The missing piece

Winkle, Melanie

CHAPTER 3

Long noncoding RNA expression profiling in normal B-cell subsets and Hodgkin lymphoma reveals Hodgkin and Reed-Sternberg cell-specific long noncoding RNAs

Mina Tayari†*, Melanie Winkle†*, Gertrud Kortman†, Jantine Sietzema†, Debora de Jong†, Martijn Terpstra††, Pieter Mestdagh#, Frans Kroese§, Lydia Visser†, Klaas Kok††, Anke van den Berg†, Joost Kluiver†

Department of †Pathology and Medical Biology, ††Genetics, §Rheumatology and Clinical Immunology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. #Center for Medical Genetics, Ghent University, Ghent, Belgium.

*contributed equally to this work

The American Journal of Pathology September 2016
(doi: 10.1016/j.ajpath.2016.05.011)

Abstract

Hodgkin lymphoma (HL) is a malignancy of germinal center (GC) B-cell origin. To explore the role of long noncoding RNAs (lncRNAs) in HL, we studied lncRNA expression patterns in normal B-cell subsets, HL cell lines, and tissues. Naïve and memory B cells showed a highly similar lncRNA expression pattern, distinct from GC-B cells. Significant differential expression between HL and normal GC-B cells was observed for 475 lncRNA loci. For two validated lncRNAs, an enhanced expression was observed in HL, diffuse large B-cell lymphoma, and lymphoblastoid cell lines. For a third lncRNA, increased expression levels were observed in HL and part of Burkitt lymphoma cell lines. RNA fluorescence in situ hybridization on primary HL tissues revealed a tumor cell-specific expression pattern for all three lncRNAs. A potential *cis*-regulatory role was observed for 107 differentially expressed lncRNA-mRNA pairs localizing within a 60-kb region. Consistent with a *cis*-acting role, we showed a preferential nuclear localization for two selected candidates. Thus, we showed dynamic lncRNA expression changes during the transit of normal B cells through the GC reaction and widely deregulated lncRNA expression patterns in HL. Three lncRNAs showed a tumor cell-specific expression pattern in HL tissues and might therefore be of value as a biomarker.

1 Introduction

Hodgkin lymphoma (HL) is a B-cell neoplasm characterized by a minority of neoplastic cells within an extensive inflammatory background. The incidence of HL is about three per 100,000 per year in Western countries and it is most common in adolescents and young adults[1]. HL has been categorized into two disease entities, classical Hodgkin lymphoma (cHL) and nodular lymphocyte predominant Hodgkin lymphoma (NLPHL). cHL accounts for approximately 95% of all cases while NLPHL is less common. The neoplastic cells of cHL, i.e. Hodgkin and Reed-Sternberg (HRS) cells, originate from germinal center (GC) B cells. The HRS cells show a loss of B-cell phenotype with no or strongly reduced expression levels of the B-cell receptor, common B-cell markers and transcription factors[2, 3]. In contrast, the neoplastic cells of NLPHL, i.e. the lymphocyte predominant (LP) cells, have retained their B-cell phenotype[4]. HRS cells can actively influence their microenvironment through the production of various cytokines and a variety of cell surface receptors. Among the factors and pathogens that contribute to cHL pathogenesis are NF-kB and Epstein-Barr virus (EBV)[5]. NF-κB signaling efficiently activates transcription of a variety of genes, which are involved in survival, proliferation and inflammation[6].

The impact of non-coding (nc)RNAs on hematological malignancies including HL, has been well described for microRNAs (miRNAs)[7]. Next to miRNAs another class of ncRNAs, i.e. the long (l)ncRNAs, have recently received a lot of attention as important regulators of gene expression[8]. LncRNAs are defined as transcripts >200 nucleotides in length that lack protein-coding potential. They are mostly categorized based on their genomic location and orientation compared to protein-coding genes, e.g. sense, antisense, intronic or intergenic[9]. Previously, a subset of lncRNAs has been shown to either positively[10] or negatively[11] regulate neighboring protein-coding genes (*cis*-acting lncRNAs)[12]. For instance, some nuclear lncRNAs were shown to function as transcriptional regulators of protein-coding genes in *cis*. This mechanism involves three-dimensional folding of chromatin juxtaposing regulatory sequences located several kilobases apart into close spatial proximity (reviewed by[12]). There is increasing evidence that alterations in the expression levels of lncRNAs are linked to tumorigenesis (reviewed by[13, 14]). It is currently unknown to what extent lncRNAs are regulated in B cells during the GC reaction and possibly deregulated in HL.

In this study, we generated lncRNA expression profiles of normal mature B-cell subsets and HL cell lines. RNA fluorescence *in situ* hybridization (FISH) was used to confirm expression in tumor cells of primary cHL tissues. Finally, we identified putative *cis*-regulatory lncRNAs based on differential expression patterns of nearby protein-coding genes.

2 Material and Methods

2.1 Tissue samples and cell lines

Five FFPE cHL tissue samples were randomly selected from the pathology files of the UMCG tissue bank. All five cHL cases were young adults (20-30 years), four were of the nodular sclerosis subtype and one was of the mixed cellularity subtype according to WHO classification[15]. EBER in situ hybridization revealed that one of the NS cases was EBV+. Normal B-cell subsets were sorted by FACS from three different tonsil samples as described previously[16]. The procedures were according to the guidelines of the medical ethics board of the University Medical Center Groningen. Written informed consent was obtained for the use of the tonsil samples from the parents of the children. The HL cell lines were cultured at 37°C under an atmosphere containing 5% CO2 in RPMI-1640 medium (Cambrex Biosciences, Walkersville, USA) supplemented with ultraglutamine (2mM), penicillin (100U/ml), streptomycin (0.1mg/ml; Cambrex Biosciences) and 20% (DEV, L540, SUP-HD1), 10% (L1236, KM-H2) or 5% (L428) fetal calf serum (Cambrex Biosciences. We routinely confirmed the identity of our cell lines using the PowerPlex® 16 HS System (Promega, Leiden, The Netherlands).

2.2 RNA isolation from cell lines, B-cell subsets and nuclear and cytoplasmic fractions

RNA was isolated from cell lines and B-cell subsets as described earlier[17]. RNA concentration was measured with a NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific Inc., Waltham, USA) and integrity was assessed by analysis of the 18S/28S bands on a 1% agarose gel. Nuclear and cytoplasmic fractions were isolated from the cell lines L1236, L428 and L540 as described previously[17].

2.3 Microarray analysis

Design of the custom lncRNA microarray, labeling and hybridization procedures, data analysis and generation of heatmaps were performed as described previously[17]. Naïve, memory, GC-B and HL cell lines were Cy3 labeled using 50-100ng total RNA. Probes consistently flagged as present and expressed in the 10-100th percentile in at least 1 of the 3 conditions for the comparison of naïve, GC-B and memory B cells were included in the statistical analyses. For the comparison of the GC-B and HL samples probes consistently expressed 1 out of 2 conditions were included. We detected 10,469 lncRNA and 17,885 mRNA probes with consistent signals above the background for the 3 B-cell subsets and 9,955 lncRNA and 17,551 mRNA probes for the HL vs GC-B cell comparison. Significant differences in transcript abundance were determined by one-way ANOVA (normal B-cell subsets) or unpaired T-test (HL vs GC-B cells). Benjamini-Hochberg multiple testing correction was applied. Microarray data used for this publication were

deposited at NCBI Gene Expression Omnibus under accession GSE81086 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE81086).

2.4 Quantitative RT-PCR

First strand cDNA was synthesized with 500ng RNA in a total reaction volume of 20µL using the following reagents: random hexamer primers, 10mM dNTP mix, 5× first-strand buffer, 0.1M DTT Solution, 1µL RNaseOUT™ Recombinant Ribonuclease Inhibitor and 1µL superscript II reverse transcriptase (Life Technologies). PCR reactions were performed in triplicate on a Lightcycler 480 system (Roche, Penzberg, Germany). Each qPCR reaction contained 1ng cDNA, SYBR Green Master Mix (Applied Biosystems) and 3µM primers in a reaction volume of 10µL. U6 served as endogenous control for normalization. Relative expression levels are calculated as $2^{-\Delta Ct}$. For analysis of the subcellular localization expression was normalized to 18S when the cytoplasmic fraction was compared to the total fraction and to U3 snoRNA when the nuclear fraction was compared to the total fraction. TABLE S1 contains the sequences of all primers used in this study.

2.5 Processing of the cancer genome atlas (TCGA) RNA-Seq data

Fastq files for 20 randomly selected samples per cancer type were mapped to the human genome (build hg19) using tophat. Gene expression values were quantified using Cufflinks software version 2.1.1 (http://cole-trapnell-lab.github.io/cufflinks/releases/v2.1.1) on the basis of the Ensembl reference transcriptome annotation (version 75).

2.6 IHC and RNA-FISH

An immunohistochemical staining for CCL17 was performed as described previously[18] to confirm that the tumor cells were positive for CCL17. FISH probe sets and reagents (QuantiGene ViewRNA ISH Tissue 2-Plex Assay kit) were purchased from Affymetrix (Santa Clara, California, USA). RNA-FISH was performed according to the instructions provided by the manufacturer using formalin-fixed paraffin-embedded tissue sections of 5µm. A probe set for CCL17 was used as a positive control and to identify the HRS tumor cells in combination with one of the three probe sets for the selected lncRNAs. CCL17 was visualized with a TYPE 6 probe set (Fast Blue) and the lncRNAs were visualized with TYPE 1 probe sets (Fast Red). Custom probe sets designed by the manufacturer covered the exons that are common to all isoforms. Slides were imaged on a Leica SP8 confocal microscope.

2.7 Identification of differentially expressed lncRNA-mRNA pairs mapping within a 60kb region

To identify putative *cis*-acting lncRNAs we determined for each differentially expressed lncRNA whether there was a differentially expressed protein-coding gene in close

vicinity (directly adjacent) using Galaxy (https://usegalaxy.org)[19, 20]. The genomic coordinates of all differentially expressed probes were uploaded using the get data tool. We next defined the neighboring differentially expressed mRNAs using operate on genomic intervals tools, fetch closest non-overlapping feature for every interval and joint datasets. The resulting dataset was subsequently filtered for lncRNA-mRNA pairs within 60kb of each other.

3 Results

3.1 Highly dynamic lncRNA expression changes during the germinal center reaction

Statistical analysis of the gene expression profiles generated for naïve, GC and memory B-cells revealed 401 significantly differentially expressed lncRNA probes (251 lncRNA loci) (ANOVA, FDR<0.05) with a fold change of at least 2 in expression levels. The naïve and memory B-cell subsets showed a highly similar expression pattern with a significantly differential expression for only 2 probes. In contrast, GC-B cells showed a distinct lncRNA expression pattern with 377 differentially expressed probes compared to naïve and 376 compared to memory B cells. Unsupervised hierarchical clustering using the 401 lncRNA probes revealed two distinct clusters, one cluster with the GC-B cell samples and a second cluster with the naïve and memory B-cell subsets (**FIGURE 1A**). Analysis of the differentially expressed mRNAs (2,908 probes, 2,505 loci) revealed a clustering pattern similar to that of lncRNAs (**FIG. S1A**).

Next, we clustered the 6 HL cell lines together with the B-cell subsets using the 401 lncRNA probes that were significantly different between the three normal B-cell subsets. This revealed a pair-wise clustering of the GC-B cells and HL cell lines in one tree and the naïve and memory B-cells in a separate tree (**FIGURE 1B**). The nodular lymphocyte predominant Hodgkin lymphoma (NLPHL) cell line DEV clustered in the same branch as the GC-B cells, while the 5 cHL cell lines clustered in a separate branch next to the GC-B cells. This is in line with the GC-B cell origin of the tumor cells of HL, with a loss-of-B-cell phenotype in cHL but not in NLPHL. A similar result was observed using the 2,908 mRNA probes that were differentially expressed between the normal B-cell subsets (**FIG. S1B**).

3.2 LncRNA expression is widely deregulated in cHL

Comparison of the expression profiles of cHL (excluding the NLPHL cell line DEV) to GC-B cells revealed a significantly differential expression for 639 lncRNA probes (475 lncRNA loci) (**FIGURE 1C**). Comparison of the mRNA expression profiles revealed 2,402

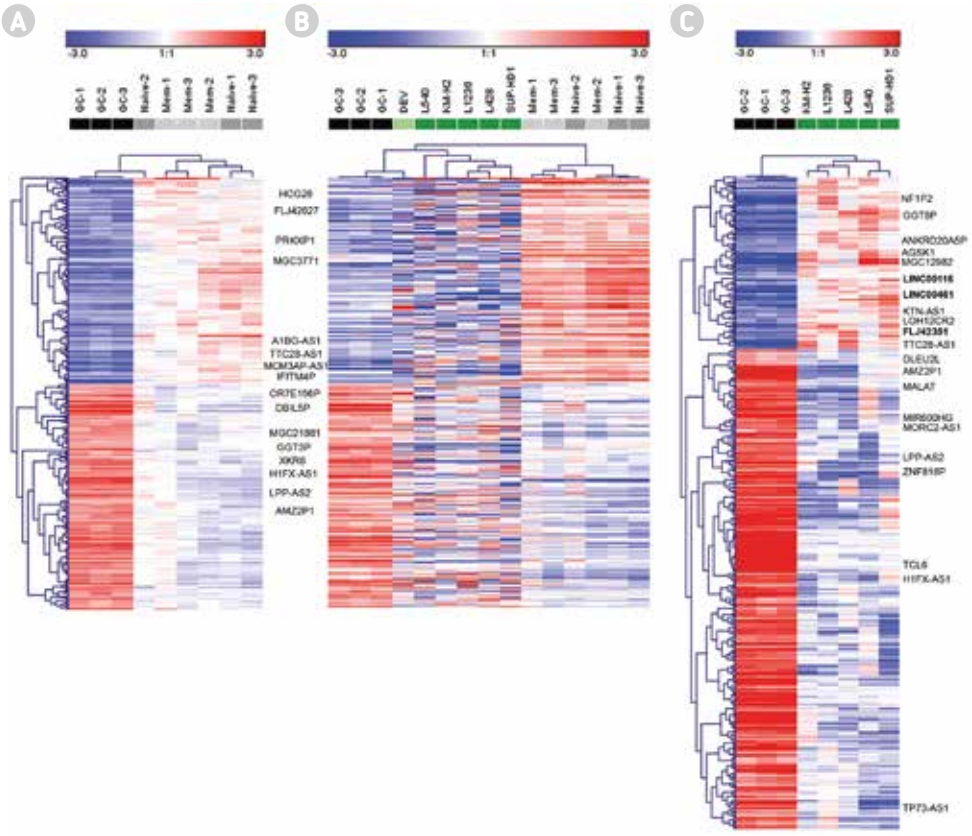

FIGURE 1 Unsupervised hierarchical clustering of differentially expressed lncRNAs. **(A)** Heatmap of the 401 lncRNA probes (251 lncRNA loci) differentially expressed between germinal center, naïve and memory B-cells. **(B)** Heatmap of the same 401 human lncRNA probes, now including the cHL cell lines L540, KM-H2, L1236, L428 and SUP-HD1 and the NLPHL cell line DEV. **(C)** Unsupervised clustering of the 639 lncRNAs probes (475 lncRNA loci) differentially expressed between cHL and GC-B cells. The positions of some known and well-annotated lncRNAs are indicated in the heatmaps. In bold are the three lncRNA candidates which were further studied by RNA-FISH.

differentially expressed probes (2,023 loci, FIG. S1C). LncRNAs were predominantly downregulated in HL (74% down vs 26% up), whereas mRNAs showed a more equal distribution between up and downregulated transcripts (56% down vs 44% up). Comparison of the lncRNAs and mRNAs differentially expressed between the normal B-cell subsets and between cHL cell lines and GC-B cells revealed a limited overlap of 70 lncRNA and 581 mRNA probes.

To validate the differential expression patterns we designed qRT-PCR primer sets for 9 up- and 5 downregulated lncRNAs randomly selected from the set of lncRNAs differentially expressed between cHL and GC-B cells. All 9 upregulated lncRNAs

showed a higher expression in cHL consistent with the microarray data, with significant differences for 6 of them (FIG. S2). For the downregulated lncRNAs we could confirm significant changes for 3 out of 5 lncRNAs. For the remaining 2 lncRNAs we were not able to design primer sets that could amplify these lncRNA transcripts.

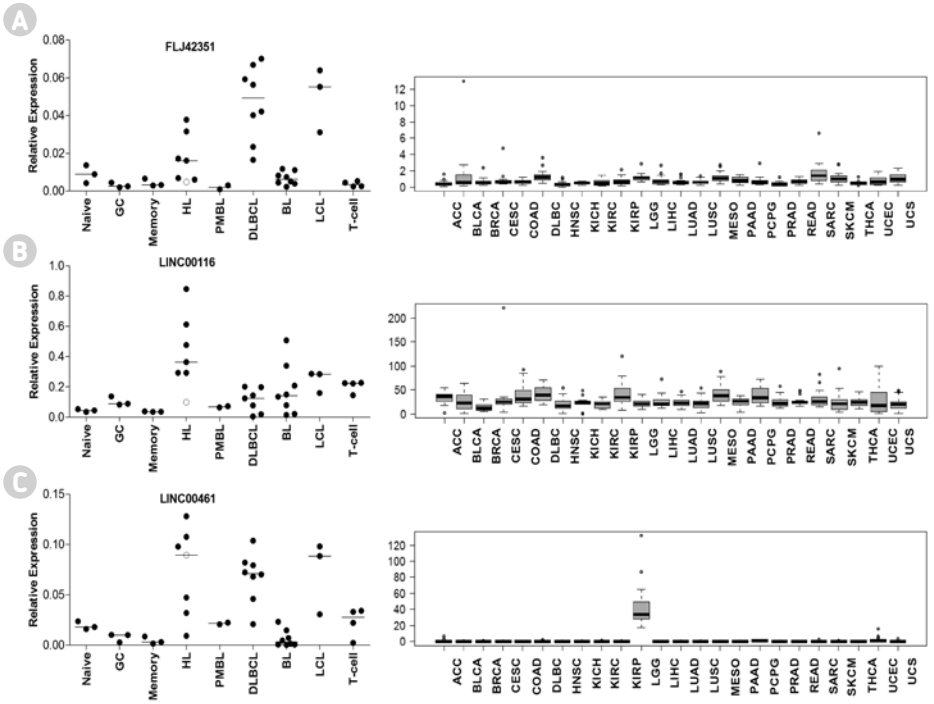

FIGURE 2 LncRNA expression levels in normal B-cells, lymphoma and other cancer tissues. QRT-PCR analysis (left panels) and re-analysis of TCGA RNA-seq data (right panels) for **(A)** FLJ42531, **(B)** LINC00116 and **(C)** LINC00461. Groups consisted of: Naïve, CD19+IgD+CD38- cells (3); GCB, CD19+IgD-CD38+ (3); Memory, CD19+ IgD-CD38- (3); HL, L428, L540, L1236, KM-H2, L591, SUP-HD1 and NLPHL, DEV (open circle); PMBL, K1106P and MEDB1; DLBCL, Karpas 422, SU-DHL-4, SU-DHL-5, SU-DHL-6, SU-DHL-10, OCILy-3, U-2932, DOHH2; BL, ST486, DG75, RAMOS, CA46, BL65, RAJI, Jiyoye and Nawalma; LCL, LCL6A, LCL39, LCL89; T-cell, Jurkat, HUT-78, KARPAS 299, SR678. Left panel abbreviations: HL, Hodgkin lymphoma, NLPHL, Nodular lymphocyte predominant Hodgkin lymphoma; PMBL, Primary Mediastinal B-cell lymphoma; DLBCL, Diffuse large B-cell lymphoma; BL, Burkitt lymphoma; LCL, Lymphoblastoid cell line; T-cell, T-cell leukemia and lymphoma cell lines. TCGA abbreviations: ACC, Adrenocortical carcinoma; BLCA, Bladder Urothelial Carcinoma; BRCA, Breast invasive carcinoma; CESC, Cervical squamous cell carcinoma and endocervical adenocarcinoma; COAD, Colon adenocarcinoma; DLBC, Diffuse large B-cell Lymphoma; HNSC, Head and Neck squamous cell carcinoma; KICH, Kidney Chromophobe; KIRC, Kidney renal clear cell carcinoma; KIRP, Kidney renal papillary cell carcinoma; LGG, Brain Lower Grade Glioma; LIHC, Liver hepatocellular carcinoma; LUAD, Lung adenocarcinoma; LUSC, Lung squamous cell carcinoma; MESO, Mesothelioma; PAAD, Pancreatic adenocarcinoma; PCPG, Pheochromocytoma and Paraganglioma; READ, Rectum adenocarcinoma; SARC, Sarcoma; SKCM, Skin Cutaneous Melanoma; THCA, Thyroid carcinoma; UCEC, Uterine Corpus Endometrial Carcinoma; UCS, Uterine Carcinoma.

3.3 LncRNA expression in B cells, lymphoma cell lines and primary cHL tissues

We selected three of the validated upregulated lncRNAs, i.e. FLJ42351, LINC00116 and LINC00461 (FIG. S3) for further expression analysis in normal B-cell subsets and a panel of lymphoma cell lines. This revealed high levels in HL, DLBCL and LCL cell lines for FLJ42351 and LINC00461 (FIGURE 2A AND C). For LINC00116, high levels were mainly observed in the HL cell lines and in part of the BL cell lines (FIGURE 2B). Reanalysis of TCGA RNA-seq data across 24 different cancer types revealed relative low expression levels for FLJ42351 and LINC00116 without clear differences between different cancer tissues (right panels of FIGURE 2). For LINC00461, higher levels were observed only for lower grade glioma.

Next, we used RNA-FISH to study the expression of the three in cHL cell lines upregulated lncRNAs in tissue sections of primary cHL cases. The five randomly selected cHL cases showed positive staining for CCL17 by immunohistochemistry (data not shown). RNA-FISH using the CCL17 probe-set as a positive control revealed staining of the HRS cells in four of the five cases. RNA-FISH using the lncRNA probe-sets revealed a tumor cell specific staining in the four remaining cHL cases for all three lncRNAs (FIGURE 3). The probe-sets for FLJ42351 and LINC00461 showed positive signals in three cases and weak positive signals in one case, the probe set for LINC00116 showed positive signals in all four cases.

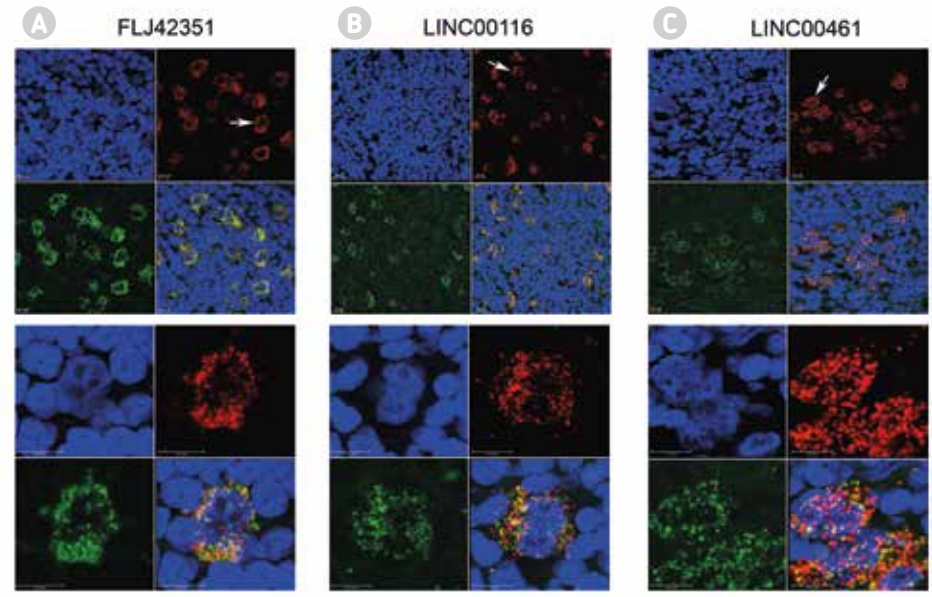

FIGURE 3 **Expression of FLJ42351, LINC00116 and LINC00461 is restricted to the HRS cells of cHL cases.** Dual-color RNA-FISH was performed using a probe-set for CCL17 (tumor cell specifically expressed) in combination with probe-sets for FLJ42351 **(A)**, LINC00116 **(B)** or LINC00461 **(C)**. Shown are representative images of the DAPI, CCL17 (in red, fast blue) and lncRNA (in green, Cy3/fast red) signals and the overlay of the three. Arrows in the top row panels indicate the HRS tumor cell that was used for the zoom-in pictures that are shown in the bottom row panels. Scale bar 10µm.

TABLE 1 **Co-regulation of lncRNAs and mRNAs in HL vs GCB.**

	lncRNA			mRNA			lncRNA to mRNA		Direction
Chr	Probe name	Probe location	Transcript(s)	Probe name	Probe location	Gene Symbol	Distance (kb)	Orientation	lncRNA / mRNA
chr1	CUST_4772_PI427622066	212837956-212838015	TCONS_I2_00001761-2, 5	A_23_P34915	212793856-212793915	ATF3	-44	tail-to-tail	up/ up
chr1	CUST_4757_PI427622066	212805800-212805859	TCONS_00002580 / TCONS_00001808-11	A_23_P34915	212793856-212793915	ATF3	-11	tail-to-tail	up/ up
chr1	CUST_4772_PI427622066	212837956-212838015	TCONS_I2_00001761-2, 5	A_23_P160720	212860087-212860028	BATF3	22	sense	up/ up
chr1	CUST_4751_PI427622066	212805846-212805905	TCONS_00001808 -11	A_23_P160720	212860087-212860028	BATF3	54	sense	up/ up
chr1	A_33_P3411328	203269703-203269644	NR_034151.1 LINC01136	A_23_P62901	203278454-203278513	BTG2	8	head-to-head	down/ down
chr1	CUST_1604_PI427622066	28573319-28573378	TCONS_00001440	A_23_P35082	28608194-28608253	SESN2	34	head-to-head	down/ up
chr2	CUST_9360_PI427622066	113401676-113401735	TCONS_00002672 / TCONS_00005231 / FLJ42351	A_23_P165657	113421288-113421347	SLC20A1	19	head-to-head	down/ down
chr2	CUST_10333_PI427622066	169957576-169957635	TCONS_00003045	A_23_P56559	169952375-169952434	DHRS9	-5	sense	up/ up
chr2	**CUST_8107_PI427622066**	**64313597-64313656**	**TCONS_00003301**	**A_33_P3316928**	**64320119-64320060**	**PELI1**	**6**	**sense**	**down/ down**
chr2	CUST_8408_PI427622066	74639943-74640002	TCONS_I2_00014653	A_23_P68072	74652257-74652316	WDR54	12	head-to-head	down/ down
chr2	CUST_6868_PI427622066	6970109-6970168	TCONS_00004118 / TCONS_00003184	A_33_P3401826	6988526-6988467	CMPK2	18	sense	down/ down
chr3	CUST_11860_PI427622066	48767088-48767147	TCONS_00006507	A_33_P3288023	48716057-48715998	NCKIPSD	-51	sense	down/ down
chr3	A_33_P3368771	129036811-129036870	NR_026991.1 H1FX-AS1	A_23_P251377	129020861-129020920	C3orf37	-15	sense	down/ down
chr4	CUST_19108_PI427622066	183798702-183798761	TCONS_00009180-2	A_24_P337334	183812095-183812036	DCTD	13	sense	up/ up
chr4	CUST_12908_PI427622068	115603442-115603501	TCONS_00008872	A_24_P103264	115597201-115597260	UGT8	-6	sense	down/ down
chr5	A_32_P127350	87960601-87960542	NR_024384. LINC00461	A_23_P320739	88016876-88016817	MEF2C	56	sense	down/ down
chr5	CUST_22952_PI427622066	150555338-150555397	TCONS_00011070	A_24_P97825	150561538-150561479	CCDC69	6	sense	up/ down
chr6	CUST_20144_PI427622068	32870040-32870099	TCONS_I2_00024170-2	A_32_P351968	32902563-32902504	HLA-DMB	32	tail-to-tail	down/ up
chr8	CUST_25512_PI427622068	9009366-9009425	TCONS_00014600 / TCONS_00014599 / TCONS_00015212	A_24_P201064	8993872-8993813	PPP1R3B	-15	head-to-head	down/ down
chr9	A_33_P3574055	37087746-37087805	TCONS_I2_00028716 , 18-22	A_33_P3348001	37038728-37038787	LOC100130458	-49	head-to-head	up/ up
chr9	CUST_34071_PI427622066	37384372-37384431	TCONS_00015807	A_23_P124855	37357730-37357789	ZCCHC7	-26	tail-to-tail	up/ up
chr9	A_23_P216935	125872202-125872143	NR_026677.1 MIR600HG	A_24_P252575	125866353-125866412	RABGAP1	-5	tail-to-tail	up/ down
chr9	A_23_P216935	125872202-125872143	NR_026677.1 MIR600HG	A_23_P252155	125887911-125887852	STRBP	15	sense	down/ down
chr10	CUST_40617_PI427622066	72370073-72370132	TCONS_00017839	A_23_P161481	72327843-72327902	KIAA1274	-42	sense	down/ down
chr11	CUST_42984_PI427622066	71162959-71163018	TCONS_00019374	A_23_P24444	71145573-71145514	DHCR7	-17	head-to-head	down/ down
chr11	A_23_P127367	67119197-67119138	NR_046412. POLD4	A_23_P323227	67166439-67166297	PPP1CA	47	sense	down/ down
chr12	CUST_44640_PI427622066	47643152-47643211	TCONS_00020398	A_33_P3215647	47629913-47629972	FAM113B	-13	sense	down/ down
chr12	**CUST_45746_PI427622066**	**125511162-125511221**	**TCONS_00020260**	**A_32_P831181**	**125497122-125497181**	**BRI3BP**	**-14**	**tail-to-tail**	**down/ down**
chr12	CUST_44512_PI427622066	32112102-32112161	TCONS_00021299	A_23_P98930	32145764-32145823	C12orf35	33	tail-to-tail	down/ down

chr13	CUST_38642_PI427622068	42080252-42080311	TCONS_00021766	A_24_P10137	42042937-42044635	C13orf15	-36	sense	up/ down
chr14	A_23_P25746	32544949-32544890	NR_027263. ARHGAP5-AS1	A_33_P3386459	32586421-32586480	ARHGAP5	41	head-to-head	down/ down
chr14	CUST_40389_PI427622068	71148640-71148699	TCONS_00022548	A_24_P273742	71141238-71141297	TTC9	-7	sense	down/ down
chr14	CUST_39858_PI427622068	96181979-96182038	TCONS_00022357	A_23_P357717	96176337-96176290	TCL1A	-5	head-to-head	down/ down
chr15	CUST_41903_PI427622068	40617065-40617124	TCONS_I2_00008581/TCONS_I2_00008583	A_33_P3260614	40580198-40580139	PLCB2	-36	head-to-head	down/ down
chr15	CUST_52013_PI427622066	85133449-85133508	TCONS_I2_00008800	A_33_P3229276	85147070-85147129	ZSCAN2	13	sense	down/ down
chr15	CUST_50180_PI427622066	56921713-56921772	TCONS_I2_00009507	A_33_P3262799	56947003-56946944	ZNF280D	25	sense	down/ down
chr16	A_24_P865226	29876156-29876499	NR_015396.1 CDIPT-AS1	A_33_P3294608	29859281-29859340	MVP	-17	tail-to-tail	down/ down
chr16	A_24_P865226	29876156-29876499	NR_024370.1 CDIPT-AS1	A_23_P10194	29882918-29882859	SEZ6L2	6	head-to-head	up/ up
chr16	CUST_54455_PI427622066	89980491-89980550	TCONS_I2_00010376-80	A_24_P193582	90025433-90025492	DEF8	44	head-to-head	down/ down
chr17	CUST_55786_PI427622066	47316514-47316573	TCONS_00025420/TCONS_00025926	A_23_P141429	47300191-47300250	ABI3	-16	sense	down/ down
chr17	CUST_46171_PI427622068	66202319-66202378	TCONS_I2_00010943/TCONS_I2_00010942	A_33_P3344648	66196091-66196150	LOC440461	-6	sense	down/ up
chr17	A_32_P19000	76107395-76107336	NR_040071.1 TNRC6C-AS1	A_33_P3302428	76101674-76101733	TNRC6C	-5	antisense	down/ down
chr17	A_32_P19000	76107395-76107336	NR_040071.1 TNRC6C-AS1	A_23_P101013	76113361-76109658	TMC6	4	sense	down/ down
chr17	CUST_46165_PI427622068	66184876-66184935	TCONS_00025471	A_33_P3344648	66196091-66196150	LOC440461	11	sense	down/ down
chr17	CUST_56085_PI427622066	62969576-62969635	TCONS_I2_00011403, 6- 9, 12-16 / TCONS_I2_00011685-7	A_24_P941441	63007307-63007248	GNA13	37	sense	down/ down
chr18	A_32_P150876	5245815-5245874	NR_015389. Homo sapiens LINC00667	A_24_P74064	5289958-5289899	ZFP161	44	tail-to-tail	up/ up
chr19	A_33_P3685216	58858293-58858234	NR_049780. ZNF837	A_33_P3411427	58879052-58878993	ZNF837	20	sense	down/ down
chr19	CUST_48657_PI427622068	53700394-53700453	TCONS_I2_00012685-7	A_24_P310224	53643774-53643715	ZNF347	-56	head-to-head	up/ up
chr19	CUST_59058_PI427622066	44405536-44405595	TCONS_00027367-8 / TCONS_00026816 / TCONS_00027759	A_23_P90333	44376876-44376817	ZNF404	-28	sense	down/ down
chr19	CUST_48777_PI427622068	56905693-56905752	TCONS_00026798-99 / TCONS_00027118-26	A_33_P3418731	56891060-56891119	ZNF542	-14	sense	up/ down
chr19	A_24_P941487	53960813-53960872	NR_003148. TPM3P9	A_33_P3216433	53947715-53947774	LOC147804	-13	sense	down/ down
chr19	CUST_48637_PI427622068	53095380-53095439	TCONS_I2_000126-69	A_23_P130444	53087417-53087476	ZNF701	-7	sense	down/ down
chr19	A_33_P3646051	37294755-37294814	NR_040027.1 ZNF790-AS1	A_33_P3423401	37308486-37308427	ZNF790	13	antisense	up/ up
chr19	CUST_59260_PI427622066	53102902-53102961	TCONS_I2_00013339 / TCONS_I2_00012668	A_33_P3256868	53116330-53116271	ZNF83	13	tail-to-tail	down/ down
chr20	**CUST_50118_PI427622068**	**55857991-55858050**	**TCONS_00027871**	**A_24_P91566**	**55803439-55803380**	**BMP7**	**-54**	**head-to-head**	**down/ down**
chr20	CUST_61077_PI427622066	56175225-56175284	TCONS_00028073	A_24_P413126	56223548-56223489	PMEPA1	48	sense	down/ down
chr22	CUST_63826_PI427622066	35850592-35850651	TCONS_00029585 / TCONS_00029795	A_23_P132277	35819307-35820178	MCM5	-30	sense	down/ down
chrX	CUST_30264_PI427622068	46405002-46405061	TCONS_00016919	A_23_P319617	46457408-46457467	CHST7	52	sense	up/ up
chrX	CUST_30840_PI427622068	134531578-134531637	TCONS_00017014	A_23_P309865	134496752-134496811	ZNF449	-34	sense	down/ up

* Probe to probe distance in kb, positive distance means the lncRNA probe is upstream of the protein-coding gene, negative distance indicates the reverse. # The direction of the lncRNA is indicated first, the direction of the mRNA is indicated second. Direction is based on cHL cell lines compared to GC-B cells. In bold the lncRNA-mRNA pairs are indicated that overlap with the putative *cis*-regulated lncRNA-mRNA pairs identified in the transcriptome comparisons of the normal B-cell subsets (Table S2).

3.4 LncRNAs and neighboring protein-coding genes

To identify lncRNAs that potentially affect the expression of nearby protein-coding genes we determined which of the differentially expressed lncRNAs was close to a differentially expressed mRNA. Within the 251 lncRNAs differentially expressed between the normal B-cell subsets we identified 51 putative *cis*-regulatory lncRNAs with a probe-to-probe distance of up to 60kb to the differentially expressed mRNA (TABLE S2). Of the 475 lncRNAs differentially expressed between cHL and GC-B cells 59 were in close vicinity of a differentially expressed mRNA (TABLE 1).

Three of the putative *cis*-regulated lncRNA-mRNA pairs were detected in both the normal B-cell subsets and the HL versus GC-B cell analysis (indicated in bold). For all three pairs a positive correlation between the lncRNA and mRNA expression was observed in both analyses.

To support a putative *cis*-regulatory role we determined for two (LINC00461 and FLJ42351) of the lncRNAs for which we already designed qPCR primers whether they were preferentially localized in the nucleus. Fractionation procedures of L428, L1236 and L540 cell lines were validated by qPCR for six transcripts with a known subcellular localization. This revealed the expected enrichment of RPPH1, DANCR and MT-TK transcripts in the cytoplasmic fractions and of MIAT1, ANRIL and XIST transcripts in the nuclear fractions (FIGURE 4A AND 4B). In line with a *cis*-regulatory role, LINC00461 and FLJ42351 indeed showed enrichment in the nuclear fractions and depletion in the cytoplasmic fractions. Analysis of the putative *cis*-regulated protein-coding genes by qRT-PCR confirmed the inverse expression pattern between *LINC00461* and the nearby upstream protein-coding gene, *MEF2C* (FIGURE S2C). For FLJ42351, we could confirm the comparable expression patterns for this lncRNA and the neighboring protein-coding gene *SLC20A1*.

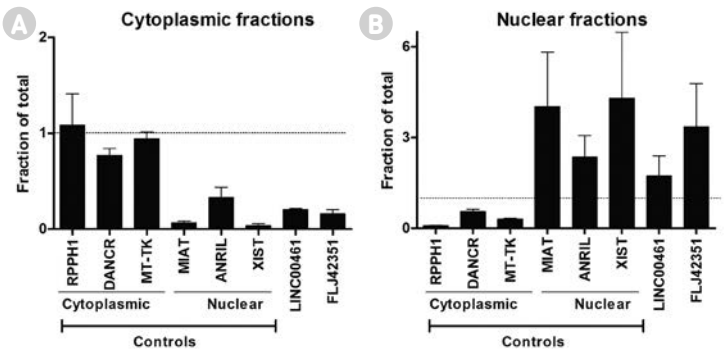

FIGURE 4 LINC00461 and FLJ42351 are nuclear enriched lncRNAs. As a validation for the isolation of cytoplasmic **(A)** and nuclear **(B)** fractions the enrichment of three cytoplasmic control RNAs (RPPH1, DANCR and MT-TK) in the cytoplasmic fraction and three nuclear control (MIAT, ANRIL and XIST) RNAs in the nuclear fraction were confirmed by qRT-PCR. Analysis of LINC00461 and FLJ42351 shows a clear enrichment in the nuclear fraction and depletion in the cytoplasmic fraction. ΔΔCt values are shown with the average of the total fractions of L428, L1236 and L540 HL cell lines set to 1 (dashed line). 18S (cytoplasmic) and U3 (nuclear) were used for normalization of cytoplasmic and nuclear fractions, respectively.

4 Discussion

In this study, we showed a dynamic regulation of lncRNA expression during the transition of B cells through the GC, i.e. from naïve B cell to GC-B cell resulting in memory B cells. Furthermore, the comparison of cHL cell lines with GC-B cells revealed a significant differential expression pattern for a considerable number of lncRNAs. This indicates that lncRNAs play a role in normal B-cell maturation and that deregulated lncRNA expression is a prominent feature of HL.

We showed that the levels of 401 probes corresponding to 251 lncRNA loci change during the transition from naïve to GC-B cells and reverse to levels comparable to naïve B cells once the cells differentiate to memory B cells. Recently, four studies analyzed lncRNA expression in B-cell populations[21-24]. Only one of the four studies also analyzed naïve, memory and germinal center B-cell populations isolated from tonsil. In line with our study, they showed that the lncRNA expression profiles of memory and naïve B cells were similar and more distinct from GC-B cells[21]. However, none of the studies provided overviews of the lncRNAs differentially expressed between similar B-cell subsets as used in our study preventing a direct comparison to our results. At the mRNA level we and others observed a similar pattern as we observed for lncRNAs, i.e. similar profiles for naïve and memory B cells and significantly different profiles for GC-B cells[25]. The NLP cell line DEV clustered with normal GC-B cells and the cHL cell lines cluster in a sub-branch next to the GC-B cells for both the lncRNAs as well as the mRNA probes. This indicates that the expression profile of LP cells of NLPHL more closely resembles that of GC-B cells than that of the HRS cells of cHL. These results are consistent with previous gene expression studies showing a loss-of-B-cell phenotype in HRS cells[26] and intermediate expression patterns in LP cells[27].

A direct comparison of cHL cell lines with GC-B cells revealed significant differences for 639 probes (475 lncRNA loci). Almost 75% of these lncRNAs were downregulated in the cHL cell lines, while the percentage of up and downregulated mRNAs was quite similar. This suggests that lncRNAs might, similar to microRNAs,[28] show a global downregulation in cancer. Previous studies comparing the expression profiles of microdissected primary HRS cells and cHL cell lines to GC-B cells showed a considerable overlap between the differentially expressed genes in primary HRS cells and cHL cell lines. However, a subset of the genes differentially expressed in primary HRS cells, including chemotaxis and surface receptor signaling related genes, were not differentially expressed in the cell lines[26, 28]. This indicates that it is likely that the lncRNA expression profile of the HL cell lines will also not fully reflect the lncRNA profile of primary HRS cells.

Three of the upregulated lncRNAs, i.e. LINC00116, LINC00461 and FLJ42351, were studied in more detail, including expression analysis in primary cHL tissues. RNA-FISH for LINC00116 revealed a remarkable specific expression pattern, with fluorescent

signals being restricted to the HRS cells. The qRT-PCR across normal B-cell subsets and lymphoma cell lines showed a similar pattern with a preferential expression in HL cells. TCGA data did not show a preferential expression pattern across multiple other types of cancer. LINC00116 has three known transcript variants, two short variants and one long variant. The longer transcript (LINC00116-001) might encode for a 138 amino acid uncharacterized protein (http://www.uniprot.org/uniprot/Q8NCU8), whereas the shorter variants represent true lncRNA transcripts without protein-coding potential. The expression pattern of the probes present on our array pointed to a predominant expression of the shorter isoforms (LINC00116-002 LINC00116-003, **FIG. S3**). Moreover, the probes for RNA-FISH and the primers used for the qPCR were also designed based on these short non-coding transcript isoforms. Thus, our data support a HRS cell specific expression pattern of the two short LINC00116 transcript variants.

LINC00461 also showed a highly specific expression pattern in the HRS cells by RNA-FISH. We also noticed enhanced expression of LINC00461 in HL, DLBCL and LCL cell lines. Analysis of the TCGA data revealed expression only in low-grade gliomas and not in DLBCL tissues. The difference between the qPCR and TCGA results for DLBCL may be related to differences that sometimes can be observed between cell lines and primary tissues. The LINC00461 locus contains several isoforms of which the array probe, the qRT-PCR primer set and the FISH probe-set all detect both the 3.6kb (NR_024384) and the 3.4kb (NR_024383) isoforms (**FIG. S3**). The last exon of both lncRNA isoforms contains the pre-miR-9-2 sequence. Whether this lncRNA, besides serving as a primary transcript for miR-9, also has other functions remains to be determined. It has been reported that miR-9 expression is increased in HL and glioma and that inhibition of miR-9 inhibits growth of HL cell lines in a xenograft mouse model[29-32]. The expression pattern of the nuclear enriched LINC00461 showed an inverse correlation with the expression of the nearby upstream protein-coding gene, *MEF2C*. MEF2C is essential for B-cell proliferation and survival in response to BCR stimulation *in vitro*[33]. *In vivo*, MEF2C loss results in reduced antibody responses to T-cell-dependent antigens and impaired GC formation[33, 34]. Studies in rats have shown that MEF2C can stimulate expression of the miR-9-2 locus by binding to a highly conserved site in the promoter of the miR-9-2 host gene[35]. On the other hand, MEF2C is a predicted miR-9 target via a broadly conserved 8-mer site in the 3'UTR. How LINC00461 and MEF2C regulate each other's expression and to what extend this is relevant for cHL remains to be investigated.

FLJ42351, the third lncRNA studied in more detail, also showed a remarkable HRS cell specific expression in primary HL tissues. qPCR confirmed expression in HL cell lines and showed in addition expression in DLBCL and LCL cell lines. TCGA RNA-seq data did not show substantial expression in other types of cancer tissues. The expression pattern of FLJ42351 is similar to that of the immediately downstream phosphate transporter SLC20A1. In line with a *cis*-regulatory role of FLJ42351 we did observe a preferential nuclear localization in the cHL cell lines. SLC20A1 knockdown in HeLa and hepatic cell lines reduced cell proliferation[36] and SLC20A1 knockout mice showed

increased apoptosis rates in erythroid cells[37]. We previously identified FLJ42351 as a MYC-induced transcript in P493-6 cells[17]. In support of a putative *cis*-regulatory role of this lncRNA, we also identified SLC20A1 as a positively correlating neighboring gene of FLJ42351 in our MYC-regulated lncRNA study. Thus, FLJ42351 is a HRS cell expressed lncRNA whose expression pattern is similar to that of the neighboring protein-coding gene *SLC20A1*, which is known to have important functions in cell growth.

Despite several attempts we were not able to efficiently knockdown LINC00461 and FLJ42351 with siRNA and lentiviral shRNA based approaches. This precluded a further analysis to provide more definitive proof that these two lncRNAs can indeed regulate the expression of the neighboring genes. Interestingly, both lncRNAs do have a DNase I hypersensitivity site (DHS), which is indicative of regulatory sequences, i.e. ~15 kb upstream of the transcriptional start site (TSS) of *LINC00461* in both L1236 and L428 cell lines[38] and ~45 kb downstream of the *FLJ42351* TSS in L1236.

In summary, our comprehensive expression analysis shows dynamic regulation of lncRNA expression during the GC transition of B cells and a widely deregulated expression in HL. These lncRNAs, together with the putative *cis*-acting data, provide a valuable source for further studies aiming at understanding the role of lncRNAs in normal B-cell biology and in the pathogenesis of HL.

5 References

1 Hjalgrim, H. & Engels, E. A. Infectious aetiology of Hodgkin and non-Hodgkin lymphomas: a review of the epidemiological evidence. *J. Intern. Med.* **264**, 537-548 (2008).

2 Schwering, I. *et al.* Loss of the B-lineage-specific gene expression program in Hodgkin and Reed-Sternberg cells of Hodgkin lymphoma. *Blood* **101**, 1505-1512 (2003).

3 Hertel, C. B., Zhou, X. G., Hamilton-Dutoit, S. J. & Junker, S. Loss of B cell identity correlates with loss of B cell-specific transcription factors in Hodgkin / Reed-Sternberg cells of classical Hodgkin lymphoma. *Oncogene* **21**, 4908-4920 (2002).

4 Tsai, H. K. & Mauch, P. M. Nodular lymphocyte-predominant hodgkin lymphoma. *Semin. Radiat. Oncol.* **17**, 184-189 (2007).

5 Kuppers, R. B cells under influence: transformation of B cells by Epstein-Barr virus. *Nat. Rev. Immunol*. **3**, 801-812 (2003).

6 Jost, P. J. & Ruland, J. Aberrant NF-kappaB signaling in lymphoma: mechanisms, consequences, and therapeutic implications. *Blood* **109**, 2700-2707 (2007).

7 Plattel, W., Kluiver, J., Diepstra, A., Visser, L. & van den Berg, A. in *MicroRNAs in Medicine* (ed Lawrie, C. H.) 435-447, 2013).

8 Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775-1789 (2012).

9 Ma, L., Bajic, V. B. & Zhang, Z. On the classification of long non-coding RNAs. *RNA Biol*. **10**, 925-933 (2013).

10 Orom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46-58 (2010).

11 Nagano, T. *et al.* The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717-1720 (2008).

12 Guil, S. & Esteller, M. *Cis*-acting noncoding RNAs: friends and foes. *Nat. Struct. Mol. Biol.* **19**, 1068-1075 (2012).

13 Kung, J. T., Colognori, D. & Lee, J. T. Long noncoding RNAs: past, present, and future. *Genetics* **193**, 651-669 (2013).

14 Prensner, J. R. & Chinnaiyan, A. M. The emergence of lncRNAs in cancer biology. *Cancer. Discov.* **1**, 391-407 (2011).

15 Campo, E. *et al.* The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood* **117**, 5019-5032 (2011).

16 Tan, L. P. *et al.* miRNA profiling of B-cell subsets: specific miRNA profile for germinal center B cells with variation between centroblasts and centrocytes. *Lab. Invest.* **89**, 708-716 (2009).

17 Winkle, M. *et al.* Long noncoding RNAs as a novel component of the Myc transcriptional network. *FASEB J.* **29**, 2338-2346 (2015).

18 Maggio, E. M. *et al.* Common and differential chemokine expression patterns in rs cells of NLP, EBV positive and negative classical Hodgkin lymphomas. *Int. J. Cancer* **99**, 665-672 (2002).

19 Blankenberg, D. *et al.* Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* **Chapter 19**, Unit 19.10.1-21 (2010).

20 Goecks, J., Nekrutenko, A., Taylor, J. & Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11**, R86-2010-11-8-r86. Epub 2010 Aug 25 (2010).

21 Petri, A. *et al*. Long Noncoding RNA Expression during Human B-Cell Development. *PLoS One* **10**, e0138236 (2015).

22 Ranzani, V. *et al.* The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nat. Immunol.* **16**, 318-325 (2015).

23 Bonnal, R. J. *et al*. De novo transcriptome profiling of highly purified human lymphocytes primary cells. *Sci. Data* **2**, 150051 (2015).

24 Verma, A. *et al.* Transcriptome sequencing reveals thousands of novel long non-coding RNAs in B cell lymphoma. *Genome Med.* **7**, 110-015-0230-7 (2015).

25 Klein, U. *et al*. Transcriptional analysis of the B cell germinal center reaction. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2639-2644 (2003).

26 Tiacci, E. *et al*. Analyzing primary Hodgkin and Reed-Sternberg cells to capture the molecular and cellular pathogenesis of classical Hodgkin lymphoma. *Blood* **120**, 4609-4620 (2012).

27 Brune, V. *et al.* Origin and pathogenesis of nodular lymphocyte-predominant Hodgkin lymphoma as revealed by global gene expression analysis. *J. Exp. Med.* **205**, 2251-2268 (2008).

28 Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-838 (2005).

29 Leucci, E. *et al.* Inhibition of miR-9 de-represses HuR and DICER1 and impairs Hodgkin lymphoma tumour outgrowth in vivo. *Oncogene* **31**, 5081-5089 (2012).

30 Van Vlierberghe, P. *et al.* Comparison of miRNA profiles of microdissected Hodgkin/Reed-Sternberg cells and Hodgkin cell lines versus CD77+ B-cells reveals a distinct subset of differentially expressed miRNAs. *Br. J. Haematol.* **147**, 686-690 (2009).

31 Nie, K. *et al*. MicroRNA-mediated down-regulation of PRDM1/Blimp-1 in Hodgkin/Reed-Sternberg cells: a potential pathogenetic lesion in Hodgkin lymphomas. *Am. J. Pathol.* **173**, 242-252 (2008).

32 Nass, D. *et al.* MiR-92b and miR-9/9* are specifically expressed in brain primary tumors and can be used to differentiate primary from metastatic brain tumors. *Brain Pathol.* **19**, 375-383 (2009).

33 Wilker, P. R. *et al.* Transcription factor Mef2c is required for B cell proliferation and survival after antigen receptor stimulation. *Nat. Immunol.* **9**, 603-612 (2008).

34 Homminga, I. *et al.* Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia. *Cancer. Cell.* **19**, 484-497 (2011).

35 Davila, J. L. *et al*. A positive feedback mechanism that regulates expression of miR-9 during neurogenesis. *PLoS One* **9**, e94348 (2014).

36 Beck, L. *et al.* Identification of a novel function of PiT1 critical for cell proliferation and independent of its phosphate transport activity. *J. Biol. Chem.* **284**, 31363-31374 (2009).

37 Liu, L., Sanchez-Bonilla, M., Crouthamel, M., Giachelli, C. & Keel, S. Mice lacking the sodium-dependent phosphate import protein, PiT1 (SLC20A1), have a severe defect in terminal erythroid differentiation and early B cell development. *Exp. Hematol.* **41**, 432-43. e7 (2013).

38 Kreher, S. *et al.* Mapping of transcription factor motifs in active chromatin identifies IRF5 as key regulator in classical Hodgkin lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4513-22 (2014).

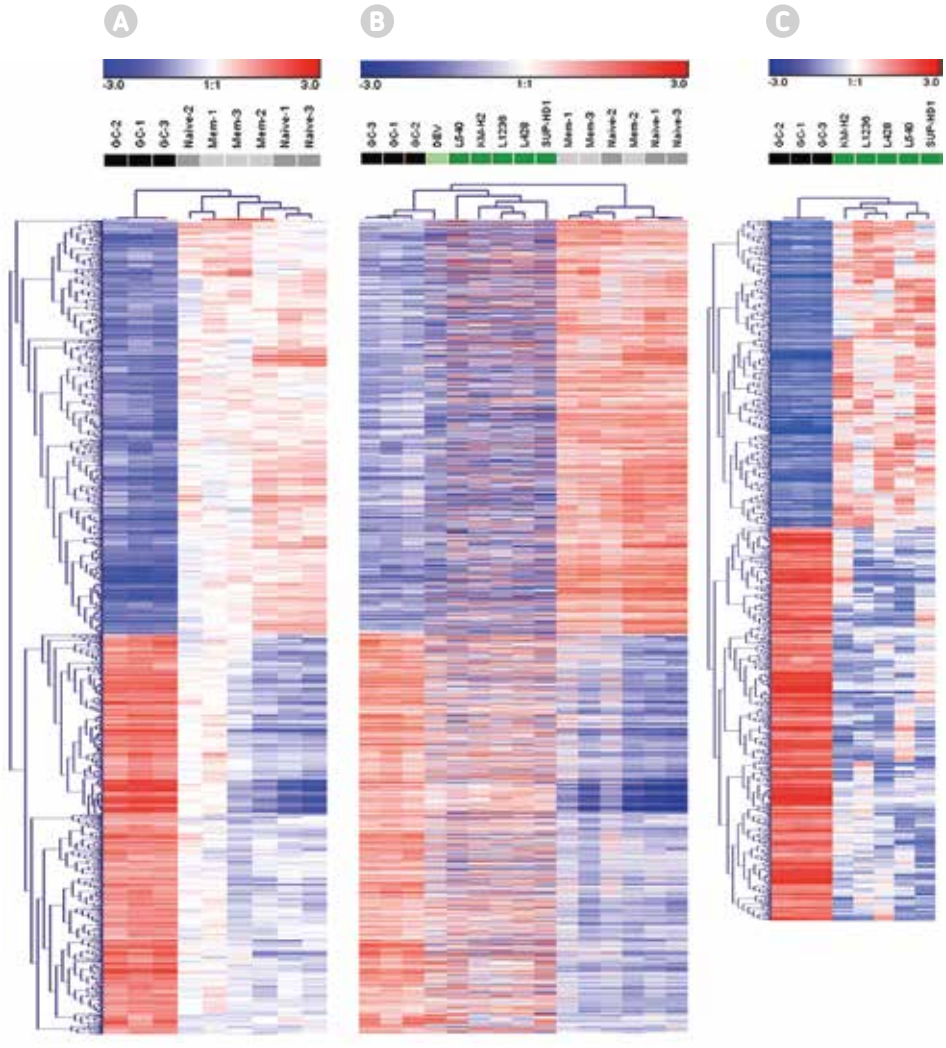

FIGURE S1 Unsupervised hierarchical clustering of differentially expressed mRNAs. (A) Heatmap of the 2,908 differentially expressed mRNA probes between germinal center, naïve and memory B-cells. **(B)** Heatmap of the same 2,908 human mRNA probes, now including the cHL cell lines L540, KM-H2, L1236, L428 and SUP-HD1 and the NLPHL cell line DEV. **(C)** Unsupervised clustering of the 2,402 mRNA probes differentially expressed human between cHL cell lines and GC-B cells.

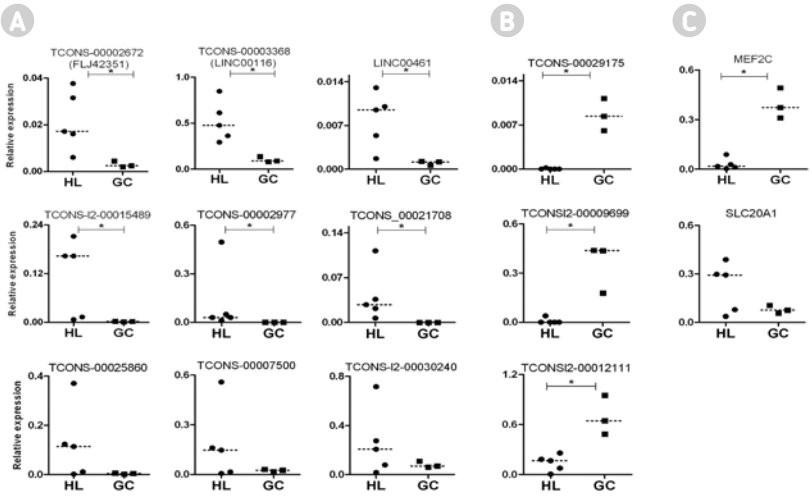

FIGURE S2 LncRNA expression validation. Confirmation of array results by real-time PCR analysis for selected upregulated **(A)** and downregulated **(B)** lncRNAs using the same RNA samples as used for the microarray study. Significant differences in the levels of six up- and three downregulated lncRNAs were confirmed by qRT-PCR. The remaining three upregulated lncRNAs did not show significant differences but did show changes in the expected direction. **(C)** The expression pattern of MEF2C is opposite to that of the nearby lncRNA LINC00461. The expression pattern of SLC20A1 is comparable to that of the neighboring lncRNA FLJ42351. Significance was calculated using the one-tailed t-test. Dashed line indicates the median.

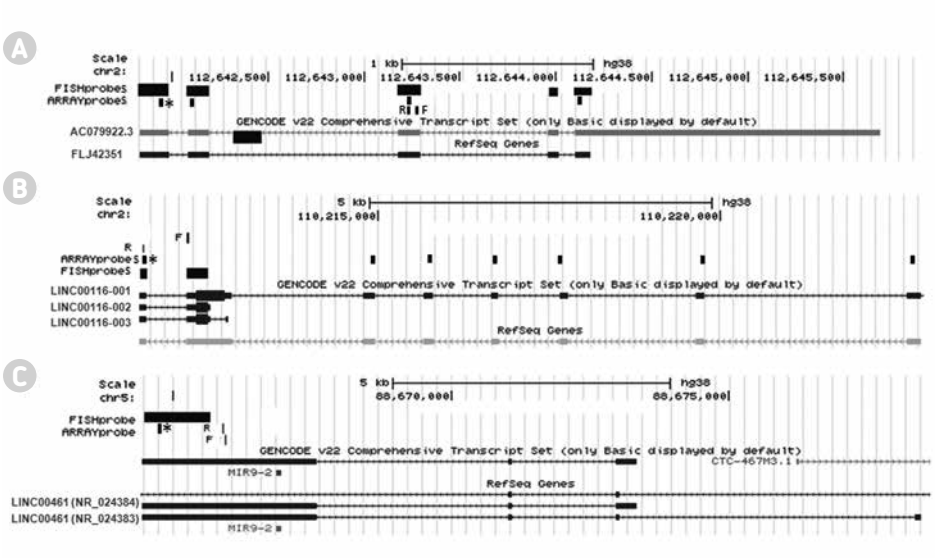

FIGURE S3 Overview of the three candidate lncRNAs selected for RNA-FISH. In the locus overviews the locations of the array probes, primer sets and RNA-FISH probes are given for **(A)** FLJ42351, **(B)** LINC00116 and **(C)** LINC00461. * indicates the array probe that was significant different between the cHL cell lines and GC-B cells. For FLJ42351, all the other indicated array probes were not differentially expressed and for LINC00116 all other probes did not show signals above the detection level.

TABLE S1 Primer sequences.

18S	F: 5'-CGGCTACCACATCCAAGGA-3' R: 5'-CCAATTACAGGGCCTCGAAA-3'
RPPH1	F: 5'-AGCTTGGAACAGACTCACGG-3' R: 5'-AATGGGCGGAGGAGAGTAGT-3'
U6	F: 5'-TGGAACGATACAGAGAAGATTAGCA-3' R: 5'-AAAATATGGAACGCTTCACGAATT-3'
U3 SNORNA	F: 5'-AACCCCGAGGAAGAGAGGTA-3' R: 5'-CACTCCCCAATACGGAGAGA-3'
ANRIL	F: 5'-AAGCCGCTCCGCTCCTCTTCT-3' R: 5'-GCCGTGTCCAGATGTCGCGT-3'
MIAT	F: 5'-TGGAGGCATCTGTCCACCCATGT-3' R: 5'-CCCTGTGATGCCGACGGGGT-3'
DANCR	F: 5'-CGTCTCTTACGTCTGCGGAA-3' R: 5'-TGGCTTGTGCCTGTAGTTGT-3'
MT-TK	F: 5'-CGGCTAGCTCAGTCGGTAGA-3' R: 5'-CCAACGTGGGGCTCGAAC-3'
XIST	F: 5'-GTCCTTTCTTTTGACCCCAGAA-3' R: 5'-GAGCCTGGCACTTTTTTTTCC-3'
TCONS_00002977	F: 5'-AAGCAGATGCTGTGCCTGATAC-3' R: 5'-TTCTCGACCCAGAAGCTCAAG-3'
TCONS_00025860	F: 5'-CGCGATTTGCAGGATTCC-3' R: 5'-CAAATGTGGGCACTTAAAAGCA-3'
TCONS_00007500	F: 5'-TGGGTTAACACTGCTTTTATGAGTTG-3' R: 5'-GCTGGCTCAGGAGTGAAGCT-3'
TCONS_00003368 (LINC00116)	F: 5'-CATGGCCGGCTCTTCCT-3' R: 5'-TCATAAAGTGCAAGAAGAAGTCATTTC-3'
TCONS _00029175	F: 5'-AGTTCATTCAACTGGTGATCTTAAGC-3' R: 5'-GCTGAGTCTACCTGGAGTCCATTATT-3'
TCONS_L2-00009699	F: 5'-GCTATTTTAAAAGGGTGTCCAG-3' R: 5'-CTGTACTAAGCCTCCCCCAGACT-3'
TCONS_L2_00012111	F: 5'-CTAAACCTCCTGCAAAAGTGGAA-3' R: 5'-TGTTTGCACTTTTTTGTCTGAAGAT-3'
TCONS _ L2_00015489	F: 5'-GGCTGCAGATGGCAGGATT-3' R: 5'-TGCTGTACAGATACACCACGGAAT-3'
FLJ42351 (TCONS_00002672)	F: 5'-TTGTGGCTCATGCCATATGAA-3' R: 5'-CAAAGATCCTGTGGGCAGTCA-3'
TCONS_L2_00030240	F: 5'-CACACTCCAAGGAAACGCAA-3' R: 5'-TGCTGTACAGATACACCACGGAAT-3'
TCONS_00021708	F: 5'-TGGATTTTACAGGCCCTCTTCA-3' R: 5'-GCTCCTGCCTCTGTTTTGCT-3'
LINC00461	F: 5'-CTTAAGCGCGGCAAGTATCC-3' R: 5'-GCCAGACTCCAGGTCCTGATC-3'
SLC20A1	F: 5'-CTGGCTCCGGTCCAAGAA-3' R: 5'-TGTGACAAACCAGGCCATAAAA-3'
MEF2C	F: 5'-CAAATGCAGGGCCCCTT-3' R: 5'-GATATGCACTTACTGAATTCCA-3'

TABLE S2 Table of putative *cis*-regulated lncRNA-mRNA pairs identified in the transcriptome comparisons of the normal B-cell subsets.

	lncRNA		mRNA		lncRNA to mRNA		
Chr	Probe name	Transcript(s)	Probe name	Gene Symbol	Distance (kb)	Orientation	Direction
1	CUST_2086_PI427622066	TCONS_I2_00001240, -41	A_23_P96872	HSPB11	-52	head-to-head	up / up
1	CUST_2688_PI427622066	TCONS_00001570~74 / TCONS_00002404 / TCONS_00000019	A_33_P3406030	LRRC8C	54	head-to-head	down / down
1	CUST_2655_PI427622068	TCONS_00002142	A_33_P3256685	TTF2	-2	tail-to-tail	up / up
2	CUST_7987_PI427622066	TCONS_00003683	A_33_P3258472	SPTBN1	-38	sense	down / down
2	**CUST_8107_PI427622066**	**TCONS_00003301**	**A_33_P3316928**	**PELI1**	**6**	**sense**	**up/up**
2	CUST_4864_PI427622068	TCONS_00003306, -7 / TCONS_00003712~4	A_24_P191312	SLC1A4	-8	sense	up / up
2	CUST_7020_PI427622068	TCONS-00004836	A_23_P136573	ST3GAL5	-49	head-to-head	down / down
3	CUST_10529_PI427622068	TCONS-00006023	A_23_P314584	MAPKAPK3	-22	sense	down / down
3	**CUST_13544_PI427622066**	**TCONS-00006602**	**A_23_P6935**	**CD47**	**20**	**sense**	**down/down**
3	CUST_13581_PI427622066	TCONS_I2_00018738	A_23_P121480	CD200	55	sense	up / down
3	A_33_P3368771	NR_026991.1 / H1FX-AS1	A_23_P96087	H1FX	-2	antisense	up / up
4	A_33_P3447441	TCONS_00007979	A_33_P3360426	WDR1	29	tail-to-tail	down / up
4	CUST_14305_PI427622068	TCONS_00008144	A_24_P183128	PLAC8	3	tail-to-tail	down / down
5	CUST_19589_PI427622066	TCONS_00010926 / TCONS_00010320, -1 / TCONS_00009277	A_33_P3299279	C5orf39	53	sense	down / down
5	CUST_21197_PI427622066	TCONS_00009935	A_33_P3299279	C5orf39	43	sense	down / down
5	A_33_P3331282	NR_034085.1 / LOC648987	A_33_P3299279	C5orf39	23	sense	down / down
6	CUST_24388_PI427622066	TCONS_00011697	A_32_P52018	PHACTR1	-48	sense	up / down
6	CUST_24782_PI427622066	TCONS_I2_00025363	A_33_P3379939	HLA-F	-25	sense	down / down
6	CUST_20150_PI427622068	TCONS_00011294	A_24_P288836	HLA-DPB2	-19	sense	down / down
7	A_33_P3249674	NR_131935.1 / LOC541472	A_23_P71037	IL6	5	antisense	down / down
7	CUST_22947_PI427622068	TCONS_I2_00025783	A_23_P93750	LSM5	15	tail-to-tail	up / up
7	CUST_23831_PI427622068	TCONS_00014161	A_23_P93823	RFC2	-19	head-to-head	up / up
7	CUST_24079_PI427622068	TCONS_I2_00026087, 8	A_24_P406334	STEAP1	36	sense	up / down
9	CUST_34071_PI427622066	TCONS_00015807	A_23_P60225	GRHPR	45	head-to-head	up / up
10	A_33_P3645465	NR_026932.1 / PDCD4-AS1	A_33_P3212092	PDCD4	30	head-to-head	down / down
11	CUST_41848_PI427622066	TCONS_00019608	A_23_P385126	DEPDC7	29	head-to-head	down / up
11	CUST_34279_PI427622068	TCONS_I2_00004565~7 / TCONS_I2_00004572, -3, -6, -7	A_24_P298360	LTBP3	37	sense	down / down
11	**CUST_34183_PI427622068**	**TCONS_I2_00004704**	**A_33_P3268472**	**CTSC**	**-9**	**head-to-head**	**down/down**
12	CUST_36487_PI427622068	TCONS_00021304 / TCONS_00020768 / TCONS_00020391	A_23_P218079	SLC38A2	-23	sense / head-to-head	down / down

12	CUST_43855_PI427622066	TCONS_00020810, -2, -3	A_23_P53363	XRCC6BP1	16	head-to-head	up/up
12	A_33_P3210363	NR_027157.1 / TMPO-AS1	A_23_P325040	TMPO	21	antisense	up/up
12	CUST_37379_PI427622068	TCONS_00020238	A_33_P3263867	P2RX7	-3	sense	down/down
12	A_23_P116743	NR_002809.2 / LINC01089	A_23_P319895	SETD1B	36	head-to-head	down/down
12	**CUST_45746_PI427622066**	**NR_110375.1 / THRIL**	**A_32_P831181**	**BRI3BP**	**-14**	**tail-to-tail**	**up/up**
12	CUST_37427_PI427622068	TCONS_00020615	A_23_P204395	AACS	-8	sense	up/up
13	CUST_46932_PI427622066	TCONS_I2_00007498	A_32_P219116	CENPJ	-43	sense	up/up
14	CUST_40877_PI427622068	TCONS_00023060 / TCONS_00022650~54	A_23_P128808	GPR132	-46	head-to-head	down/down
15	CUST_41915_PI427622068	TCONS_00023374, -5 / TCONS_00023923	A_33_P3224105	C15orf23	-41	sense	down/up
15	CUST_51373_PI427622066	TCONS_00023935	A_24_P160874	DUT	21	sense	up/up
17	CUST_54932_PI427622066	TCONS_I2_00011115	A_23_P84705	TNFRSF13B	23	sense	down/down
17	CUST_45687_PI427622068	TCONS_00025383 / TCONS_00025120	A_23_P343398	CCR7	26	tail-to-tail	down/down
19	CUST_48607_PI427622068	TCONSI2_00012645	A_33_P3382498	SIGLEC14	39	tail-to-tail	down/down
19	A_33_P3501900	TCONS_00026798, -9 / TCONS_00027118~21, -23, -26	A_23_P67432	ZNF583	24	sense	down/down
20	CUST_60816_PI427622066	TCONS_00028695 / TCONS_00028061	A_33_P3284029	CSE1L	54	head-to-head	up/up
20	**CUST_50118_PI427622068**	**TCONS_00027871**	**A_24_P91566**	**BMP7**	**-54**	**head-to-head**	**up/up**
20	CUST_48971_PI427622068	TCONS_00027984	A_24_P678104	STMN3	11	tail-to-tail	down/down
21	CUST_50998_PI427622068	TCONS_00029009	A_23_P325501	MORC3	-8	sense	down/down
21	CUST_50998_PI427622068	TCONS_00029009	A_23_P57306	CHAF1B	31	sense	down/up
21	CUST_50488_PI427622068	TCONS_00029038, -9	A_23_P143535	WDR4	-29	head-to-head	up/up
22	CUST_52438_PI427622068	TCONS_00029808, -9	A_23_P132316	APOBEC3D	-37	sense	down/down
X	CUST_30332_PI427622068	TCONS_00017181	A_23_P217528	KLF8	-5	sense	down/down

* Probe to probe distance in kb, positive distance means the lncRNA probe is upstream of the protein-coding gene, negative distance indicates the reverse.

** The direction of the lncRNA is indicated first, the direction of the mRNA is indicated second. Direction is based on GC-B cells compared to naïve/memory B cells.

In bold the lncRNA-mRNA pairs are indicated that overlap with the putative cis-regulated lncRNA-mRNA pairs identified in the transcriptome comparisons of the cHL cell lines with the GC-B cell subsets (Table I).