Full-length RNA-Seq of the RHOH gene in human B cells reveals new exons and splicing patterns

Frédéric Leprêtre, Jean-Pascal Meneboo, Celine Villenet, Laure Delestré, Bruno Quesnel, Carl Simon Shelley, Martin Figeac, Sylvie Galiègue-Zouitina

scientific reports

Check for updates

OPEN

Full-length RNA-Seq of the *RHOH* gene in human B cells reveals new exons and splicing patterns

Frédéric Leprêtre[1]✉, Jean-Pascal Meneboo[1], Céline Villenet[1], Laure Delestré[2], Bruno Quesnel[3], Carl Simon Shelley[4], Martin Figeac[1] & Sylvie Galiègue-Zouitina[5]✉

The RhoH protein is a member of the Ras superfamily of guanosine triphosphate-binding proteins. RhoH is an atypical Rho family member that is always GTP-bound and thus always activated. It is restrictively expressed in normal hematopoietic cells, where it is a negative regulator of cell growth and survival. We previously analyzed the *RHOH* gene structure and demonstrated that this gene is composed of 7 exons, one single encoding exon located at the 3′ extremity of the gene, preceded by 6 noncoding exons. To further understand the transcription events associated with this gene, we performed full-length RNA-Seq on 12 B-cell lines. We identified new exons, new splice events and new splice sites, leading to the discovery of 38 *RHOH* mRNA molecules, 27 of which have never been described before. Here, we also describe new fusion transcripts. Moreover, our method allowed quantitative measurements of the different mRNA species relative to each other in relation to B-cell differentiation.

Keywords *RHOH* gene, Full-length targeted mRNA-Seq, Oxford Nanopore Technologies, New exons, New transcripts, Gene fusion, B-cell differentiation

We previously isolated the human *RHOH* (*TTF*) gene by fusion to the *BCL6* (*LAZ3*) gene, which was induced by a t(3;4) secondary recurring translocation, in follicular B non-Hodgkin's lymphoma cells[1,2]. Fluorescence in situ hybridization (FISH) experiments allowed us to refine *RHOH* gene mapping to 4p13-14[3], which was subsequently accurately assigned to 4p14 by the human genome sequencing project (hg19, chr4:40198527–40246281). We further showed that in normal tissues, *RHOH* is restrictively expressed in hematopoietic cells[1,4], in agreement with the findings of other studies[5], and that the 21 kDa hemato-specific RhoH protein is encoded by a single exon containing a major ORF of 576 bp[1]. We finally characterized the *RHOH* gene structure[3,4].

In normal T cells, the RhoH protein regulates activation and differentiation[5–9] and TCR signaling[10–12]. In hemopoietic progenitors, RhoH modulates intercellular interactions, proliferation, survival, homing and chemotaxis[13,14]. It also decreases leukotriene production in neutrophils[15] and negatively regulates eosinophilopoiesis[16]. Despite a lack of knowledge about RhoH function in normal B cells, *RHOH* gene deregulation has been extensively studied in tumor B cells, especially in B-cell pathologies such as hairy cell leukemia[17], chronic lymphocytic leukemia[18,19], and B-nonHodgkin's lymphoma of numerous subtypes, first in diffuse large B-cell lymphoma (DLBCL), where aberrant somatic hypermutation of the *RHOH* gene has been found[20]. The same hypermutation mechanisms have been described in many other subtypes of B-cell lymphoma[21–23]. Additionally, in the B-cell lineage, RhoH functions as a negative regulator of IL3-induced signaling through modulation of the JAK-STAT pathway[23]. Notably, the *RHOH* gene is also expressed in myeloid cells, where its induction could drive terminal differentiation and might represent a means of affecting acute myeloid leukemia (AML) differentiation[24].

RhoH is an atypical member of the Rho family of small GTP-binding proteins[5,25], suggesting its important role in both normal and disease states and suggesting that the *RHOH* gene is a potential target in therapeutic treatments. In contrast to classic members, atypical members either exhibit a high intrinsic guanine nucleotide exchange rate (RhoU and RhoV) or lack the ability to hydrolyze GTP[5,26]. Consequently, the atypical RhoH member is constitutively active and is regulated by mechanisms such as posttranslational modifications, such as phosphorylation[21], subcellular localization and lysosomal degradation[22], or by quantitative mRNA expression[5].

[1]Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, US 41-UAR 2014-PLBS, 59000 Lille, France. [2]Inserm U1170, 94800 Villejuif, France. [3]CHU Lille, UMR-S 1277-Canther-Cancer Heterogeneity, Plasticity and Resistance to Therapies, Université de Lille, 59000 Lille, France. [4]Leukemia Therapeutics, LLC, Hull, MA 02045, USA. [5]Institut pour la Recherche sur le Cancer de Lille, Place de Verdun, 59054 Lille, France. ✉email: Frederic.lepretre@univ-lille.fr; Sylvie.zouitina@outlook.fr

Therefore, transcriptional and posttranscriptional events targeting the *RHOH* gene seem to be key points to be considered. In particular, the patterns of transcripts and modulation of their expression as a function of several physiological events, such as differentiation, are important. In a previous study aimed at identifying new *RHOH* exons in the 5′ region of the gene possibly involved in the modulation of RhoH protein translation, our strategy was to use 5′Race-PCR starting from the unique encoding exon, previously named exon 2, in the B-cell line Raji[4]. This led us to analyze *RHOH* transcript patterns in different hematopoietic cells from the B, T, and myeloid lineages, allowing us to identify six 5′ noncoding exons. We first named these genes X1, 1a, X2, 1b, X3, and X4 (Table S1, for accession numbers). We thus described the *RHOH* 5′-UTR as characterized by complex splicing patterns involving up to 6 upstream noncoding exons alternatively spliced to the unique encoding exon at the 3′ extremity of the gene. Moreover, our study emphasized differences in transcript patterns as a function of the hematopoietic lineage. In particular, some noncoding exons were shown to be specific to B or T cells[4]. Next, we simplified the *RHOH* exon nomenclature from exons X1, 1a, X2, 1b, X3, X4, and 2 to exons 1–7[27]. According to this new nomenclature, exons 1a and 1b are now called exon 2 and exon 4, respectively, and the unique RhoH-encoding exon is called exon 7. Primer extension assays revealed the presence of transcription initiation sites upstream of exons 1 (X1), 2 (1a) and 4 (1b)[4], which was confirmed by further studies; in particular, by the cloning and complete characterization of one *RHOH* promoter next to exon 4, we called this promoter P3[27]. Our work therefore revealed that the *RHOH* gene can be transcribed from 3 distinct promoters, P1, P2 and P3, upstream of exons 1, 2 and 4, respectively[4,27].

Our present work focused on 12 human B-cell lines generated from different hematopoietic disorders and mimicking 4 differentiation stages, from pre-B to terminal plasmocytic differentiation, using up-to-date genomic studies for discovering and characterizing additional *RHOH* gene transcript patterns.

Materials and methods

Cell culture

Pre-B-cell line models representing the early stage of B-cell differentiation (Nalm-6 and Lila-1, derived from acute lymphoblastic leukemia patients) were obtained from the German Collection of Microorganisms and Cell Culture (DSMZ, Braunschweig, Germany) and kindly provided by Martin Dyer (MRC Toxicology Unit, Hodgkin Building, Leicester, United Kingdom). The pre-B-cell line Lila-1 was treated with phorbol myristate acetate (PMA) to induce differentiation. Both cell lines were cultured in RPMI 90% supplemented with 10% FBS, and when indicated, Lila-1 cells were induced to differentiate by treatment with 20 ng/mL PMA or the same volume of the corresponding vehicle DMSO for 24 to 72 h. Lila-1 ± PMA-treated cells were analyzed by flow cytometry to determine the increase in the expression of CD19 surface markers following PMA treatment (from 10% after 24 h to 40% after 72 h), and cell counting was used to determine whether the cells stopped dividing when they were differentiated. Mature Burkitt B cells (Raji) were obtained from American Type Culture Collection (ATCC; Manassas, USA). Mature B cells Burkitt BJA-B, diffuse large B-cell lymphoma (DLBCL), HBL-1 and U-2932 were obtained from Dimitar Efremov (International Centre for Genetics Engineering and Biotechnology, ICGEB, Trieste, Italy) and were grown in RPMI 90% supplemented with 10% FBS. The DLBCL cell lines OCI-Ly1 and OCI-Ly3 were obtained from Mark Minden MD, PhD (Princess Margaret Cancer Centre, University Health Network, Toronto, Canada) and were grown in 80% IMDM supplemented with 20% FBS. Multiple myeloma cell lines, RPMI 8226 and U266, which represent the late stage of B-cell differentiation, were obtained from the ATCC and cultured in RPMI 90% supplemented with 10% FBS. Finally, the myeloid cell line control HL60 was obtained from the DSMZ and was grown in the same way. All cells were grown in the presence of 100 units/ml penicillin and 100 µg/ml streptomycin.

RNA extraction

Total RNA was isolated using an RNeasy Kit from Qiagen (Düsseldorf, Germany) following DNAseI treatment. All RNA samples were subsequently tested for integrity on an Agilent 2100 Bioanalyzer (Santa Clara, California, USA) for RIN calculation, and concentrations were assessed on a NanoDrop 2000 from Thermo Fisher (Waltham, Massachusetts, USA). All the RINs reached 10.0 and were therefore included in further analysis.

NanoString nCounter Expression Assay

The nCounter expression assay was performed according to the manufacturer's instructions (NanoString Technologies, Seattle, USA). A custom panel was designed comprising 3 pairs of probes targeting the *ABL1*, *GAPDH* and *TBP* genes used for normalization and 18 pairs of probes targeting known parts of the *RHOH* transcripts and *RHOH* predicted exon sequences (extracted from the UCSC Table Browser for Unigene and SIB databases). The names, sequences and positions of the probes in the hg19 genome are given in Table S1.

The custom CodeSet was hybridized to 300 ng of total RNA from the 12 cell lines for 20 h at 67 °C in a VERITI thermal cycler (Thermo Fisher Scientific, Waltham, Massachusetts, USA). After hybridization, the samples were purified and immobilized in a sample cartridge using the automated nCounter® Prep Station. This cartridge was scanned in an nCounter® Digital Analyzer for quantification of mRNA targets in each sample. Analysis of the assays was performed with nSolver version 4.0 software (NanoString Technologies, Seattle, USA), including quality control of the raw data and normalization to the counts of *ABL1*, *GAPDH* and *TBP* (Table S2). The results were visualized through IGV software (Broad Institute, Cambridge, USA).

3′RACE-PCR

We performed 3'RACE PCR using the 3′RACE System for Rapid Amplification of cDNA Ends Kit from Invitrogen according to the manufacturer's protocol (Waltham, Massachusetts, USA). The primer sequences are given in Table S1. Briefly, cDNA synthesis was initiated using an oligo-dT adapter primer (AP) on 1 µg of total RNA per cell line. Then, amplification of the target cDNA fragments was performed using a custom primer

annealed to a specific site of interest (SSI) and one universal primer that targeted the polyA region (AUAP). PCR was subsequently performed to amplify cDNA fragments starting from exon 2 or exon 4 to downstream sequences up to the 3′ end of the *RHOH* gene (Fig. 1).

The following forward primers were used: ex2RACElong or ex4RACElong, each of which was combined with the AUAP reverse primer. To increase specificity, nested PCR was conducted by combining *RHOH* gene-specific primers adjacent to the first SSI (ex2RACEnested or ex4RACEnested) with the universal AUAP primer. The sequence products started 49 bp and 24 bp downstream of the documented transcription initiation sites of exon 2 and exon 4, respectively, to ensure the accuracy of the sequencing reactions. Finally, the PCR products were purified using AMPure Beads XP (Beckman, Brea, USA) and quantified on an Agilent 2100 bioanalyzer. The primers used were designed using Oligo7.6 software (Molecular Biology Insights, Cascade, USA) to allow homogeneous amplification, and the positions at the most 5' regions in the exons were selected.

Library preparation

The 3′RACE-PCR products (starting at exon 2 or exon 4) were pooled in equimolar amounts, and 500 ng was used to prepare barcoded libraries using a Ligation Sequencing Kit (SQK-LSK109) and a Native Baring Expansion Kit from Oxford Nanopore Technologies, according to the manufacturer's protocol (1D Native Barped Genomic DNA Rev E). After qualification and quantitation on an Agilent 2100 bioanalyzer, all the barcoded libraries were mixed in equimolar amounts.

MinION sequencing and analysis

As previously described, sequencing was performed on an Oxford Nanopore MinION sequencer with a SpotON Flowcell (MIN106-R9.4.1) for 24 h using the MinKNOW program. The run was followed by base-calling and demultiplexing via the guppy tool (version 2.3.5). Quality control was carried out by using fastQC (version 0.11.5) and MinIONQC (version 1.3.5). Demultiplexing by use of barcodes led to the identification of transcript samples, starting from exon 2 or exon 4, as mentioned above. The sequences were filtered by use of cutadapt (version 2.1) and Nanoflit (version 2.2.0) to keep sequences of interest only, filtering out sequences without barcodes, those below a length of 600 bp and those of low quality (phred read quality score under 10). Each read of interest was also selected to obtain 5′ and 3′ primers in the sequence. These filters allowed us to keep the complete sequences of the transcripts after trimming the barcode and adaptor sequences.

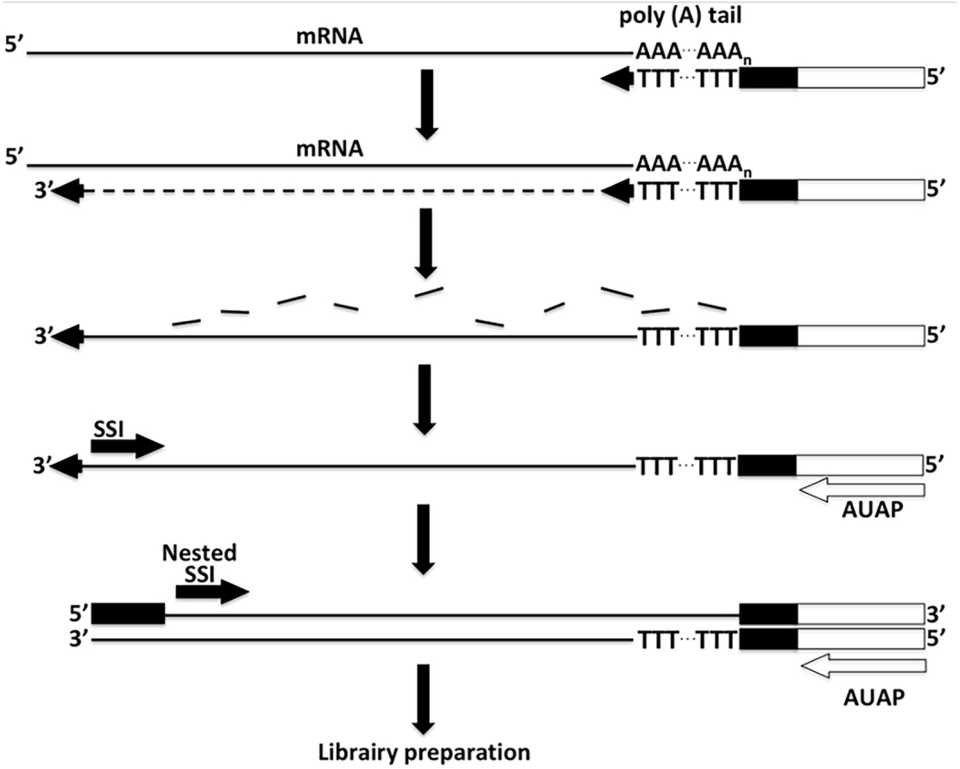

Fig. 1. Experimental workflow used for full-length *RHOH* transcript sequencing. First-strand cDNA synthesis was initiated at the poly(A) tail of mRNAs using an adapter primer (AP). After first-strand cDNA synthesis, the original mRNA template is degraded with RNase H, which is specific for RNA-DNA heteroduplex molecules. Amplification was performed using 3 primers, the AUAP (Abridged Universal Amplification Primer), which is homologous to the adapter sequence for first strand cDNA synthesis; the SSI (Single Sequence of Interest), for the amplification of all targeted sequences for 2 cycles; and, third, a nested SSI to allow more specific amplification of the *RHOH* transcripts for 35 cycles. The primer sequences are given in Table S1.

Sequence reads were subsequently aligned to the human genome (hg19) using minimap2 software (version 2.1) and reference gene annotations (GRCh37/hg19 release 87). To discover new transcripts, we used pinfish software (version 0.1) to cluster reads with similar exon/intron structures to create a consensus sequence. Transcript expression was quantified from the alignment files (.bam files) and gene annotations (pinfish) using flair software (Fig. 2).

Finally, to ensure the reliability of our methods and findings, we applied a minimum depth threshold of 500 to the read count matrix.

Prediction of potential proteins

According to the length of the discovered transcripts (529–1759 bp), we used ORFfinder software from the NCBI (https://www.ncbi.nlm.nih.gov/orffinder/) to search for open reading frames (ORFs) for potential protein -encoding segments and their predicted proteins/peptides. For complete analysis and to fit to biological process, we chose to add to each transcript sequence the 49 and 24 bp between the transcript initiation sites and the start of sequencing, respectively, for exons 2 and 4. Using standard "ATG" and alternative noncanonic initiation codons, ORFs were generated and filtered by selecting proteins with weights above 7 kDa. In addition, we used the protein families database (Pfam) tool (https://pfam.xfam.org) from EMBL-EBI for sequence alignment.

Results

On the basis of our previous study[27], we showed that the P3 promoter upstream of exon 4 is predominant and active mainly in normal T and B cells, while the P1 promoter is silent in normal B cells. In contrast, P2 promoter activity upstream of exon 2 seems to be restricted to the B-cell lineage, where it shows similar activity to that

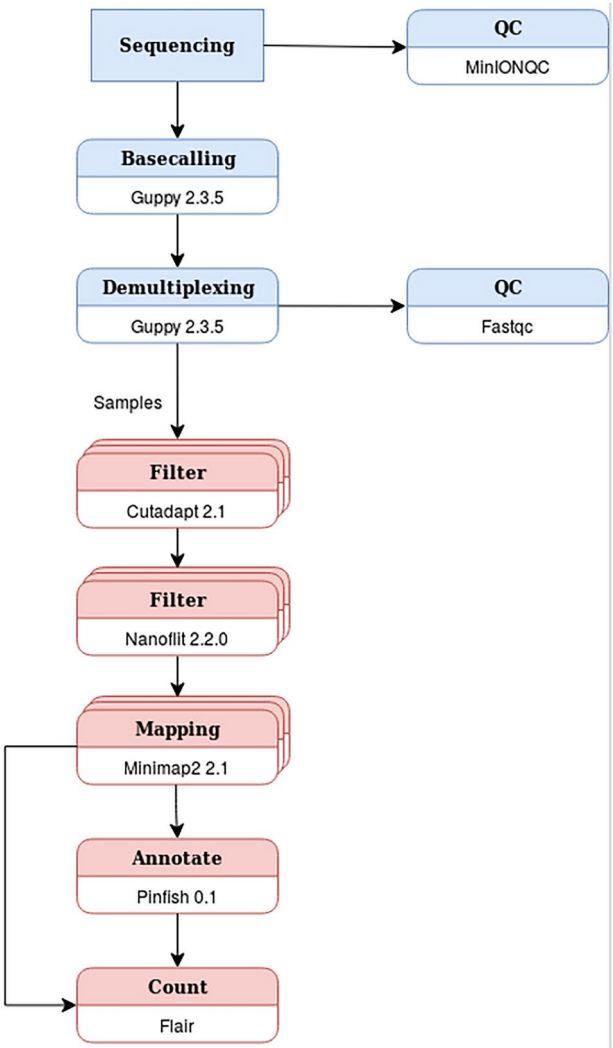

Fig. 2. Bioinformatics workflow used for full-length *RHOH* transcript sequencing. The bioinformatics workflow is separated into 2 parts. The first part (in blue) corresponds to the primary analysis, with basecalling and demultiplexing of the samples. The second part (in red) corresponds to the secondary analysis involving filtering, mapping, annotation and counting for each sample.

of P3 (Fig. S1). Additionally, PCR were loaded on 2% agarose gels stained with BET (Fig. S2). We then focused on the P2 and P3 active promoters to perform our present study. For this purpose, we analyzed the total RNA of the nCounter (NanoString Technologies, Seattle, USA) potential new *RHOH* exons, which were predicted or potential exons from different sources (UCSC, Swiss Institute of Bioinformatics, Unigene and Ref-Seq databases), to ensure the future of new discoveries. We subsequently performed a full-length *RHOH*-targeted mRNA sequencing approach by using 3′Race-PCR downstream of either exon 2 or exon 4, followed by nested PCR and MinION sequencing (Oxford Nanopore Technologies, Oxford, UK).

NanoString analysis of B cells allowed the discovery of potential new *RHOH* exons

On the basis of our previous work using a 5′Race-PCR[4] approach, "priming" inside the RhoH gene encoding an exon allowed the upstream identification of new exons. Furthermore, several other *RHOH* exon predictions were mentioned in the University of California, Santa Cruz (UCSC), the Swiss Institute of Bioinformatics (SIB) and Unigene. We selected from UCSC 8 ESTs we called them EST1-8 (5 predicted exons and 3 identified as exons: EST2, -6, and -7), and from Unigene, 3 predicted exons we called them Uni1-3 to search for their presence in total RNA by use of the nCounter. Therefore, we designed 18 pairs of probes for EST1-8 and Uni1-3 for 6 out of the 7 *RHOH* exons we had previously identified; exons 1–4, 6, and 7; and the *RHOH* 3′-UTR as a control. Notably, EST7 was identified as *RHOH* exon 5 in our protocol. Three genes (*ABL1*, *GAPDH*, and *TBP*) were added for normalization of the raw data via nSolver software.

Figure 3 shows that our assay allowed the identification of all exons we had previously identified (in blue), except for exon 1, which was not expressed in normal B cells (Fig. S1).

Exons 4 and 6 and the encoding exon 7, together with the 3′-UTR, presented the highest counts in all 12 B-cell lines, in agreement with previous observations leading us to call them "major exons" in hematopoietic cell lines from all lineages[4]. In contrast, both exons 2 and 3 exhibited weaker counts but were detected in all B-cell lines but not in the myeloid cell line control HL60 (Fig. 3). Interestingly, all the predicted exons, EST1-8 and Uni1-3, were detected, with weak counts (Fig. 3, in red), similar to exons 2 and 3 (Table S2, showing the normalized counts that were obtained). Thus, this approach highlighted the possible existence of 10 potential new exons in RHOH mRNAs (Fig. 3).

Full-length mRNA-Seq analysis of the *RHOH* gene in B cells

In accordance with the results of our nCounter study, we performed a targeted full-length mRNA-Seq analysis of the *RHOH* gene. This study aimed to characterize new transcript patterns governed by both the P2 and P3 promoters in B cells and to determine the presence of possible new exons, as detected in our nCounter study. With the Oxford Nanopore MinION sequencer, we achieved a sequence of 7.63 gigabases, leading to a total of 7,774.901 reads with median and maximum lengths of 819 and 8886 bases, respectively, with 99.5% read quality > Q7 (the nanopore default passed quality). The exons discovered in our project were named with respect to the exon nomenclature in Delestré's paper or adapted with respect to the known closest one. Figures 4 and 5 present the positions and exon successions of the 38 *RHOH* mRNA molecules that were identified, numbered from T01 to T38, and their relative percentages in the 12 B-cell lines.

Figures 6, 7, 8 and 9 precisely illustrates 21 of the mRNA molecules that we identified.

All sizes and positions of exons, in addition to transcript structures, are given for additional information (Figs. 4 and 5 and Table S1).

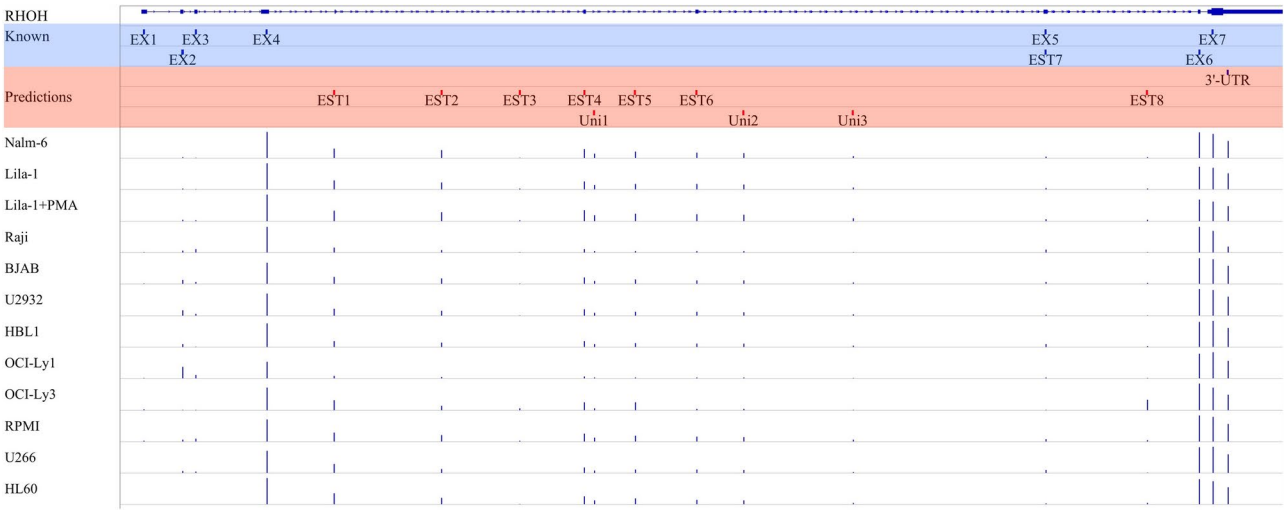

Fig. 3. NanoString-normalized counts of probes within the *RHOH* gene for all cell lines. Normalized counts are visualized as bar charts on the Integrative Genomics Viewer (IGV version 2.8.3; Broad Institute, Cambridge, USA) with the autoscale option turned on and the hg19 genome. Known sequences of the *RHOH* gene and predicted sequences are represented under the gene track with blue and red marks, respectively. Samples are shown with respect to increasing differentiation status.

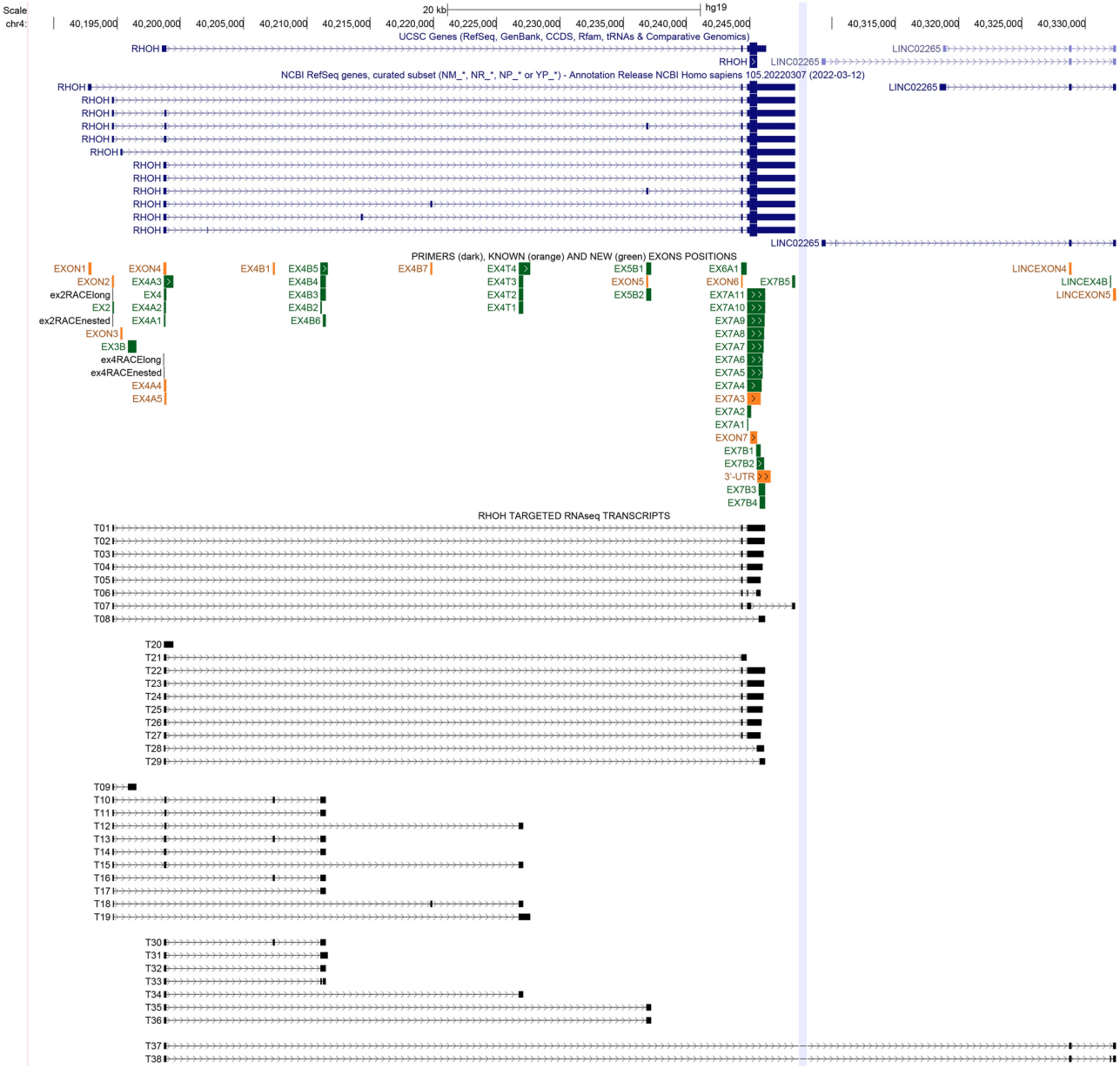

Fig. 4. Representation of the *RHOH* gene region with all *RHOH* mRNAs detected by mRNA sequencing in B-cell lines. The *RHOH* gene region is described here in the GRCh37-hg19 genome assembly. The primers used for 3'RACE-PCR were used; former exons (orange) and new exons (green) are shown in the third track. Finally, the observed transcripts are presented in black below the corresponding exons. We used the multiregional option of the UCSC genome browser to locate the *RHOH* and *LINC02265* genes more closely related to each other and resize the 63 kb gap between both genes. This is symbolized by a blue bar.

First, we identified *RHOH* transcript patterns containing known exons. Eighteen mRNA molecules (Fig. 4, in black), 8 of which were initiated from exon 2 (T01–08) and 10 from exon 4 (T20–29), contained either exons 6–7 (T01–07 and T22–27, respectively, for 7 and 6 mRNAs), exon 7 alone (T08, T28–29), or exon 6 alone (T21), which were spliced to exon 2 or exon 4. Both T06 and T07 contained split versions of exon 7 with almost total excision of the RhoH protein ORF (Fig. 6); consequently, they might not be templates for RhoH protein synthesis, as did T08, T28 and T29, which contained truncated versions of exon 7, or T21, which contained only exon 6 spliced to exon 2 (Fig. 6). In contrast, in transcripts T01–05 and T22-27, both exons 6–7 were entirely spliced (Figs. 4, 6: T01 and T22 as examples), indicating that these 11 mRNAs could encode the RhoH protein. Notably, in these mRNAs, exon 7 presented 9 different 3'-UTRs, leading to the exon 7 configuration EX7A3 to EX7A11 (Fig. 5; Tables S1, S3). Both T01 and T22, which had the largest 3'-UTRs, were found to be major transcripts, representing nearly 20% of the total transcripts in 10 out of the 12 cell lines (Fig. 5).

Strikingly, the T20 and T21 mRNAs contained larger versions of either exon 4 alone or exon 6 spliced to exon 4 (Fig. 6), possibly corresponding to a hypothetical "intron retention"[28] or to another mechanism leading to

Transcript	Exon	Nalm-6	Lila-1	Lila-1+PMA	Raji	BJAB	U2932	HBL1	OCI-Ly3	OCI-Ly1	RPMI	U266	HL60
T01	EX2-EX6-EX7A10	16,0	19,8	18,0	10,5	22,7	17,9	19,7	14,7	34,5	18,4	15,8	22,4
T02	EX2-EX6-EX7A9	11,6	14,0	12,9	7,2	12,0	8,8	9,5	8,7	11,3	11,0	9,8	13,1
T03	EX2-EX6-EX7A7	7,9	9,5	11,2	5,9	8,6	6,5	6,9	3,8	6,6	7,8	8,4	7,2
T04	EX2-EX6-EX7A6	7,8	9,5	11,3	6,1	8,5	6,2	6,2	3,2	6,2	7,7	7,9	5,8
T05	EX2-EX6-EX7A3	6,6	8,8	9,8	5,7	7,0	5,1	5,3	2,5	4,8	6,4	7,3	5,8
T06*	EX2-EX6-EX7A1-EX7B1	1,1	13,2	0,4	0,0	1,7	1,3	2,0	2,7	0,1	0,0	0,4	3,3
T07	EX2-EX6-EX7A2-EX7B5	15,3	0,0	0,1	0,0	0,2	0,2	2,0	0,3	0,1	0,0	0,1	3,6
T08*	EX2-EX7B3	1,0	9,6	0,1	0,1	1,5	1,5	2,7	15,3	0,3	0,2	0,4	18,1
T09*	EX2-EX3B	0,7	0,7	19,4	4,5	5,9	1,4	4,5	13,1	0,9	0,3	1,4	8,4
T10	EX2-EX4A5-EX4B1-EX4B4	2,5	0,1	0,1	2,1	0,3	3,7	1,0	0,4	0,8	0,1	1,9	0,2
T11	EX2-EX4A5-EX4B4	0,6	0,1	0,3	13,2	2,6	10,4	7,7	2,9	4,6	4,4	19,6	0,9
T12*	EX2-EX4A5-EX4T3	0,3	0,1	0,0	19,4	0,2	3,3	0,2	1,0	1,8	10,0	9,4	0,5
T13*	EX2-EX4A4-EX4B1-EX4B4	0,5	3,6	0,2	2,1	3,0	4,7	1,3	1,2	2,1	1,7	1,6	0,5
T14*	EX2-EX4A4-EX4B4	15,8	0,3	13,4	10,1	11,3	15,3	9,5	2,4	7,7	3,1	8,1	1,0
T15*	EX2-EX4A4-EX4T2	0,4	2,3	0,2	10,4	1,6	2,0	0,6	1,2	3,9	12,8	1,8	0,5
T16	EX2-EX4B1-EX4B3	0,2	0,1	0,1	0,1	0,4	2,9	11,0	0,7	2,2	6,0	0,1	0,1
T17*	EX2-EX4B3	10,6	8,0	0,6	0,9	5,3	2,7	5,1	1,5	8,9	0,2	1,3	1,4
T18	EX2-EX4B7-EX4T1	0,3	0,1	0,0	0,1	0,2	2,2	0,5	0,9	0,8	3,8	0,7	0,3
T19*	EX2-EX4T4	0,6	0,4	1,9	1,5	7,0	3,7	4,4	23,4	2,5	6,3	3,8	6,6
T20*	EX4A3	12,2	8,0	1,4	6,6	5,8	3,5	5,9	3,6	1,8	15,3	2,4	2,5
T21	EX4-EX6A1	18,3	17,8	21,9	3,7	3,8	13,8	0,2	0,2	1,5	11,0	3,9	19,2
T22	EX4-EX6-EX7A11	13,4	22,5	24,1	23,6	24,3	18,4	19,2	22,2	22,8	23,4	25,9	24,1
T23	EX4-EX6-EX7A8	5,4	7,6	8,7	9,4	8,2	5,6	5,3	7,0	6,5	8,0	8,3	10,3
T24	EX4-EX6-EX7A7	3,9	4,3	5,2	6,6	5,0	3,4	3,3	4,5	3,6	4,6	5,6	7,6
T25	EX4-EX6-EX7A5	3,2	3,7	4,5	5,8	3,9	2,8	2,8	4,1	2,9	3,5	5,3	7,0
T26	EX4-EX6-EX7A4	3,0	3,2	4,2	5,2	3,7	2,6	2,5	3,5	2,5	3,3	4,8	6,8
T27	EX4-EX6-EX7A3	3,0	3,1	3,9	4,9	3,4	2,6	2,5	3,5	2,5	3,1	4,6	6,8
T28	EX4A1-EX7B2	0,8	0,6	0,5	0,0	0,3	0,6	0,5	5,2	0,1	1,7	0,4	1,3
T29	EX4A2-EX7B4	0,1	0,0	0,0	0,3	0,2	0,2	0,1	0,0	0,0	0,0	0,1	2,7
T30	EX4-EX4B1-EX4B4	1,3	0,5	0,6	1,2	1,1	1,4	1,0	2,2	1,6	0,2	1,4	0,0
T31	EX4-EX4B5	12,7	9,8	9,2	5,9	11,5	20,2	21,4	17,0	22,0	8,6	16,1	3,2
T32	EX4-EX4B4	12,9	9,5	9,4	4,6	10,7	18,1	19,6	17,2	19,7	8,7	14,1	2,1
T33	EX4-EX4B2-EX4B6	0,2	0,0	0,1	0,6	0,9	0,5	3,7	0,1	3,5	0,5	0,3	0,0
T34	EX4-EX4T3	2,2	2,2	2,4	1,7	1,3	3,4	2,5	0,6	6,0	3,3	2,5	1,0
T35	EX4-EX5B1	2,2	0,1	0,2	1,9	0,2	0,4	0,7	0,3	0,6	0,4	0,2	2,4
T36	EX4-EX5B2	0,5	0,5	0,6	1,8	2,4	1,0	2,9	0,7	0,5	3,4	0,7	1,2
T37*	EX4-LINCEX4-LINCEX5	3,2	6,3	2,8	15,0	11,7	1,3	5,2	5,0	1,6	0,6	3,0	1,5
T38	EX4-LINCEX4-LINCEX4B-LINCEX5	1,4	0,3	0,1	1,1	1,7	0,2	0,3	2,9	0,2	0,4	0,5	0,0

Fig. 5. Heatmap of the observed transcript counts in the 12 cell lines. The read counts for each sample were computed as percentages for the start of exon 2 or exon 4 sequencing. The percentages range from pale blue (low) to red (high). Samples are shown with respect to increasing differentiation status. Red asterisks indicate variations in Lila-1 cells following PMA treatment, whereas blue asterisks indicate variations in the terminal differentiation of plasmocytic cells.

this event. Both were notably expressed, especially T21, in 6 out of the 12 cell lines studied (Fig. 5; Table 1). The emergence of the T20 and T21 mRNAs from the *RHOH* gene, as well as the T06-08, T28 and T29 transcripts, has never been described before. Thus, among the 18 *RHOH* mRNA molecules described above, 11 might be templates for RhoH protein synthesis (T01–05, T22–27), and 7 (T06–07–08, T20–21, T28–29) known exons have never been detected before. Specifically, the T06–07–08 transcript, which exhibited a large split of exon 7, was weakly expressed, while the T20 and T21 genes without any exon 7 sequence were notably expressed, all of which eliminate the possibility of RhoH protein synthesis. Several internal cryptic splice sites were identified, which may lead to the synthesis of 6 transcripts. We observed these sites inside exon 7, leading to its splitting accompanied by the loss of almost the entire RhoH protein coding region (T07). A 5′-sided shorter exon 7 due to a splice exonic acceptor shift (Fig. 6, Table 1) was also observed for HBOC genes[29]. In addition, some of them were found inside exon 4 (T28 and T29), exhibiting a 3′-sided shorter exon 4 due to a splice exonic donor shift in this exon. Moreover, skipping of exon 6 was observed, accompanied by exon 7 splitting and by loss of part or the entire exon 7 sequence (T08, T20, T28–29; Fig. 6, Table 1). These cryptic splice sites are highlighted in Table S3.

To summarize, 18 *RHOH* mRNA molecules containing exons 6 and 7 are hypothesized to be potential templates for RhoH protein synthesis or to be involved in its regulation. These mRNAs included 13 molecules containing exons 2 and 4 and four molecules containing exons 2 or 4, in addition to a single molecule containing a large version of exon 4 alone.

Second, we detected 5 new *RHOH* exons, which led to the generation of 18 new transcripts that did not contain RhoH, which encodes exon 7 or exon 6 (Fig. 4). These new exons located downstream of exon 3 or exon 4 were named EX3B, EX4B1, EX4B2–6, EX4B7 and EX4T1–4. These fragments were alternatively spliced downstream of exon 2 (11 mRNAs, T09–19) or exon 4 (7 mRNAs, T30–36). Two of these 5 new exons were already identified as exons in the SIB database (EX4B1 and EX4B7; Fig. 4), but no transcripts have been described in the literature or in other databases.

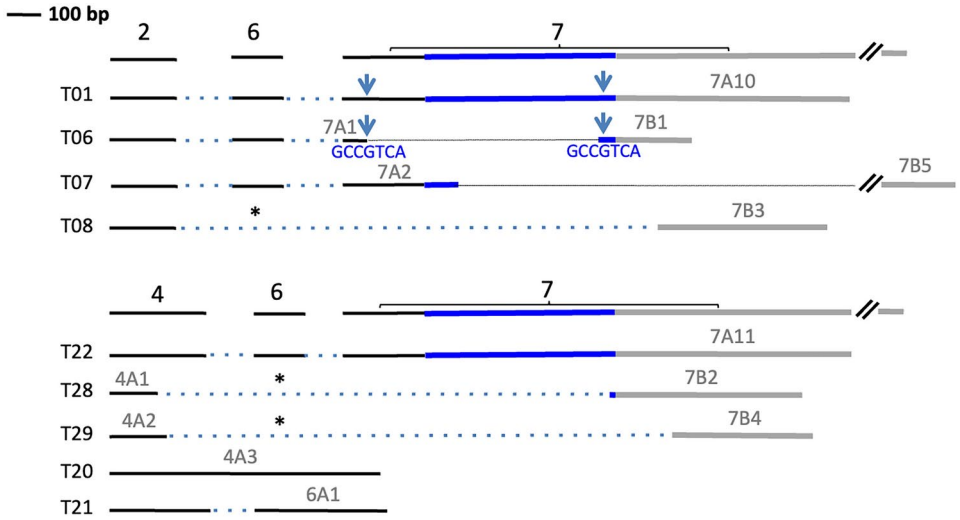

Fig. 6. New *RHOH* mRNAs with known *RHOH* exons. Previously known exons 2, 4, and 6 of RHOH are depicted in black, and exon 7 is depicted in black, blue and gray. Exon 7 contains the RhoH protein ORF (in blue), framed by its 5′-extremity and the 3′-UTR regions (both presented in gray). Exon 6 skipping (in T08, T28 and T29) is indicated by a black asterisk. Shorter 3′-UTRs in exon 7 are named 7B1 to 7B5, and the largest 3′-UTRs are named 7A10 and 7A11 (for T01 and T22, respectively). In T06, exon 7 was split; two identical sequence motifs, "GCCGTCA", frame the deletion.

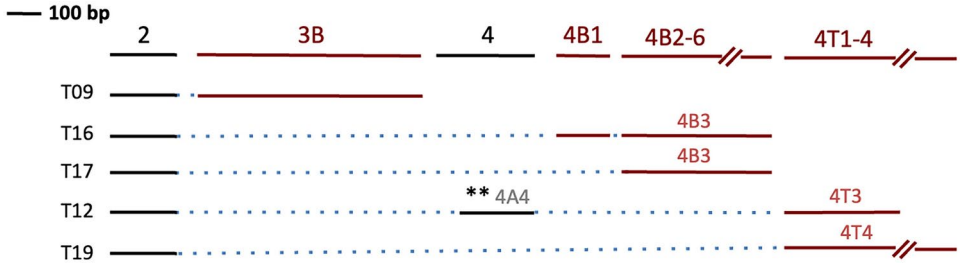

Fig. 7. New *RHOH* mRNAs with new *RHOH* exons and starting from *RHOH* exon 2. Four new exons of RHOH (3B; 4B1; 4B2-6, 4T1-4) are depicted in brown. An example of exon 4 inclusion between exon 2 and exon 4T1-4 (4A4 in T12) is indicated by a double asterisk.

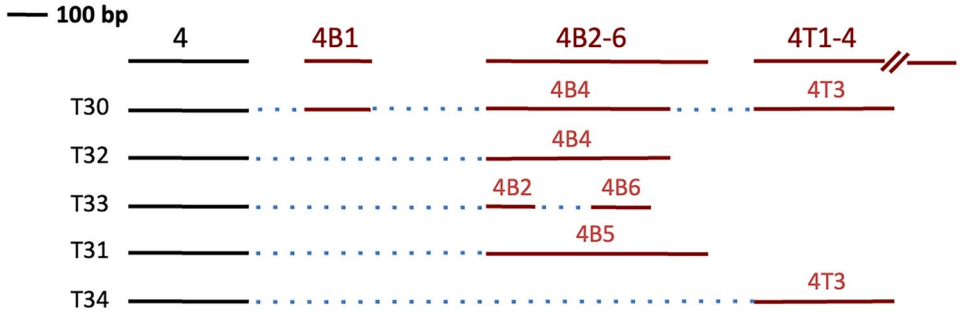

Fig. 8. New *RHOH* mRNAs with new *RHOH* exons and starting from *RHOH* exon 4. In the T33 mRNA, splitting of exon 4B2-6 was accompanied by a deletion of 84 bp, leading to the 4B2 and 4B6 versions.

Among the 11 *RHOH* transcripts with new exons that started from exon 2 (T09–19), T09 contained a new single exon that was never predicted before; we named this gene EX3B, which was spliced to exon 2. Six out of the 10 remaining mRNAs in Fig. 4 contained either the new exon EX4B2–6 alone (T11, T14, T17) or both the new exons EX4B1 and EX4B2–6 (T10, T13, T16), indicating the possible inclusion or skipping of EX4B1. In addition, among these 6 mRNAs, 4 (T10–11–13–14) presented an inclusion of exon 4 immediately downstream

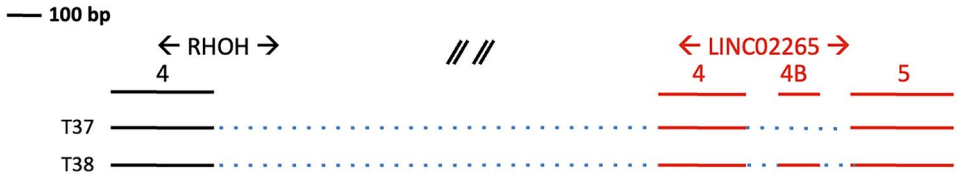

Fig. 9. Diagrams of *RHOH* fusion mRNAs. The T37 and T38 mRNAs are fusion transcripts generated by long-distance splicing between *RHOH* exon 4 and *LINC02265* exons 4 to 5. The *LINC02265* exon, which includes the new exon 4B, is depicted in red.

of exon 2, and exon 4 exhibited a shorter length upon inclusion: EX4A5in T10–11 or EX4A4 in T13–14, in agreement with our previous results[4]. The last 4 mRNAs included either EX4T1-4 alone (T12, T15, T19) or both EX4B7 and EX4T1–4 (T18) and were spliced either to exon 2 alone (T18–19) or to exon 2 directly followed by exon 4 (T12–15) (Figs. 4 and 7; Table 1). As stated in a previous study[30], selection of the 3′ splice site, highlighted by different exon sizes, might be due to internal cryptic splice sites in the EX4T1–4 sequence (Table S3). In addition, the inclusion of exon 4 was observed again (EX4A5 in T12 and EX4A4 in T15), between exon 2 and the new exon EX4T1–4. Importantly, 7 of the 11 new mRNAs, starting from exon 2 and therefore reflecting promoter P2 activity in B cells, were notably expressed in some cell lines (Fig. 5, Table 1).

The 7 following new *RHOH* mRNA molecules, T30 to T36, started from exon 4 and contained 3 of the new *RHOH* exons: EX4B1, EX4B2-6, and EX4T1–4 (Figs. 4, 5 and 8). Only one mRNA contained the new exon EX4T1–4 spliced to exon 4 (EX4T3 in T34; Figs. 5 and 8, Table 1). The T35 and T36 transcripts contained exon 5 and were named EX5B1–2. Remarkably, these cells exhibited two larger sizes, EX5B1 and EX5B2 (Fig. 4, Tables S1 and S3). However, T35–36 were very weakly expressed in all 12 cell lines, similar to T30, T33 and T34 (Fig. 5, Table 1). In contrast, both the T31 and T32 transcripts, containing the new exon EX4B2–6, were notably expressed in almost all 12 cell lines, mostly in the 4 DLBCL mature B cells (U2932, HBL1, OCI-Ly1 and OCI-Ly3; Fig. 5). The percentage of these transcripts was 2- to 5-fold greater in these cells than in other cells, as observed for T01 and T22, 2 transcripts that are likely "good" templates for RhoH protein translation (Table 1, Fig. 5). This finding suggested that this new exon, EX4B2–6, might play an important role in the control of RhoH protein synthesis. Indeed, it seems important to emphasize that these 5 new *RHOH* exons did not associate with the previously identified *RHOH* exons 6–7, which are required for RhoH protein synthesis[4,27].

Finally, the *RHOH* gene was framed by 4 neighboring genes (Fig. S3). We observed 2 fusion transcripts, T37 and T38, likely generated by long-distance splicing. Strikingly, these 2 last *RHOH* mRNAs, starting from *RHOH* exon 4, containing exons from *LINC02265* (*LINC02265* exons 4 and 5 in T37 and *LINC02265* exons 4, 4B and 5 in T38), located approximately 60 kb downstream of the *RHOH* gene, were never described before (Figs. 4 and 9; Table 1). These mRNA molecules may have emerged from tandem chimerism induced by a long-distance splicing mechanism, as previously described[31–34]. Regarding the relative amounts of these particular mRNA molecules in the 12 B-cell lines, only T37 was notably expressed in both Burkitt mature cell lines (15% in Raji and 11.7% in BJAB; Fig. 5). Notably, we discovered exon 4B as a new exon in this long noncoding RNA.

Quantitative analysis of *RHOH* transcript pattern variation during B-cell differentiation

As in traditional RNA-Seq studies, we searched for new transcripts and investigated their abundance between samples (we assumed these results according to the use of only one sample per cell line). The expression of a total of 10 mRNAs varied quantitatively upon differentiation, as observed in Lila-1 cells following PMA treatment (+PMA) or not (−PMA). Figure 5 first illustrates a series of 8 out of these *RHOH* mRNA molecules (highlighted by red asterisks). Previously unknown, they contained either exons 6–7 (in T06 and T08) or new exons, EX3B, EX4B1, EX4B2–6, and EX4T1–4 (in T09, T13–15, T17, and T19), and started all from exon 2, reflecting the B-cell lineage-specific activity of the P2 promoter upstream of this exon (Fig. S1). Notably, T06 and T08, which contained split versions of the major ORF in exon 7 and therefore did not lead to RhoH protein synthesis, presented markedly decreased expression levels upon B-cell differentiation (Table 2). More importantly, 6 of the 8 transcripts contained new *RHOH* exons spliced to exon 2 (T09, T13–15, T17, and T19), indicating their possible function during such a biological phenomenon (Fig. 5; Table 2). In contrast, the expression of most of the mRNAs starting from exon 4 did not vary upon differentiation, except for T20, which contained exon 4 alone and had a greater size and presented a 5.7-fold decrease in expression after PMA treatment. In addition, T37, which contains *LINC02265* exons fused to *RHOH* exon 4, presented a 2.25-fold decrease upon differentiation of Lila-1 cells (Fig. 5; Table 2). Another mRNA molecule, T12, exhibited increased expression upon B-cell terminal differentiation, as indicated by RPMI and U266 plasmocytic cells, compared to pre-B and mature Burkitt cells (Table 2; Fig. 5, highlighted by a blue asterisk).

Potential protein/peptide formation from *RHOH* mRNA molecules

The *RHOH* transcripts might serve as templates for the synthesis of the atypical RhoH protein or may be involved in its regulation. In particular, new transcripts could be templates for the synthesis of new regulatory peptides/proteins. We used the web version of ORFfinder software from NCBI to search for open reading frames (ORFs; Table S4). In this chapter, we will only present potential proteins for which we found homology to known protein families in Pfam.

Description	RNAseq name	New or known	Nanostring study used name or observed modification	Transcript name	Position in transcript (in bold)	Average transcript frequency (%)
New exons in new transcripts	EX3B	New		T09	EX2-EX3B	4.8
	EX4B1	New	EST2	T10	EX2-EX4A5-EX4B1-EX4B4	1.18
	EX4B1	New	EST2	T13	EX2-EX4A4-EX4B1-EX4B4	2
	EX4B1	New	EST2	T16	EX2-EX4B1-EX4B3	2.16
	EX4B1	New	EST2	T30	EX4-EX4B1-EX4B4	1.14
	EX4B3	New	EST3	T16	EX2-EX4B1-EX4B3	2.16
	EX4B3	New	EST3	T17	EX2-EX4B3	4.1
	EX4B4	New	EST3	T10	EX2-EX4A5-EX4B1-EX4B4	1.18
	EX4B4	New	EST3	T11	EX2-EX4A5-EX4B4	6.04
	EX4B4	New	*EST3*	T13	EX2-EX4A4-EX4B1-EX4B4	2
	EX4B4	New	EST3	T14	EX2-EX4A4-EX4B4	8.81
	EX4B4	New	EST3	T30	EX4-EX4B1-EX4B4	1.14
	EX4B4	New	EST3	T32	EX4-EX4B4	13.14
	EX4B5 (long)	New	EST3	T31	EX4-EX4B5	14.04
	EX4B2-B6 (split)	New	EST3	T33	EX4-EX4B2-EX4B6	0.94
	EX4B7	New	EST6	T18	EX2-EX4B7-EX4T1	0.87
	EX4T1	New	Uni3	T18	EX2-EX4B7-EX4T1	0.87
	EX4T2	New	Uni3	T15	EX2-EX4A4-EX4T2	3.38
	EX4T3	New	Uni3	T12	EX2-EX4A5-EX4T3	4.15
	EX4T3	New	Uni3	T34	EX4-EX4T3	2.55
	EX4T4 (long)	New	Uni3	T19	EX2-EX4T4	5.04
	EX5B1	Known (ex5)	EST7	T35	EX4-EX5B1	0.65
	EX5B2	Known (EX5)	EST7	T36	EX4-EX5B2	1.36
Known exons in new transcripts	EX4A3	EX4	EX4 long	T20	EX4A3	6.04
	EX6A1	EX6	EX6 long	T21	EX4-EX6A1	8.74
	EX4A1	EX4	EX4 short	T28	EX4A1-EX7B2	0.97
	EX4A2	EX4	EX4 short	T29	EX4A2-EX7B4	0.09
	EX7A1-EX7B1	EX7	EX7 splitting	T06	EX2-EX6-EX7A1-EX7B1	2.08
	EX7A2-EX7B5	EX7	EX7 splitting	T07	EX2-EX6-EX7A2-EX7B5	1.66
	EX7B3	EX7	EX7 splitting	T08	EX2-EX7B3	2.97
	EX7B2	EX7	EX7 splitting	T28	EX4A1-EX7B2	0.97
	EX7B4	EX7	EX7 splitting	T29	EX4A2-EX7B4	0.09
	EX6	EX6	EX6 skipping	T28	EX4A1-EX7B2	0.97
	EX6	EX6	EX6 skipping	T29	EX4A2-EX7B4	0.09
	EX6	EX6	EX6 skipping	T08	EX2-EX7B3	2.97
Long-distance splicing	LINCEX4-LINCEX5	New	Genes fusion	T37	EX4-LINCEX4-LINCEX5	5.06
	LINCEX4B	New	Genes fusion	T38	EX4-LINCEX4-LINCEX4B-LINCEX5	0.83

Table 1. New *RHOH* transcript patterns identified by mRNA-Seq. The 5 new *RHOH* exons that were identified in new transcript patterns are listed, as are the previously identified EX5 exons, with the details of each transcript structure and their relative abundance in the different cell lines. Already known *RHOH* exons exhibiting size/sequence modifications and included in seven new *RHOH* transcripts are listed, with detailed structures of these new mRNAs. One of these mutations started from exon 2 (T06-07–08), and the other started from exon 4 (T20-21; T28-29). All observed modifications for each of these exons are listed. Notably, EX2 is never modified in this pattern but rather is modified mainly by EX4, EX7 and EX6. The two last mRNAs, T37 and T38, emerged from the gene fusion of *RHOH* and *LINC02265* resulting from long-distance splicing of *RHOH* exon 4 to *LINC02265* exons 4 and 5 (T37) or exons 4, 4B, and 5 (T38). All average transcript frequencies were calculated without taking into account the HL60 control cell values. Significant values are in italics.

First, the RhoH protein is predicted to be normal in size (21.3 kDa) or to harbor 2 possible N-terminal extensions, 29 or 39AA, leading to 24.4 or 25.6 kDa proteins, the first of which is likely initiated by the previously described noncanonical initiation codon TTG ([35,36]; Table S4). All 3 isoforms are predicted to be translated from transcripts T01–05 or T22–27, respectively, initiated by exon 2 or exon 4, and containing both exons 6–7 entirely spliced to each of them.

Comparisons	Transcript name/cell lines	without PMA	with PMA	Average expression values	Fold change
Lila-1 cells following PMA treatment	T06	13.2	0.4	–	−32
	T08	9.6	0.1	–	−96
	T09	0.7	19.4	–	+27.7
	T13	3.6	0.2	–	−18
	T14	0.3	13.4	–	+44.7
	T15	2.3	0.2	–	−11.5
	T17	8.0	0.6	–	−13.3
	T19	0.4	1.9	–	+4.7
	T20	8.0	1.4	–	−5.7
	T37	6.3	2.8	–	−2.25
T12 transcript in B-cell lineage sub-types	Plasma cells (RPMI; U266)	–	–	9.7	1
	Pre-B (Nalm-6; Lila-1)	–	–	0.2	−48.5
	B mature Burkitt (BJAB; HBL1)	–	–	0.2	−48.5
	B mature DLBCL (Ly3; Ly1; U2932)	–	–	2	−4.8

Table 2. New *RHOH* transcript levels as a function of differentiation in the Lila-1 cell line following PMA treatment. Transcript level (%) differences of 10 transcripts in Lila-1 cells following PMA treatment or not, and the resulting fold changes ranged from −96 to +44.7. Downregulation of the T12 transcript was observed in most of the tested B-cell lines beginning at earlier differentiation stages compared to that in plasma cells.

In the second group of potential proteins, we found homology between one ORF from T07 and the Ras family in the Pfam database. Notably, this ORF comprises the 39AA region of the N-terminal extension followed by the 36AA "classic" RHOH N-terminus fused to the p4 C-terminal region (T07; Table S4).

The peptides possibly arising from transcripts containing the new *RHOH* exons are also listed in Table S4. First, transcript T09, which contains the rather large EX3B gene, presented 2 possible ORFs, p5- and p6-ORF, starting inside EX3B or EX2, respectively, that generated 4 peptides. p5 proteins are homologous to T2SSM (Type II secretion system, protein M; Table S4). Importantly, the T09 mRNA was notably expressed in some cell lines and exhibited a 27.7-fold increase in expression when Lila-1 cells differentiated (Table 2).

Finally, fusion of the T37 and T38 transcripts containing *RHOH* exon 4 fused to the *LINC02265* exons allowed the synthesis of 5 peptides from the p16, p17, and p18 ORFs (Table S4). Both peptides from p17-ORF contained a D-ribitol-5-phospho-cytidylyltransferase motif, and both emerged from p18-ORF as a restriction alleviation protein LAR motif (Table S4); these motifs are related to bacterial metabolism. As these ORFs start inside *LINC02265* exons, they do not contain *RHOH* sequences (Table S4). However, because of the fusion of both genes, peptide synthesis is governed by the *RHOH* P3 promoter.

On the basis of our analysis, several known noncanonical start codons, such as TTG, CTG and ATC, were used (in blue; Table S4). For many other ORF predictions, non-ATG start codons, not described before, were used (AGA, GAG, GGA, CAG, ACA, TGG, CTT, GCT, CCC, TTT; Table S4, in bold). To date, it is not known whether such non-ATG translation initiation codons, which have not yet been documented, may be "true" start codons[35–37].

Finally, we computed our sequenced data compared to genomic sequences and compiled exon-introns splice sites (Supplementary Table S5).

Discussion

Annotation mistakes

We detected a mistake in the EMBL numbering of one sequence. Our publication refers to the sequence of the first described *RHOH* mRNA[4], named this mRNA Z35225, whereas the number in the EMBL database is Z35227 (in the EMBL database, Z35225 refers falsely to a bacterial DNA for ribosomal spacer sequence; see Table S1).

New insights

The RhoH protein is a hemato-specific atypical member of the Rho small GTP-binding subfamily. It is always GTP bound, leading to permanent activation[5]. Therefore, the synthesis of the RhoH protein may be controlled by subtle specific mechanisms upstream and downstream of RNA synthesis. Our present work provides new insights into the properties of the *RHOH* gene complex. We first identified 7 new mRNA molecules with *RHOH* exons 6 and/or 7 that presented exon 6 skipping (T06–08, T28–29) with exon 7 splitting or skipping (T20–21) and might regulate RhoH protein synthesis. Indeed, RhoH might be negatively regulated by decreasing the amount of encoding mRNAs such as T01–05 and T22–27 and/or by increasing the amount of the 7 so-called "RhoH noncoding mRNAs" (Figs. 4, 5, 6, 7, 8, 9). Second, our method allowed us to identify a series of 10 potential new *RHOH* exons, 5 of which were included in 18 new mRNA molecules. None of these new exons were combined with exons 6–7, which are mRNA templates used for the synthesis of the RhoH protein (Figs. 4, 7, and 8). Therefore, we hypothesized that such new *RHOH* mRNA molecules containing new exons could be used for RhoH protein synthesis control, particularly during biological phenomena such as B-cell differentiation (Fig. 5; Table 2). Furthermore, T31 and T32 were notably expressed in almost all B-cell lines and contained the

new exon EX4B2-6, which was directly spliced to EX4. Additionally, transcript T14, which contained the same new exon, was notably expressed and increased by 44.7-fold upon differentiation (Fig. 5; Table 2). This finding is in agreement with possible regulatory functions and highlights the importance of this new exon, which captures EX4 and/or EX2 to create unproductive mRNAs not encoding the RhoH protein.

More elements

Analysis of the sequences led us to identify a possible regulatory motif of 7 bp (GCCGTCA) at the 5′ end and inside *RHOH* exon 7 at positions chr4:40244873–40244879 and chr4:40245514–40245520 in the hg19 genome assembly. We analyzed this sequence by the MEME and TOMTOM algorithms (https://meme-suite.org) and found it to be a potential transcriptional element binding motif[38]. Interestingly, this sequence has been described as one cAMP response element involved in the binding of cAMP response element binding protein (CREB). Notably, CREB stimulates gene expression, cell survival and migration, especially during tumorigenesis[39,40]. Moreover, multialignment analysis via the UCSC website led to high conservation of this sequence at both positions near *RHOH* exon 7 in 48 vertebrate genomes. This last piece of information highlights the importance of this motif with respect to the proximity of the splice sites we observed within T06 and the 2 CRE sites (at the 3′ end of EX7A1 and 5′ end of EX7B1; Fig. 6). T06 expression could therefore stop CREB binding by the loss of these potential binding sites and decrease RhoH protein synthesis, especially in Lila-1 cells, where T06 is highly expressed (Fig. 5).

In addition to this first motif, additional analysis led us to highlight in EX4B2-6 one sequence (GTTTTGT ATTTTTTTGTAG**AGA;** Table S3) that could have a function in mRNA secondary structure modeling, which is needed for posttranscriptional and translational control mechanisms, such as internal translation initiation facilitation[41]. The RNA structure has critical roles in processes such as the regulation of translation and splicing[42]; these results could therefore provide insight into this process.

New exons and differentiation

The expression of ten mRNAs, six of which contained new exons, varied upon differentiation in Lila-1 ± PMA (Table 2). These new exons (Figs. 4, 5) are alternatively spliced to exon 2 alone or to exons 2 and 4, leading to new *RHOH* transcript patterns never described before (T09, T13–15, T17, T19). These new transcripts presented notable variations following PMA treatment (Table 2). This might suggest a possible function of these new exons in regulating the RhoH protein level during biological events such as B-cell differentiation. Notably, exons 6–7, which are clearly required for RhoH protein translation, were absent from these new transcript patterns.

LincRNA fusion

Long noncoding RNAs (lincRNAs) are RNA types that are generally defined as transcripts of more than 200 nucleotides that are not translated into proteins. There are multiple functions of these genes in cells; they act in the regulation of gene transcription, regulate the basal transcription machinery, and are involved in posttranscriptional and epigenetic regulation, most likely in the regulation of DNA replication timing and chromosome stability[43,44]. However, this landscape of functions still seems incomplete since we found that the long noncoding RNA *LINC02265* fused to *RHOH* exon 4, generating 2 mRNA molecules never described before. To date, many fusions *involving LincRNAs* have been described across various cancers, but all of these fusions are the result of DNA damage[45]; we are the first to describe the fusion of such *LincRNAs* to their upstream neighboring genes. These mRNA molecules might constitute a new means of gene expression regulation by cis-acting long noncoding RNAs, in addition to what has been described in the literature[46]. We therefore speculate that such fusion of *LINC02265* exons to *RHOH* exon 4 might play a role in the regulation of RhoH protein synthesis by sequestering *RHOH* exon 4. This exon would then not be used as a template for RhoH protein synthesis. Alternatively, this fusion could be involved in interactions with other genes.

Nested splicing events

With respect to the P3 promoter being next to *RHOH* exon 4, which had the highest activity ([29]; Fig. S1), and to new exons downstream of P3 and exon 4, we discovered new exons. For this section, we decided to focus on splicing events in this region. Sequencing analysis confirmed that exon 4 can be spliced into exon 2-initiated transcripts (Figs. 4, 5). Due to the use of internal cryptic splice sites in addition to its 253 bp leader exon size, EX4 can exhibit 5 different sizes (EX4A1-5 of 149, 184, 798, 201 and 180 bp). EX4B2-6 also exhibited 5 different sizes 84, 433, 435, 574 and 267 bp whereas EX4T1-4 exhibited 4 different sizes 332, 337, 341 and 900 bp (Tables S1 and S3). As stated in previous studies, the regulation of the choice of the 3′ splice site, highlighted by several exon sizes differing from the 3′ side, might be due to internal cryptic splice sites regulated by elements that could control intrasplicing events to produce proteins with functional diversity[30]. Regulation of 3′ splice site choice could also arise from a hypothetical intron retention phenomenon. This last mechanism involves the overuse of the transcriptional machinery to produce unproductive mRNAs further degraded by the nonsense mRNA decay (NMD) process or by microRNA-induced cleavage[28]. These new mRNAs were then detected because they were not degraded by NMD and were able to be reliably detected. Taken together, these findings suggest that these biological engineering strategies could thereby participate in the decrease or loss of RhoH protein production. Our study revealed the extraordinary complexity of AS regulation in humans, which may be associated with the differentiation status of the targeted cell lines.

Splicing-polyA sites

In our perseverance in the discovery of new elements arising from the *RHOH* full-length RNA-Seq, we observed at the 3′ extremity of the gene that some of the 7′ ends of the exons mapped close to polyadenylation sites (Fig. S4). As shown in the picture extracted from the UCSC genome browser (hg19) with the optional track

name "reported poly(A) sites from polyA database"[47], 4 out of the 5 sites for this gene were mapped at the 3′ border of some of the exon 7 configurations we described (Hs.160673.1.15 for EX7A3, Hs.160673.1.16 for EX7A5, Hs.160673.17 differentiating EX7A8 from EX7A9, and finally Hs.160673.1.18 to distinguish EX7B3 and EX7B4 from EX7A10 and EX7A11). Previous studies have indicated that splicing and polyadenylation are two inseparable parts of one consolidated pre-mRNA processing machinery, leading to the conjecture that cotranscriptional splicing is a natural mechanism for suppressing premature transcription[48]. Recent genome-wide protein–RNA interaction studies have significantly reshaped the understanding of the role of mRNA 3′ end formation factors in RNA biology[49]. With the evolution of the understanding of the roles of 3′ end formation factors, the final picture might be more complex than originally supposed. The abundance of possible 3′ ends of the *RHOH* gene transcripts could increase the list of functions of the gene in biological pathways.

Potential peptides

Sequencing analysis led us to identify 38 single transcripts that were further analyzed via ORFfinder to generate open reading frames (ORFs) with potential translation into peptides/proteins. Although the normal RhoH protein was predicted in our analysis (with 3 lengths), many additional ORFs of many sizes were predicted (Table S4). When expressed, some of these peptides could induce a regulatory process to limit the synthesis of the RhoH protein, as their mRNA templates are notably expressed in almost all cell lines (Table S4). Such synthesis could be initiated by noncanonical codons, as suggested here (Table S4), or by human genes[35–37]. Using ORF prediction tools with new findings on the *RHOH* gene, the construction of the ORFome of this gene seems straightforward, as does the construction of the whole genome.

Outlook

Full-length mRNA sequencing of the *RHOH* gene allowed us to identify new exons and new transcripts to show quantitative discrimination of these transcripts in B-cell lines across differentiations and, more promisingly, new fusion transcripts involving the *RHOH* and *LINC02265* genes. Interestingly, both genes seem to share some regulatory elements, as shown by GeneCards via the use of the GeneHancer tool (https://www.genecards.org[49]). The link between the *RHOH-LINC02265* fusion transcripts needs to be further investigated.

Conclusion

Targeted RNA-Seq analysis of the *RHOH* gene led to the discovery of more genes than we knew about this gene in human B cells. As previously proposed, the occurrence of new exons, new splice sites, new transcripts and fusion transcripts may be linked to genome evolution in mammals and could play a major role in gene expression. In this context, increasing the variety and number of exons could lead to increased expression of the corresponding gene[50]. Using the powerful capacity of full-length mRNA sequencing, we refined the *RHOH* gene structure (through intron–exon succession) in humans and suggested new potential mechanisms for RhoH protein synthesis regulation and possible implications for controlling the regulation of differentiation in B cells. In addition, the drastic filters applied in the RNA-Seq analysis workflow allowed us to detect real mRNAs, which are known and well described in the literature, as well as new mRNAs. Finally, the present results enrich our previous work[4]. At the scale of a single gene, all those events (new exons, new splice sites, new transcripts, transcript fusions and, of course, new possible ORFs) are of primary importance. It seems clear that further investigating the involved mechanisms in the landscape of the whole genome could lead to a better understanding of normal and disease biology. Overall, our study emphasizes the importance of full-length mRNA sequencing and Pandora's box opening in understanding gene regulation.

Data availability

All raw sequencing data are available under the SRA database from NCBI website (https://www.ncbi.nlm.nih.gov/sra) with the PRJNA690664 bioproject accession number.

Received: 19 March 2024; Accepted: 7 November 2024
Published online: 16 November 2024

References

1. Dallery, E. et al. TTF, a gene encoding a novel small G protein, fuses to the lymphoma-associated LAZ3 gene by t(3;4) chromosomal translocation. *Oncogene.* **10**(11), 2171–2178 (1995).
2. Preudhomme, C. et al. Nonrandom 4p13 rearrangements of the RhoH/TTF gene, encoding a GTP-binding protein, in non-Hodgkin's lymphoma and multiple myeloma. *Oncogene.* **19**(16), 2023–2032. https://doi.org/10.1038/sj.onc.1203521 (2000).
3. Dallery-Prudhomme, E. et al. Genomic structure and assignment of the RhoH/TTF small GTPase gene (ARHH) to 4p13 by in situ hybridization. *Genomics.* **43**(1), 89–94. https://doi.org/10.1006/geno.1997.4788 (1997).
4. Lahousse, S. et al. Structural features of hematopoiesis-specific RhoH/ARHH gene: High diversity of 5'-UTR in different hematopoietic lineages suggests a complex posttranscriptional regulation. *Gene.* **343**(1), 55–68. https://doi.org/10.1016/j.gene.2004.08.022 (2004).
5. Li, X. et al. The hematopoiesis-specific GTP-binding protein RhoH is GTPase deficient and modulates activities of other Rho GTPases by an inhibitory function. *Mol. Cell Biol.* **22**(4), 1158–1171. https://doi.org/10.1128/MCB.22.4.1158-1171.2002 (2002).
6. Cherry, L. K., Li, X., Schwab, P., Lim, B. & Klickstein, L. B. RhoH is required to maintain the integrin LFA-1 in a nonadhesive state on lymphocytes. *Nat. Immunol.* **5**(9), 961–967. https://doi.org/10.1038/ni1103 (2004).
7. Baker, C. M. et al. Opposing roles for RhoH GTPase during T-cell migration and activation. *Proc. Natl. Acad. Sci. USA.* **109**(26), 10474–10479. https://doi.org/10.1073/pnas.1114214109 (2012).
8. Oda, H., Tamehiro, N., Patrick, M. S., Hayakawa, K. & Suzuki, H. Differential requirement for RhoH in development of TCRαβ CD8αα IELs and other types of T cells. *Immunol. Lett.* **151**(1–2), 1–9. https://doi.org/10.1016/j.imlet.2013.02.007 (2013).
9. Mino, A. et al. RhoH participates in a multiprotein complex with the zinc finger protein kaiso that regulates both cytoskeletal structures and chemokine-induced T cells. *Small GTPases.* **9**(3), 260–273 (2018).

10. Dorn, T. et al. RhoH is important for positive thymocyte selection and T-cell receptor signaling. *Blood.* **109**(6), 2346–2355. https://doi.org/10.1182/blood-2006-04-019034 (2007).
11. Chae, H. D., Siefring, J. E., Hildeman, D. A., Gu, Y. & Williams, D. A. RhoH regulates subcellular localization of ZAP-70 and Lck in T-cell receptor signaling. *PLoS One.* **5**(11), e13970. https://doi.org/10.1371/journal.pone.0013970 (2010).
12. Tybulewicz, V. L. & Henderson, R. B. Rho family GTPases and their regulators in lymphocytes. *Nat. Rev. Immunol.* **9**(9), 630–644. https://doi.org/10.1038/nri2606 (2009).
13. Gu, Y., Jasti, A. C., Jansen, M. & Siefring, J. E. RhoH, a hematopoietic-specific Rho GTPase, regulates proliferation, survival, migration, and engraftment of hematopoietic progenitor cells. *Blood.* **105**(4), 1467–1475. https://doi.org/10.1182/blood-2004-04-1604 (2005).
14. Chae, H. D., Lee, K. E., Williams, D. A. & Gu, Y. Cross-talk between RhoH and Rac1 in regulation of actin cytoskeleton and chemotaxis of hematopoietic progenitor cells. *Blood.* **111**(5), 2597–2605. https://doi.org/10.1182/blood-2007-06-093237 (2008).
15. Daryadel, A. et al. RhoH/TTF negatively regulates leukotriene production in neutrophils. *J. Immunol.* **182**(10), 6527–6532. https://doi.org/10.4049/jimmunol.0803846 (2009).
16. Stoeckle, C. et al. RhoH is a negative regulator of eosinophilopoiesis. *Cell Death Differ.* **23**(12), 1961–1972. https://doi.org/10.1038/cdd.2016.73 (2016).
17. Galiègue-Zouitina, S., Delestré, L., DuPont, C., Troussard, X. & Shelley, C. S. Underexpression of RhoH in hairy cell leukemia. *Cancer Res.* **68**(12), 4531–4540. https://doi.org/10.1158/0008-5472.CAN-07-5661 (2008).
18. Sanchez-Aguilera, A. et al. Involvement of RhoH GTPase in the development of B-cell chronic lymphocytic leukemia. *Leukemia.* **24**(1), 97–104. https://doi.org/10.1038/leu.2009.217 (2010).
19. Troeger, A. et al. RhoH is critical for cell-microenvironment interactions in chronic lymphocytic leukemia in mice and humans. *Blood.* **119**(20), 4708–4718. https://doi.org/10.1182/blood-2011-12-395939 (2012).
20. Pasqualucci, L. et al. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature.* **412**(6844), 341–346. https://doi.org/10.1038/35085588 (2001).
21. Fueller, F. & Kubatzky, K. F. The small GTPase RhoH is an atypical regulator of hematopoietic cells. *Cell Commun. Signal.* **6**, 6. https://doi.org/10.1186/1478-811X-6-6 (2008).
22. Troeger, A., Chae, H. D., Senturk, M., Wood, J. & Williams, D. A. A unique carboxyl-terminal insert domain in the hematopoietic specific, GTPase-deficient Rho GTPase RhoH regulates posttranslational processing. *J. Biol. Chem.* **288**(51), 36451–36462. https://doi.org/10.1074/jbc.M113.505727 (2013).
23. Voena, C. & Chiarle, R. RHO family GTPases in the biology of lymphoma. *Cells.* **8**(7), 646. https://doi.org/10.3390/cells8070646 (2019).
24. Galiègue-Zouitina, S. et al. Bimodal expression of RHOH during myelomonocytic differentiation: Implications for the expansion of AML differentiation therapy. *EJHaem.* **2**(2), 196–210. https://doi.org/10.1002/jha2.128 (2021).
25. Ahmad Mokhtar, A. M., Hashim, I. F., Mohd Zaini Makhtar, M., Salikin, N. H. & Amin-Nordin, S. The role of RhoH in TCR signaling and its involvement in diseases. *Cells.* **10**(4), 950. https://doi.org/10.3390/cells10040950 (2021).
26. Mouly, L. et al. The RND1 small GTPase: Main functions and emerging role in oncogenesis. *Int. J. Mol. Sci.* **20**(15), 3612. https://doi.org/10.3390/ijms20153612 (2019).
27. Delestré, L. et al. Repression of the RHOH gene by JunD. *Biochem. J.* **437**(1), 75–88. https://doi.org/10.1042/BJ20100829 (2011).
28. Monteuuis, G., Wong, J. J. L., Bailey, C. G., Schmitz, U. & Rasko, J. E. J. The changing paradigm of intron retention: Regulation, ramifications and recipes. *Nucleic Acids Res.* **47**(22), 11497–11513. https://doi.org/10.1093/nar/gkz1068 (2019).
29. Davy, G. et al. Detecting splicing patterns in genes involved in hereditary breast and ovarian cancer. *Eur. J. Hum. Genet.* **25**(10), 1147–1154. https://doi.org/10.1038/ejhg.2017.116 (2017).
30. Parra, M.K., Gallagher, T.L., Amacher, S.L., Mohandas, N., Conboy, J.G. (2012) Deep intron elements mediate nested splicing events at consecutive AG dinucleotides to regulate alternative 3' splice site choice in vertebrate 4.1 genes. *Mol. Cell Biol.* **32**(11), 2044–53. https://doi.org/10.1128/MCB.05716-11
31. Anderson, A. M. & Staley, J. P. Long-distance splicing. *Proc. Natl. Acad. Sci. USA.* **105**(19), 6793–6794. https://doi.org/10.1073/pnas.0803068105 (2008).
32. Parra, G. et al. Tandem chimerism as a means to increase protein complexity in the human genome. *Genome Res.* **16**(1), 37–44. https://doi.org/10.1101/gr.4145906 (2006).
33. Hoff, A. M. et al. Novel RNA variants in colorectal cancers. *Oncotarget.* **6**(34), 36587–36602. https://doi.org/10.18632/oncotarget.5500 (2015).
34. Hoff, A. M. et al. Identification of novel fusion genes in testicular germ cell tumors. *Cancer Res.* **76**(1), 108–116. https://doi.org/10.1158/0008-5472.CAN-15-1790 (2016).
35. Cao, X., Slavoff, S.A. (2020) Non-AUG start codons: Expanding and regulating the small and alternative ORFeome. *Exp. Cell Res.* **391**(1), 111973. https://doi.org/10.1016/j.yexcr.2020.111973
36. Kearse, M. G. & Wilusz, J. E. Non-AUG translation: A new start for protein synthesis in eukaryotes. *Genes Dev.* **31**(17), 1717–1731. https://doi.org/10.1101/gad.305250.117 (2017).
37. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science.* **324**(5924), 218–223. https://doi.org/10.1126/science.1168978 (2009).
38. Vuilleumier, R. et al. Control of Dpp morphogen signaling by a secreted feedback regulator. *Nat. Cell Biol.* **12**(6), 611–617. https://doi.org/10.1038/ncb2064 (2010).
39. Mizumoto, N. et al. Differential activation profiles of multiple transcription factors during dendritic cell maturation. *J. Invest. Dermatol.* **124**(4), 718–724. https://doi.org/10.1111/j.0022-202X.2005.23616.x (2005).
40. Zhang, H., Kong, Q., Wang, J., Jiang, Y. & Hua, H. Complex roles of cAMP-PKA-CREB signaling in cancer. *Exp. Hematol. Oncol.* **9**(1), 32. https://doi.org/10.1186/s40164-020-00191-1 (2020).
41. Weingarten-Gabbay, S., Elias-Kirma, S., Nir, R., Gritsenko, A.A., Stern-Ginossar, N., Yakhini, Z., Weinberger, A., Segal, E. (2016) Comparative genetics. Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science.* **351**(6270), aad4939. https://doi.org/10.1126/science.aad4939
42. Ding, Y. et al. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature.* **505**(7485), 696–700. https://doi.org/10.1038/nature12756 (2014).
43. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long noncoding RNAs: Insights into functions. *Nat. Rev. Genet.* **10**(3), 155–159. https://doi.org/10.1038/nrg2521 (2009).
44. Statello, L., Guo, C. J., Chen, L. L. & Huarte, M. Gene regulation by long noncoding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**(2), 96–118. https://doi.org/10.1038/s41580-020-00315-9 (2021).
45. Guo, M. et al. The landscape of long noncoding RNA-involved and tumor-specific fusions across various cancers. *Nucleic Acids Res.* **48**(22), 12618–12631. https://doi.org/10.1093/nar/gkaa1119 (2020).
46. Gil, N. & Ulitsky, I. Regulation of gene expression by cis-acting long noncoding RNAs. *Nat. Rev. Genet.* **21**(2), 102–117. https://doi.org/10.1038/s41576-019-0184-5 (2020).
47. Zhang, H., Hu, J., Recce, M. & Tian, B. PolyA_DB: A database for mammalian mRNA polyadenylation. *Nucleic Acids Res.* **33**(Database issue), D116–D120. https://doi.org/10.1093/nar/gki055 (2005).
48. Vlasenok, M., Margasyuk, S. & Pervouchine, D. D. Transcriptome sequencing suggests that pre-mRNA splicing counteracts widespread intronic cleavage and polyadenylation. *NAR Genom. Bioinform.* **5**(2), lqad01. https://doi.org/10.1093/nargab/lqad051 (2023).

49. Fishilevich, S. et al. GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford).* **2017**, bax028. https://doi.org/10.1093/database/bax028 (2017).
50. Fiszbein, A., Krick, K. S., Begg, B. E. & Burge, C. B. Exon-mediated activation of transcription starts. *Cell.* **179**(7), 1551-1565.e17. https://doi.org/10.1016/j.cell.2019.11.002 (2019).

Acknowledgements
We would like to dedicate this work to our regretted scientist, Jean-Pierre Kerckaert also affectionately nicknamed JPK, who significantly participated in the elaboration of the RhoH project, and kept always very supportive and helpful. We also thank the IRCL (Institut pour la Recherche sur le Cancer de Lille) for kind help and assistance. We are grateful to Emilie Floquet and Nathalie Jouy (US 41—UAR 2014—PLBS) for flow cytometry analysis of the purity of T and B lymphocyte preparations and CD19 expression analysis in Lila-1 cells following PMA treatment.

Author contributions
Conceptualization, F.L. and S.G.Z.; methodology, F.L., M.F. and S.G.Z.; software, F.L. and J-P.M; formal analysis, F.L. and S.G.Z.; investigation, L.D. and C.V.; resources, S.G.Z; data curation, F.L. and J-P.M.; writing-original draft preparation, F.L. and S.G.Z.; writing-review and editing, F.L. and S.G.Z.; visualization, F.L. and S.G.Z.; supervision, F.L. and S.G.Z.; project administration, F.L. and S.G.Z.; funding acquisition, M.F. All authors have read and agreed to the last version of the manuscript.

Funding
University of Lille (to F.L., J-P.M., C.V. and M.F.); Inserm and CHU of Lille (to B.Q.); Funding for open access charge: University of Lille; Ligue Nationale Française contre le Cancer, comité du Pas de Calais (to S.G.Z.); Inserm (to S.G.Z.).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-79307-0.

Correspondence and requests for materials should be addressed to F.L. or S.G.-Z.

Reprints and permissions information is available at www.nature.com/reprints.

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