the protein abundancy measurement must be used to infer its effect in the network.
- No actions within a cluster. The members of a given cluster (of genes or proteins) cannot be used to infer the expressions of the abundancies of other members of the same cluster. It makes since members of the same cluster share similar time profiles.
- No loop action (either for a gene on itself or for a protein on itself)

Such a weight is view as a penalty factor in the penalized regression model: it is a number that multiplies the lambda value in the minimization problem to allow differential shrinkage ([20], equation 1 p.3). If equal to 0, it implies no shrinkage, and that variable is always included in the model. Default is 1 for all variables. Infinity means that the variable is excluded from the model. Note that the weights are rescaled to sum to the number of variables.

The core of the statistical model combined a F matrix (a square matrix of size 8*62=496) that models the time dependent effects between clusters and a ω matrix (a square matrix of size 7 747) that captures a non-time dependent link (the possibility of an action) between the actors of the network.

More precisely, a cell F_{ij} (itself a square matrix of size T) of the F matrix models the effect of cluster i on cluster j. Notice that if an actor of the network n_0 belongs to cluster i and an actor of the network n belongs to cluster j, the matrix F_{ij} allows the link between the actors n_0 and n to be time dependent. To enforce temporal causality, we used the two following time constraints:

- A cluster i cannot affect another cluster j, if $C_i \geq C_j$: this ensures that a cluster i cannot affect a cluster j if the first action time C_j of cluster j precedes the first action time of cluster i. Hence, if $C_i \geq C_j$, then the F_{ij}cell of the F matrices is set to 0 (F_{ij}=0). Supplementary Fig. 4 B-C shows nonzero F_{ij} matrices for the P and NP network reverse engineering. These are the same for these data but could have happened to be different.
- If $C_i < C_j$, then the F_{ij} matrix is a lower triangular square matrix of size T. Its shape is chosen so that the measurement of an actor of the network at time t_k can influence another the measurement of an actor of the network at time t_{k_0} if and only if $k < k_0$ (Supplementary Fig. 4C). The only exception being the G2P case of the action of a cluster of genes i for which we also measured the proteins ($21 \leq i \leq 41$) on its protein cluster (the cluster of the proteins abundancies with the same Gene ID numbered $i + 21$). In that case we choose an almost diagonal matrix (Supplementary Fig. 4B). Been able to easily switch between those matrices is one of the main reasons that accounts for creating the same clusters for both genes and proteins when possible.

Hence, contrary to what was previously assumed in Vallat et al. 2013 and Jung et al. 2014 [21, 22], the size of the F_{ij} was increased in order to cope with the G2P actions that are modelled at the same timepoint and the sub-diagonals and the diagonal of matrices F were no longer supposed to be constant since the experimental timepoints were no longer equally distributed throughout the experiment (the times differences range from 30 minutes to 2 days). A constant value on a subdiagonal of the F_{ij} matrix would mean that the intensity of the effect of an actor on another one only depends on the time lag between the measurements. Nevertheless, we assumed that for similar time steps, the interactions should depend only on time index differences (i.e. time lag) rather than absolute time index. Hence some parts of the diagonal and sub-diagonal of the F_{ij} matrices are constant.

We have selected $N = 7\,747$ actors for the reverse engineering across $T = 8$ timepoints and for $P = 3$ individuals; we denote by x_{npt} observed value (gene expression of protein abundancy) of the actor n for an individual p at timepoint t. For any actor of the network n among the $N = 7\,747$ ones, the mathematical model was written:

$$\widetilde{\boldsymbol{x}}_{\boldsymbol{np}\cdot} = \sum_{n'=1}^{N} \omega_{n\prime n} \boldsymbol{F}_{\boldsymbol{m}(\boldsymbol{n}')\boldsymbol{m}(\boldsymbol{n})} \widetilde{\boldsymbol{x}}_{\boldsymbol{n}'\boldsymbol{p}\cdot} + \boldsymbol{\varepsilon}_{\boldsymbol{np}}, \qquad 1 \leq p \leq P.$$

In this model, N is the total number of actors, $k \mapsto m(k)$ is the function that maps an actor to its time-cluster, $\boldsymbol{F}_{\boldsymbol{ij}}$ is a T square matrix that describes the action of the actors belonging to cluster i on

an actor that belongs to cluster j, ω_{kl} is the strength of the connection from actor k towards actor l and $\boldsymbol{\varepsilon_{np}}$, $1 \leq p \leq P$ is a T dimensional random vector with zero mean and unit variance $\boldsymbol{I_T}$. So, in this model $\tilde{\boldsymbol{x}}_{\boldsymbol{n}..}$ is the regulated actor and $\tilde{x}_{n_0..}$, $1 \leq n_0 \neq n \leq 7\,747$, are the regulators ($n_0 \neq n$ to not allow self-regulation). It is known that genes or proteins tend to be regulated by few genes or proteins whereas a given gene or protein can regulate many other genes or proteins. Hence we chose to carry out the fitting of the model using penalized regression. In addition, in order to select only the more relevant and stable regulators for given actor, we used a weighted variant of stability selection [23] combined with nonnegative least squares and a coordinate ascent approach by alternatively supposing the $\boldsymbol{F_{ij}}$ matrices known or the $\boldsymbol{\omega}$ matrix known. To get a more robust result, the estimation of the $\widehat{\boldsymbol{F}}_{\boldsymbol{ij}}$ matrices was done several times by 'leave one subject out' cross-validation. Furthermore, to avoid computational issues, after each step of the algorithm the new $\boldsymbol{F_{ij}}$ matrices were chosen as a linear combination between the $\boldsymbol{F_{ij}}$ matrices estimated at the end of the preceding step and the $\boldsymbol{F_{ij}}$ matrices that were estimated at this step using non negative least squares. The result of the estimation process was threefold: (1) a connectivity network described by the nonzero elements of $\widehat{\boldsymbol{\omega}}$, if $\widehat{\omega}_{n'n} \neq 0$ means that an action of n' on n was detected, (2) any $\widehat{\boldsymbol{F}}_{\boldsymbol{ij}}$ matrix catches if there is an effect of cluster i on cluster j and at which time(s) those effects arise, and (3) the evolution through time of the action of actor n' on actor n can be derived by computing the product $\widehat{\omega}_{n'n}\widehat{\boldsymbol{F}}_{\boldsymbol{m(n')m(n)}}$ and by combining the $\boldsymbol{\omega}$ and $\boldsymbol{F_{ij}}$ (code available in the Patterns R package [24]).

Data simulation and algorithm validation

To simulate the measurements of the actors based on a regulatory network, we designed an algorithm that is inspired by the preferential attachment [25]. Then, we adapted it to temporal nested networks. We then used our cascade network based model to make some simulations, using Laplace laws to set the values of the measurements of the actors at the first time point.

The results of simulated data were then compared with the performances of our previous Cascade algorithm (Cascade package [21, 22]), its non-weighted, properly weighted and incorrectly weighted Lasso-versions (Patterns package [24]) and with the performances of a non-weighted and weighted Stability Selection-version algorithm [26] (Supplementary Fig. 4 D-F). We also compared these results with the performances of a weighted and non-weighted version of the SelectBoost algorithm (SelectBoost Package, results available online on the package website [27]). These results showed that correctly weighted models achieved a higher sensibility, PPV and F-score than the other ones (especially Cascade) and demonstrated the performances of this modelling approach.

References supplemental methods

1. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013 Apr 25; **14**(4)**:** R36.

2. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012 Mar 4; **9**(4)**:** 357-359.

3. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015 Jan 15; **31**(2)**:** 166-169.

4. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 2012 May; **40**(10)**:** 4288-4297.

5. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014 Feb 3; **15**(2)**:** R29.

6. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* 2014 Sep; **13**(9)**:** 2513-2526.

7. Goeminne LJ, Gevaert K, Clement L. Peptide-level Robust Ridge Regression Improves Estimation, Sensitivity, and Specificity in Data-dependent Quantitative Label-free Shotgun Proteomics. *Mol Cell Proteomics* 2016 Feb; **15**(2)**:** 657-668.

8. Chen C, Huang H, Wu CH. Protein Bioinformatics Databases and Resources. *Methods Mol Biol* 2017; **1558:** 3-39.

9. Nightingale A, Antunes R, Alpi E, Bursteinas B, Gonzales L, Liu W*, et al.* The Proteins API: accessing key integrated protein and genome information. *Nucleic Acids Res* 2017 Jul 3; **45**(W1)**:** W539-W544.

10. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 2010; **11**(3)**:** R25.

11. Papastamoulis P, Martin-Magniette ML, Maugis-Rabusseau C. On the estimation of mixtures of Poisson regression models with large number of components. *Computational Statistics & Data Analysis* 2016; **93:** 97-106.

12. Arlot S, Brault V, Baudry JP, Maugis C, Michel B. capushe: CAlibrating Penalities Using Slope HEuristics. version 1.1.1. *R package* 2016.

13. Giai Gianetto Q, Combes F, Ramus C, Bruley C, Coute Y, Burger T. Calibration plot for proteomics: A graphical tool to visually check the assumptions underlying FDR control in quantitative experiments. *Proteomics* 2016 Jan; **16**(1)**:** 29-32.

14. Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false dicovery rate. *Biometrika* 2006; **93**(3)**:** 491-507.

15. Futschik ME, Carlisle B. Noise-robust soft clustering of gene expression time-course data. *J Bioinform Comput Biol* 2005 Aug; **3**(4)**:** 965-988.

16. Kumar L, M EF. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* 2007 May 20; **2**(1)**:** 5-7.

17. Schwammle V, Jensen ON. A simple and fast method to determine the parameters for fuzzy c-means cluster analysis. *Bioinformatics* 2010 Nov 15; **26**(22)**:** 2841-2848.

18. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 2011; **45**(3)**:** 1-67.

19. Liu ZP, Wu C, Miao H, Wu H. RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database (Oxford)* 2015; **2015**.

20. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software* 2010; **33**(1)**:** 1.

21. Vallat L, Kemper CA, Jung N, Maumy-Bertrand M, Bertrand F, Meyer N*, et al.* Reverse-engineering the genetic circuitry of a cancer cell with predicted intervention in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences of the United States of America* 2013 Jan 8; **110**(2)**:** 459-464.

22. Jung N, Bertrand F, Bahram S, Vallat L, Maumy-Bertrand M. Cascade: a R package to study, predict and simulate the diffusion of a signal through a temporal gene network. *Bioinformatics* 2014 Feb 15; **30**(4)**:** 571-573.

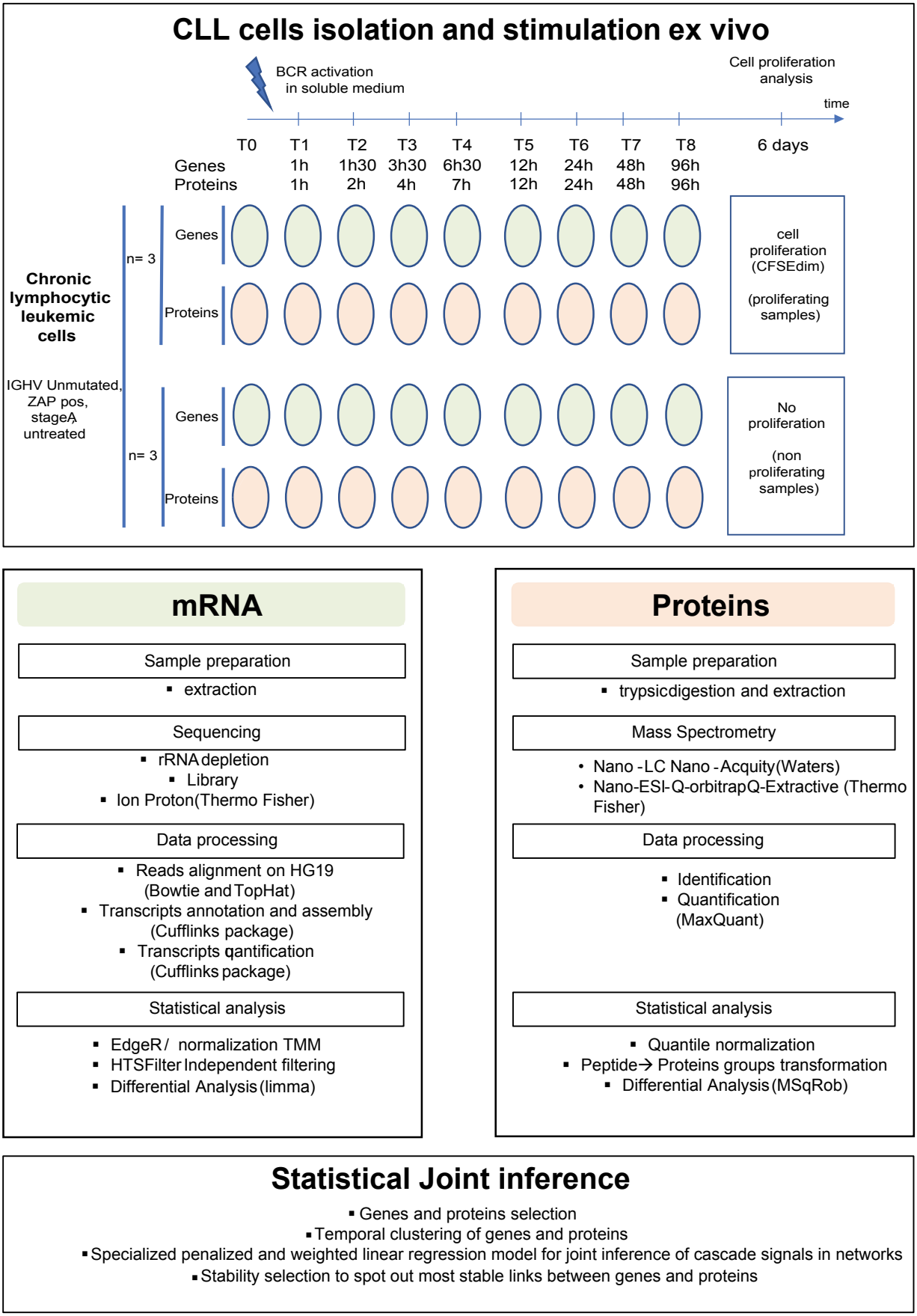
23. Meinshausen N, Bühlmann P. Stability selection. *Journal of the Royal Statistical Society: series B (Statistical methodology)* 2010; **72**(4)**:** 417-473.

24. Bertrand F, Maumy-Bertrand M. Patterns: patterned networks reverse engineering. *R package* 2019.

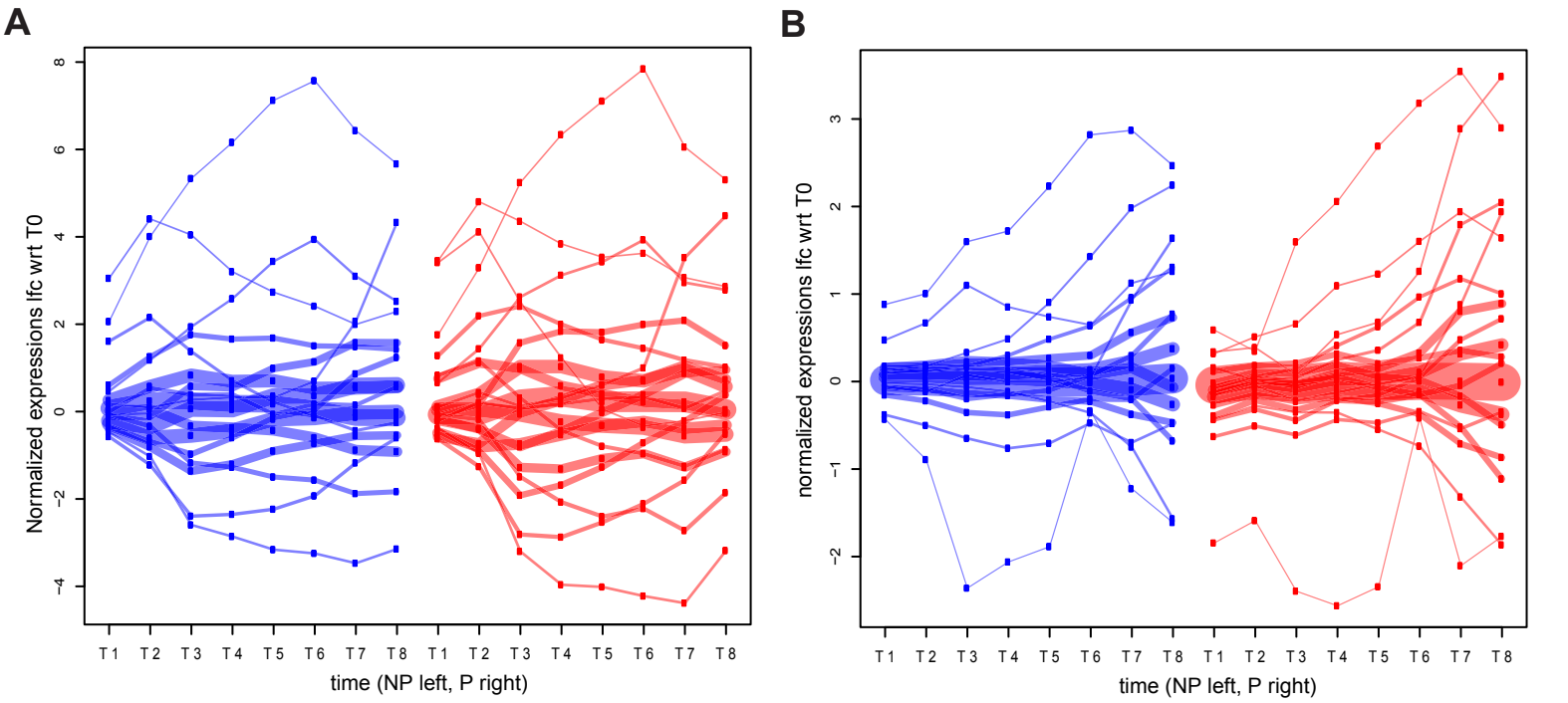
25. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004 Feb; **5**(2)**:** 101-113.

26. Meinshausen N, Bühlmann P. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2010; **72:** 417-473.

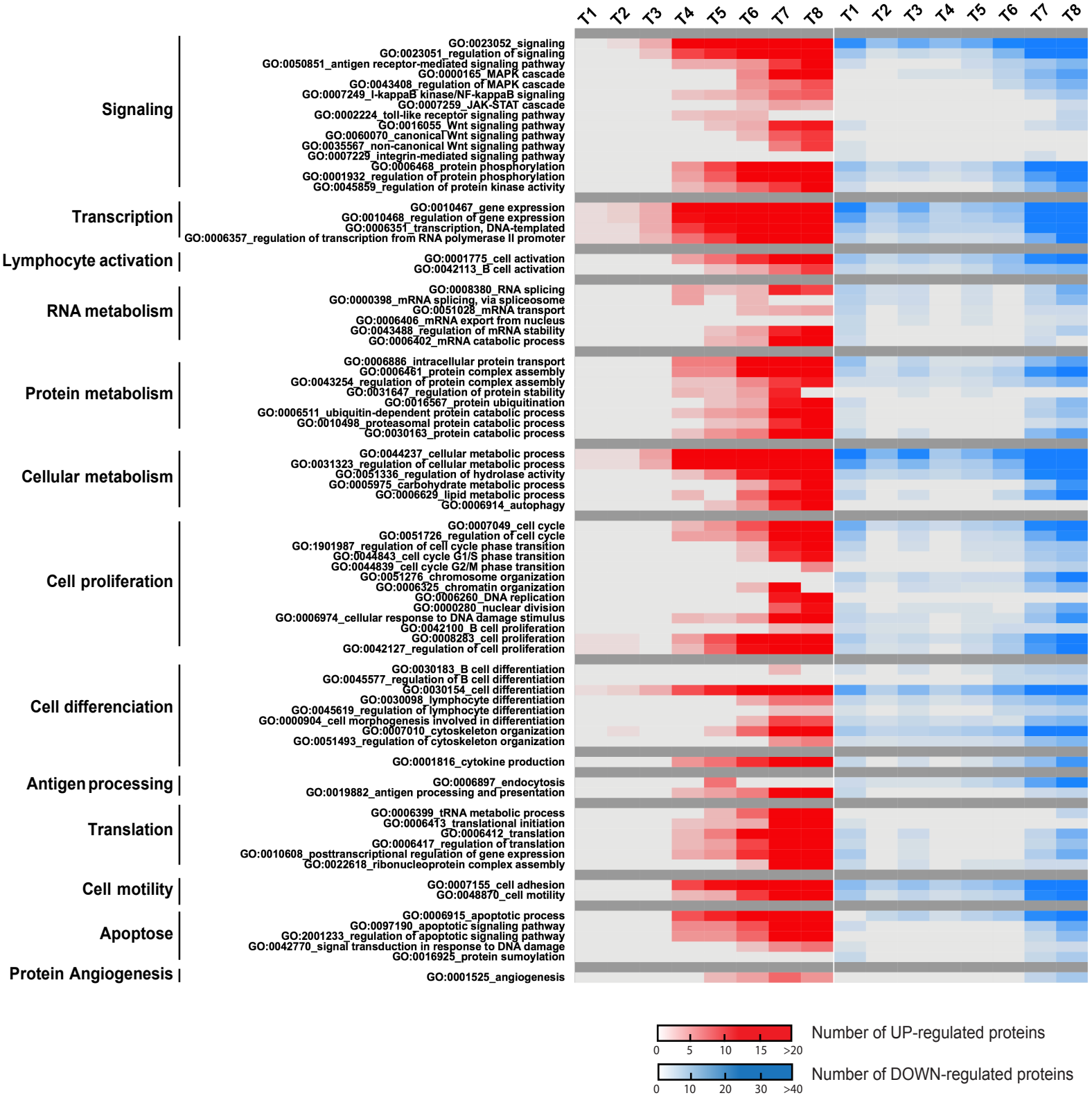
27. Bertrand F, Maumy-Bertrand M, Jung N, Aouadi I. SelectBoost : a General Algorithm to Enhance the Performance of Variable Selection Methods in Correlated Datasets. version1.3.0. *R package* 2019.



Supplementary Figure 1. Experimental design and methods outline.
Top panel. Three proliferative and three non-proliferative CLL samples have been included. BCR engagement in soluble medium was performed at T0 and cell samples were collected immediately before BCR engagement at T0 and at 8 time points after cell stimulation for gene expression (RNAseq) and protein abundance (mass spectrometry) measurements. At day six, cell proliferation has been quantified by flow cytometry after initial staining with CFSE. Bottom panel. Summary of the processing used for mRNA expression quantification, protein abundance quantification and mathematical analysis.

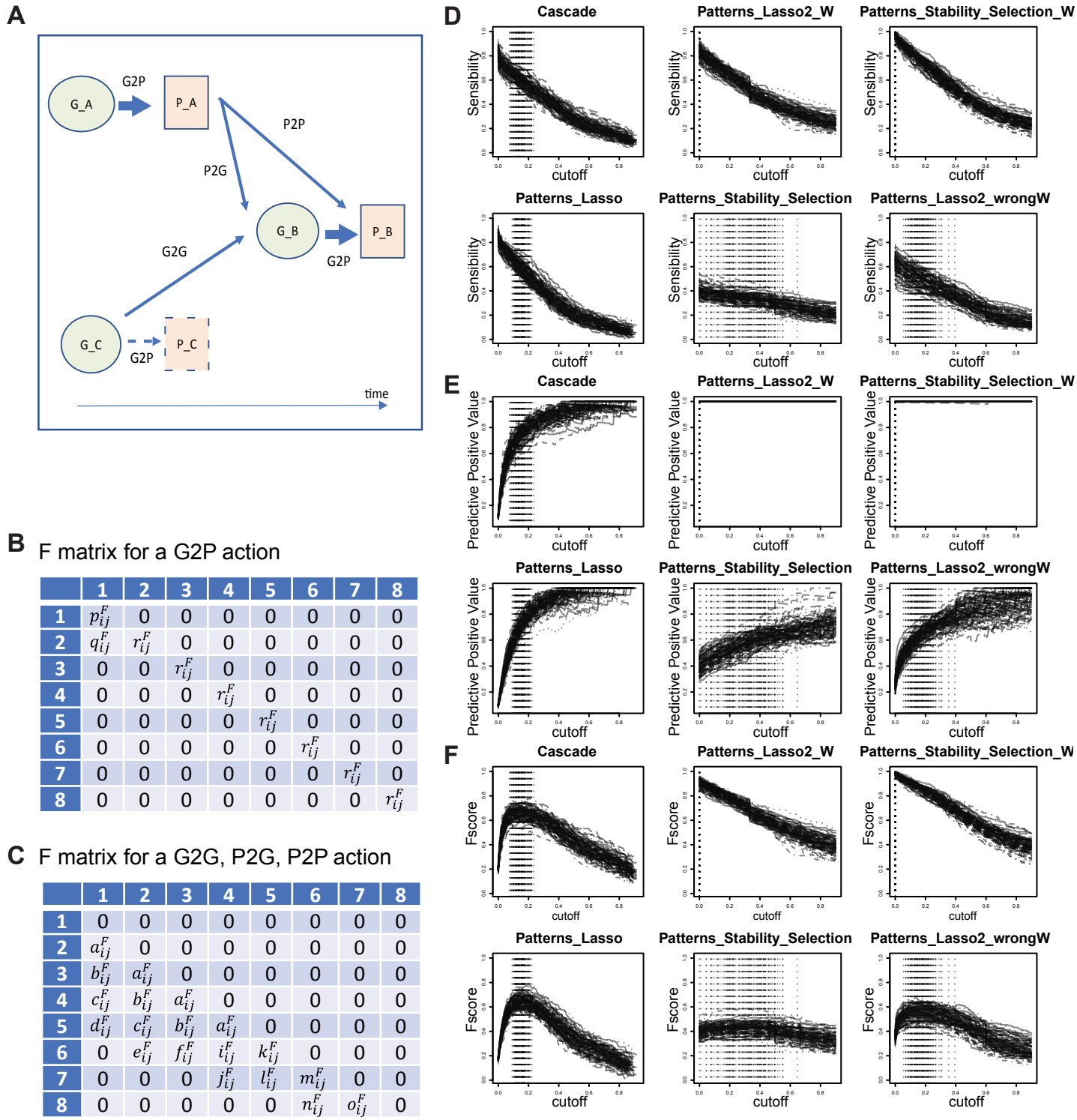


Supplementary Figure 2.
A Unsupervised temporal cluster of gene expression. Gene expression clustering of all proliferative (P, red) and non-proliferative (NP, blue) samples was unsupervised and performed separately. Median P or NP gene expression were used for a distinct representation of non-proliferative temporal cluster of gene expression or proliferative temporal cluster. The width of a line correlates with the size of the corresponding cluster. **B Unsupervised temporal cluster of protein abundance.** Same as (A) for the protein abundance clustering of the proliferative (P, red) and non-proliferative (NP, blue) samples.



Supplementary Figure 3. Temporal representation of biological functions (GO BP terms) of up- and down-regulated proteins in proliferating CLL samples.

The heat map shows with a color code, the number of proteins differentially up-regulated (red) or down-regulated (blue) assigned to a particular Gene Ontology (GO) function, at each time point (T1 to T8) after cell stimulation of the proliferating samples. Functional enrichment of biological processes (GO BP terms) has been analyses using DAVID on the 1,107 differentially up- or down-regulated proteins in the proliferating cell samples.



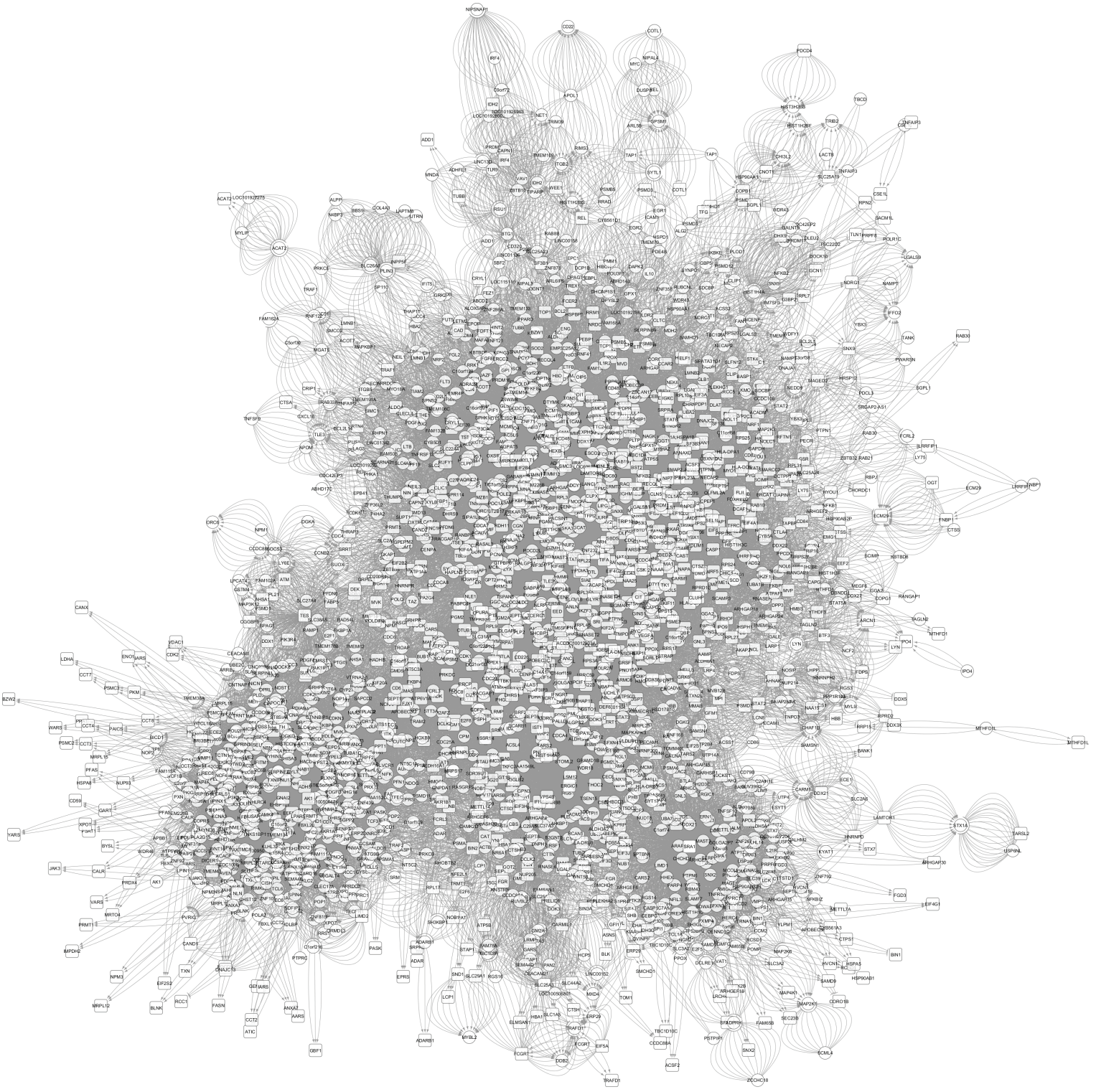
B F matrix for a G2P action

	1	2	3	4	5	6	7	8
1	p_{ij}^F	0	0	0	0	0	0	0
2	q_{ij}^F	r_{ij}^F	0	0	0	0	0	0
3	0	0	r_{ij}^F	0	0	0	0	0
4	0	0	0	r_{ij}^F	0	0	0	0
5	0	0	0	0	r_{ij}^F	0	0	0
6	0	0	0	0	0	r_{ij}^F	0	0
7	0	0	0	0	0	0	r_{ij}^F	0
8	0	0	0	0	0	0	0	r_{ij}^F

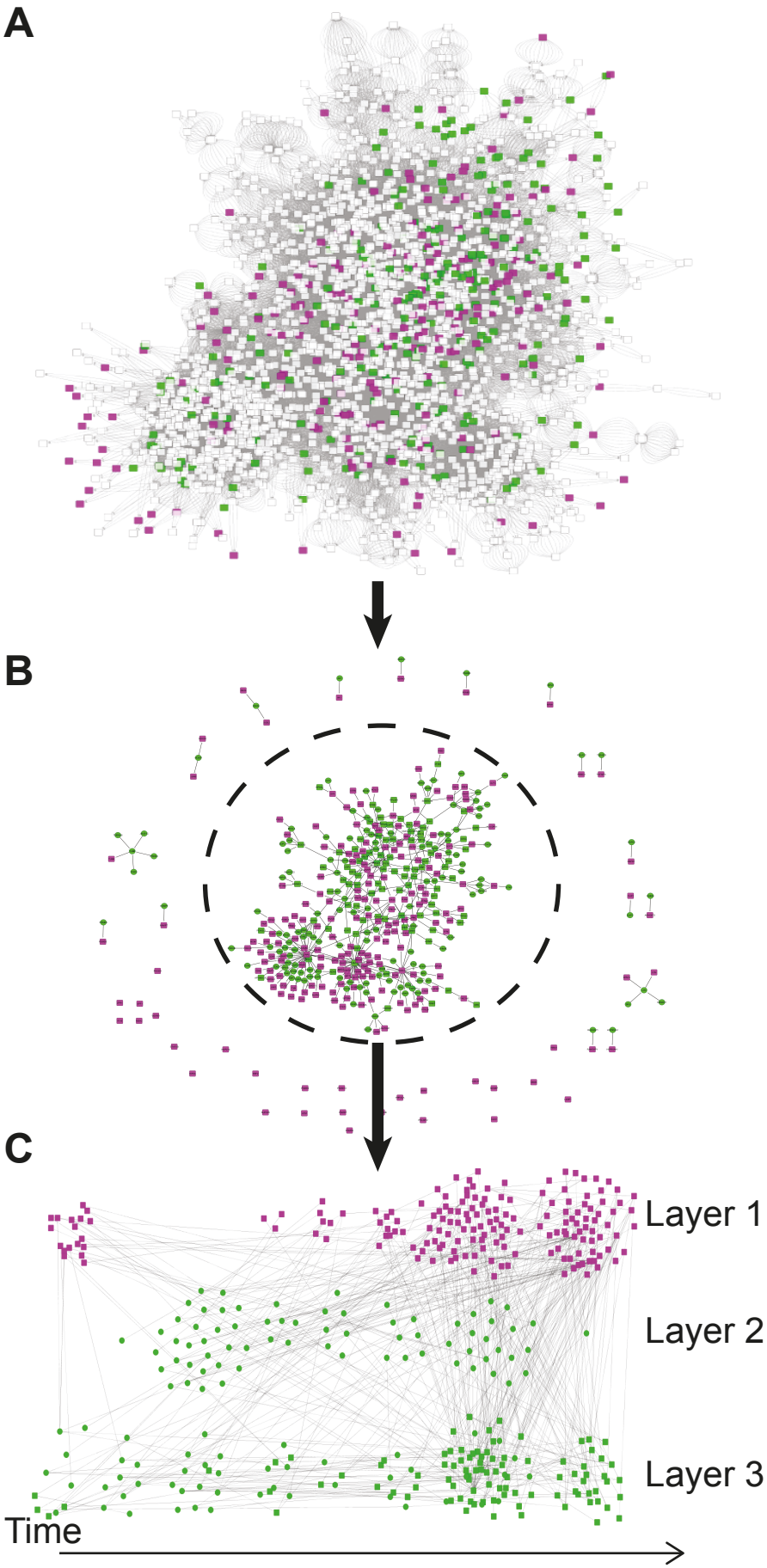
C F matrix for a G2G, P2G, P2P action

	1	2	3	4	5	6	7	8
1	0	0	0	0	0	0	0	0
2	a_{ij}^F	0	0	0	0	0	0	0
3	b_{ij}^F	a_{ij}^F	0	0	0	0	0	0
4	c_{ij}^F	b_{ij}^F	a_{ij}^F	0	0	0	0	0
5	d_{ij}^F	c_{ij}^F	b_{ij}^F	a_{ij}^F	0	0	0	0
6	0	e_{ij}^F	f_{ij}^F	i_{ij}^F	k_{ij}^F	0	0	0
7	0	0	0	j_{ij}^F	l_{ij}^F	m_{ij}^F	0	0
8	0	0	0	0	0	n_{ij}^F	o_{ij}^F	0

Supplementary Figure 4. Model formalization and validation.
A Genes and proteins are represented by circles and squares, respectively. Genes or proteins differentially expressed at least at one of the time point after cell stimulation are represented with a full lined circle or square. Genes or proteins not measured are represented with a dotted line. Authorized links between genes and/or proteins in the formalized model are represented by arrows (gene to protein (G2P), gene to gene (G2G), protein to gene (P2G), protein to protein (P2P)). Weighted links between genes and proteins with the same gene symbol (after translation of protein accession numbers in gene symbols (e.g. gene_A and protein_A) are represented in bold. Links between a protein with transcriptional factor activity and its known gene target (e.g. gene_B, or the corresponding protein_B if the gene_B is not measured in our experiment) are also weighted according to database information (e.g. RegNetwork). Matrix of temporal interactions shows authorized links between couple of time points for G2P (**B**) and for G2G, P2G and P2P (**C**). Each letter symbolizes a weight for a particular temporal link. To test our model, we analyzed the sensitivity (**D**), predictive positive value (**E**) and F-score (**F**) with Cascade algorithm, a non-weighted , an appropriately weighted and an incorrectly weighted version of our algorithm, a non-weighted stability selection based version of our algorithm (see methods). To simulate the measurements of the actors based on a regulatory network, we designed an algorithm that is inspired by the preferential attachment. Then, we adapted it to temporal nested networks. We then used our cascade network based model to make some simulations, using Laplace laws to set the values of the measurements of the actors at the first time point (see methods).

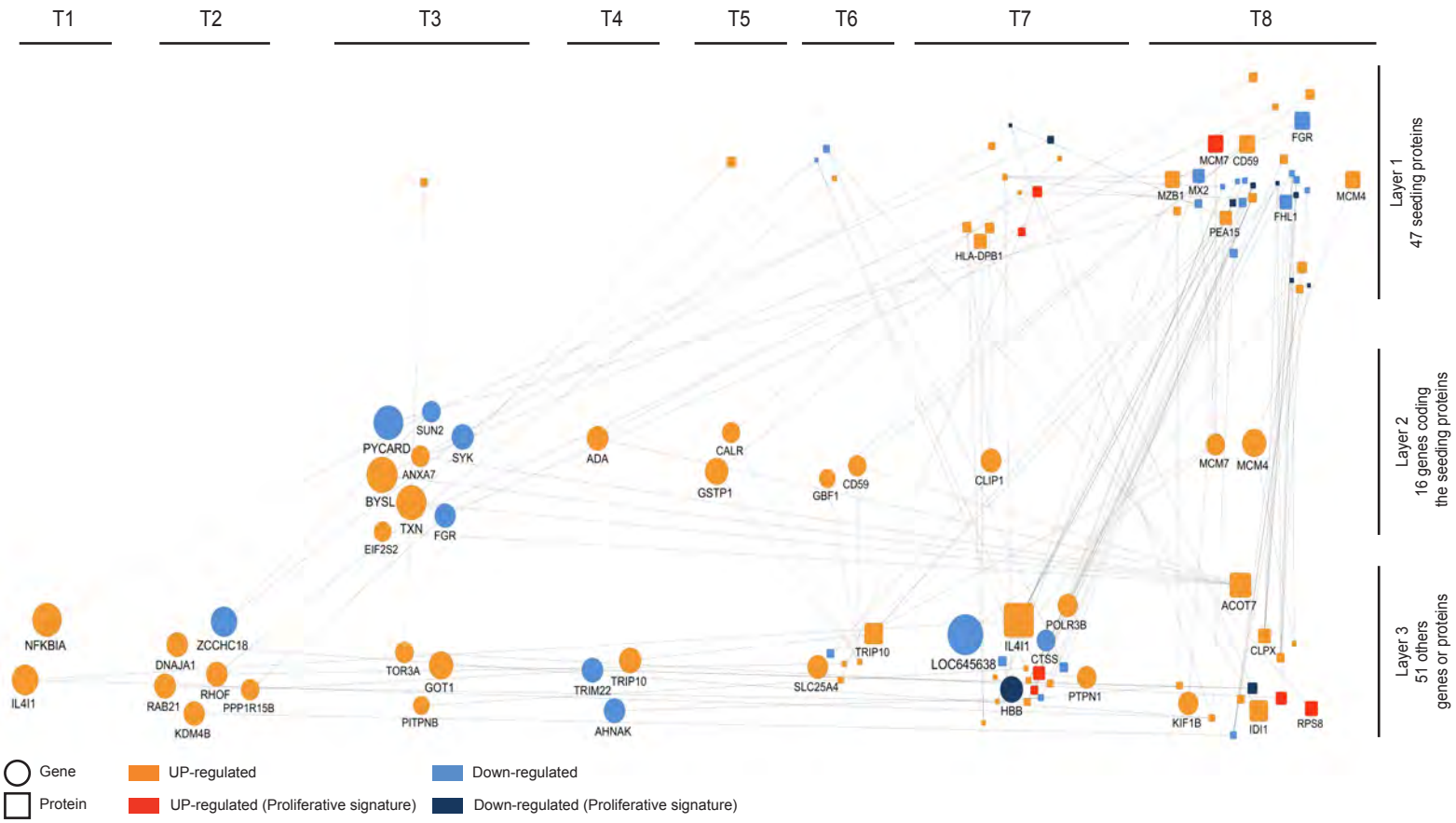


Supplementary Figure 5. Global model of genes and proteins temporal interactions after BCR stimulation in proliferating CLL samples.
Visualization of the joint inference of statistical interactions between genes and proteins differentially expressed at least at one of the time point (T v T0) after cell stimulation in proliferating cells. Circles represent genes, squares represent proteins and edges represent inferred links between genes and /or proteins across time. The graphical representation (generated with Cytoscape software) includes 2,167 genes and 1,074 proteins (for a total of 2,846 unique symbols). Potential interactions (represented in the graph when omega ≥ .01, see methods section) between genes and/or proteins expression over time are represented by time-directed arrows.



Supplementary Figure 6. A CLL proliferative program nested within the BCR-response program in proliferative CLL-cells.

A Search for protein associated with BP terms “cell-cycle regulation” or “proliferation” identified 267 seeding proteins (purple) in the global model of genes-proteins temporal interactions after BCR stimulation in proliferative CLL-cells. These seeding proteins are connected to 243 neighbors in the model (green). **B** These 267 seeding proteins and 243 neighbors organized into a major nested sub-network made of 173 of the seeding proteins and 215 neighbors. C Temporal representation of the nested sub-network, organized within three layers of actors.



Supplementary Figure 7. Non proliferating sub-network

Search for protein associated with BP terms “cell-cycle regulation” or “proliferation” identified 193 seeding proteins in the global model of genes-proteins temporal interactions after BCR stimulation in non-proliferative CLL-cells. These seeding proteins and neighbors organized into a major nested sub-network made of 47 proteins and 67 neighbors. Temporal representation of the nested sub-network, organized within three layers of actors (generated with Cytoscape software).