A comprehensive inventory of TLX1 controlled long non-coding RNAs in T-cell acute lymphoblastic leukemia through polyA+ and total RNA sequencing

Karen Verboom,[1,2] Wouter Van Loocke,[1,2] Pieter-Jan Volders,[1-4] Bieke Decaesteker,[1,2] Francisco Avila Cobos,[1,2,4] Simon Bornschein,[5,6] Charles E. de Bock,[5,6] Zeynep Kalender Atak,[5,7] Emmanuelle Clappier,[8] Stein Aerts,[5,7] Jan Cools,[5,6] Jean Soulier,[8] Tom Taghon,[2,9] Pieter Van Vlierberghe,[1,2] Jo Vandesompele,[1,2,4] Frank Speleman[1,2] and Kaat Durinck[1,2]

[1]Center for Medical Genetics, Ghent University, Belgium; [2]Cancer Research Institute Ghent, Belgium; [3]Center for Medical Biotechnology, VIB-UGent, Ghent, Belgium; [4]Bioinformatics Institute Ghent from Nucleotides to Networks, BIG N2N, Belgium; [5]KU Leuven Center for Human Genetics, Belgium; [6]VIB Center for Cancer Biology, Leuven, Belgium; [7]VIB Center for Brain & Disease Research, Laboratory of Computational Biology, Leuven, Belgium; [8]Hôpital Saint Louis, Institut Universitaire d'Hématologie, Paris, France and [9]Department of Clinical Chemistry, Microbiology and Immunology, Ghent University, Belgium

Data availability: GSE110637.

Correspondence: kaat.durinck@ugent.be
doi:10.3324/haematol.2018.190587

Online supplementary methods

Cell lines

The TLX1 positive cell line ALL-SIL was obtained from the DSMZ cell line repository. Cells were maintained in RPMI-1640 medium (Life Technologies, 52400-025) supplemented with 20 % fetal bovine serum, 1 % of L-glutamine (Life Technologies, 15140-148) and 1 % penicillin/streptomycin (Life Technologies, 15160-047).

siRNA mediated knockdown in ALL-SIL lymphoblasts

The RNA isolated after *TLX1* knockdown that has been used for microarray based gene expression profiling (Agilent SurePrint G3, 8x60k) in the previous study has been used in this study for RNA-seq.[1] In short, ALL-SIL cells were electroporated (250 V, 1000 µF) using a Genepulser Xcell device (Bio-Rad) with 400 nM of Silencer Select Negative Control 1 siRNA (Ambion, #AM4635) or siRNAs targeting TLX1 (Silencer Select, Ambion, Carlsbad, CA, USA; #4392420, s6746 (siRNA1) and s6747 (siRNA2)). ALL-SIL cells were collected 24h post electroporation.

Compound treatment in ALL-SIL lymphoblasts

ALL-SIL cells were seeded at a density of $1x10^6$ cells/ml and treated with either DMSO or 1 µM of JQ1 compound (BPS Bioscience, 27401). Cells were harvested 12h post treatment for RNA isolation.

Clinical samples

Bone marrow lymphoblasts from 60 pediatric and adult T-ALL patients (13 immature, 23 *TALR*, 17 *TLX1/TLX3* and 7 *HOXA*) were collected with informed consent according to the declaration of Helsinki from Saint-Louis Hospital (Paris, France) and the study was approved by the Institut Universitaire d'Hématologie Institutional Review Board. This primary T-ALL cohort has previously been investigated[2] and these RNA samples were used for RNA-seq in this study.

RNA isolation

Total RNA was isolated using the miRNeasy mini kit (Qiagen) with DNA digestion on-column. By means of spectrophotometry, RNA concentrations were measured (Nanodrop 1000, Thermo Scientific).

PolyA+ RNA sequencing

Polyadenylated transcripts of the *TLX1* knockdown samples were sequenced using the TruSeq stranded mRNA sample preparation kit (Illumina). The libraries were quantified using the KAPA library quantification kit (Illumina) and samples were paired-end sequenced on the NextSeq 500 sequencer (Illumina) with a read length of 75 bp. The sequencing depth per

sample is shown in **Supplementary Figure 1A (*left*)**. PolyA+ RNA-seq of the 60 primary T-ALL samples was performed using 100 ng of RNA as input material by Biogazelle (Belgium) with the TruSeq stranded mRNA sample preparation kit (Illumina). The samples were quantified using the KAPA library quantification kit and paired-end sequenced on a NextSeq 500 sequencer with a read length of 75 bp. The sequencing depth per sample is shown in **Supplementary Figure 9A (*left*)**. Dll1 data generated in the context of the paper of Durinck et al. were used as a reference dataset (GSE62006).[3]

Total RNA sequencing

Total RNA-seq of the knockdown samples and 25 primary T-ALL samples (5 immature, 5 *TALR*, 10 *TLX1/TLX3* and 5 *HOXA*) was performed using the TruSeq Stranded Total RNA (w/RiboZero Gold) sample prep kit (Illumina), involving depletion of ribosomal (rRNA) transcripts by Biogazelle (Belgium). Libraries were quantified using the Qubit 2.0 Fluorometer and paired-end sequenced on a NextSeq 500 sequencer (Illumina) with a read length of 75 bp. The sequencing depth per sample is shown in **Supplementary Figure 1A (*right*), 9A (*right*).** Total RNA-seq of the JQ1 treated ALL-SIL cells was performed using the TruSeq Stranded Total RNA sample prep kit (Illumina). The libraries were quantified using the KAPA library quantification kit and were single-end sequenced on a NextSeq 500 sequencer with a read length of 75 bp and an average sequencing depth of 38 million reads per sample.

Data processing RNA sequencing

FastQC was used for quality control of fastq files (available online at http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and RSeQC[4] was used for calculating read distribution over genomic features. All samples were aligned against GRCh38 with STAR_2.4.2a[5] using default settings and two pass methods as described in the STAR manual, using GENCODE v25 primary annotation as a guide during the first pass and the combined splice junction database of all samples generated in the first pass as a guide in the second step. Genes were quantified on GENCODE v25 during alignment with STAR. LncRNAs were defined as genes with biotype “lincRNA” or “antisense” (GENCODE v25). StringTie-1.3.3b was used for transcript assembly and transcriptome gtf files were subsequently merged with StringTie merge setting[6,7] to generate a transcriptome containing all detected transcripts over all samples. The merged file was subsequently used to quantify transcripts using HTSeq-0.6.1.[8] Differential gene expression was performed in R based on a negative binomial distribution using DESeq2.[9] Reported adjusted p-values were calculated using Wald statistic. IGV-tools[10] count was used on BAM files to generate a visualisation track for IGV.[11]

(Tumor suppressor) genes in the neighborhood of lncRNAs

BEDTools[12] was used to identify genes within a certain range from a selected lncRNA.

Protein-coding potential calculations

PhyloCSF[13] and the Coding-Potential Assessment Tool[14] (CPAT, version 1.2.2) were used to identify putative protein coding transcripts in the unannotated genes obtained by RNA-seq. PhyloCSF employs codon substitution frequencies in whole-genome multi-species alignments to distinguish between coding and non-coding loci. Multiple alignments of 45 vertebrate genomes with human (hg19) are obtained from the UCSC website (multiz46way) and processed using the PHAST package (version 1.3) to obtain the required input format for PhyloCSF. To validate our workflow and obtain the optimal threshold for the PhyloCSF score, we benchmarked PhyloCSF with transcripts annotated in RefSeq.[15] Alignments in the BED format of 5859 RefSeq lncRNAs and 6051 RefSeq mRNAs were obtained from the UCSC genome browser website to serve as negative and positive sets, respectively. After ROC analysis, an optimal threshold for the PhyloCSF score of 60.7876 was found. This corresponds to a sensitivity and specificity of 92.75 %. For CPAT, the transcript sequences were provided in the FASTA format. The hexamer frequency table and logit model provided with the algorithm were used. We used the published cutoff of 0.364 as a threshold for the CPAT output. Bed files were converted to hg19 using LiftOver (UCSC).

Guilt-by-association analysis

Normalized counts were generated for the samples of the primary T-ALL cohort (polyA+, n=60). Spearman correlations were calculated for the top 5 TLX1 up and down regulated lncRNAs. The output was used as input for a GSEA pre-ranked analysis[16] using the c5.bp.v6.0 (GO biological processes) gene set. Next, the output was used as input for the Cytoscape plugin 'Enrichment map' to create networks of enriched gene set clusters. Functional gene set clusters that are correlated (blue nodes) and anti-correlated (red nodes) with the lncRNA of interest are depicted in these networks.

Chromatin immunoprecipitation sequencing

Our previously generated TLX1 and H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) data generated in the context of the paper of Durinck et al. were used (GSE62144).[1] H3K4me1 and H3K4me3 chromatin immunoprecipitation sequencing (ChIP-seq) were performed as previously described with minor changes.[17] In brief, $1x10^7$ cells were cross-linked with 1.1 % formaldehyde (Sigma-Aldrich, F1635) at room temperature for 10 min and the cross-linking reaction was quenched with glycine (125 mM final concentration, Sigma-Aldrich, G-8790). Nuclei were isolated and chromatin was purified by chemical lysis. Next, the purified chromatin was fragmented to 200-300 bp fragments by sonication (Covaris, M220, Focused-ultrasonicator). Chromatin immunoprecipitation was performed by incubation of the chromatin fraction overnight with 20 µl of protein-A coated beads (Thermo-Scientific, 53139) and 2 µg of H3K4me1 specific (Abcam, ab8895) or H3K4me3 specific (Abcam, ab8580) antibody. The next day, beads were washed to remove non-specific binding events and enriched chromatin fragments were eluted from the beads, followed by reverse cross-linking by incubation at 65 °C overnight. DNA was subsequently purified by phenol/chloroform extraction, assisted by phase lock gel tubes (5Prime). DNA obtained from the ChIP-assays was adaptor-ligated, amplified and quantified using the KAPA library quantification kit. The

libraries were single-end sequenced on a NextSeq 500 sequencer with a read length of 75 bp and an average sequencing depth of 35 million reads.

Assay for transposase accessible chromatin sequencing

Assay for transposase accessible chromatin sequencing (ATAC-seq) was performed as previously described with minor changes.[18] In short, 50,000 cells were lysed and fragmented using Tn5 transposase (Illumina). Next, the samples were purified using the MinElute kit (Qiagen). The transposased DNA fragments were amplified and purified using the PCR Cleanup kit (Qiagen). Samples were quantified using the KAPA library quantification kit and single-end sequenced on a NextSeq 500 with a read length of 75 bp and an average sequencing read depth of 50 million reads.

Data processing ChIP and ATAC sequencing

All ChIP-seq files were aligned using STAR_2.4.2a[5] with --outFilterMultimapNmax 1 to exclude multimapping reads and --alignIntronMax 1 to turn off splice awareness. Peak calling was subsequently performed with MACS2[19] using input samples as control. Peak calling for histone marks was performed using --broad setting. BEDTools[12] was used to find overlaps between ChIP-seq tracks. All plots were generated in R using ggplot2. Any additional data manipulation was performed in R. IGV-tools[10] count was used on BAM files to generate a visualisation track for IGV.[11]

Visualization of distances between transcription start sites and closest ATAC-seq peaks

Based on the distribution of gene expression values from polyA+ RNA-seq performed on ALL-SIL lymphoblasts, we randomly selected 1000 genes belonging to the lowest decile, 1000 genes with expression levels between the 40 % and 60 % percentile and 1000 genes with expression values in the highest decile. Next, we retrieved the transcription start sites (TSSs) of those genes and built three Zipper plots showing the distribution of distances between the TSSs and the closest ATAC-seq peaks in a 5 kb window from the TSS.[20]

Motif analysis

MEME-ChIP[21] was used to find centrally enriched TF motifs in ChIP-seq peaks. To prepare peak files for input, peaks were adjusted to 500 bp centered around the peak summit.

SE calling

Super-enhancer analysis was performed using ROSE[22,23] with H3K27ac alignment file and input as control, excluding TSSs (+-2500 bp). BEDTools[12] was used to assign genes within 100 kb to the super-enhancer regions.

Statistical analyses

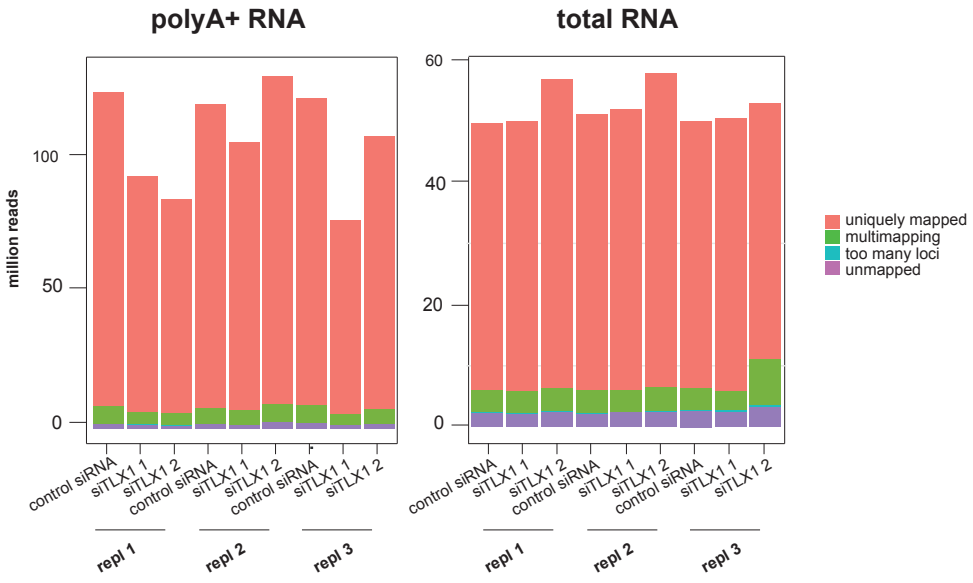
Statistical significance (pAdj.<0.05) of differences between conditions was determined by Fischer exact test using R package.

1. Durinck K, Van Loocke W, Van der Meulen J, et al. Characterization of the genome-wide TLX1 binding profile in T-cell acute lymphoblastic leukemia. Leukemia. 2015;29(12):2317–2327.
2. Wallaert A, Durinck K, Van Loocke W, et al. Long noncoding RNA signatures define oncogenic subtypes in T-cell acute lymphoblastic leukemia. Leukemia. 2016;30(9):1927–1930.
3. Durinck K, Wallaert A, Van de Walle I, et al. The notch driven long non-coding RNA repertoire in T-cell acute lymphoblastic leukemia. Haematologica. 2014;99(12):1808–1816.
4. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics. 2012;28(16):2184–2185.
5. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
6. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290–295.
7. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016;11(9):1650–1667.
8. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166–9.
9. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
10. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–92.
11. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–6.
12. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–842.
13. Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics. 2011;27(13):i275–i282.
14. Wang L, Park HJ, Dasari S, Wang S, Kocher J-P, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. 2013;41(6):e74.
15. Pruitt KD, Brown GR, Hiatt SM, et al. RefSeq: an update on mammalian reference sequences. Nucleic Acids Res. 2014;42(D1):D756–D763.
16. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545–50.
17. Lee TI, Johnstone SE, Young RA. Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc. 2006;1(2):729–48.
18. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Current Protocols in Molecular Biology.

2015;119:21.29.1-21.29.9.
19. Zhang Y, Liu T, Meyer CA, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.
20. Avila Cobos F, Anckaert J, Volders P-J, et al. Zipper plot: visualizing transcriptional activity of genomic regions. BMC Bioinformatics. 2017;18(1):231.
21. Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37:W202-8.
22. Whyte WA, Orlando DA, Hnisz D, et al. Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell. 2013;153(2):307–319.
23. Lovén J, Hoke HA, Lin CY, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell. 2013;153(2):320–34.

Figure S1

A


polyA+ RNA
total RNA
million reads
100
50
0
60
40
20
0
uniquely mapped
multimapping
too many loci
unmapped
control siRNA
siTLX1 1
siTLX1 2
repl 1
repl 2
repl 3

B

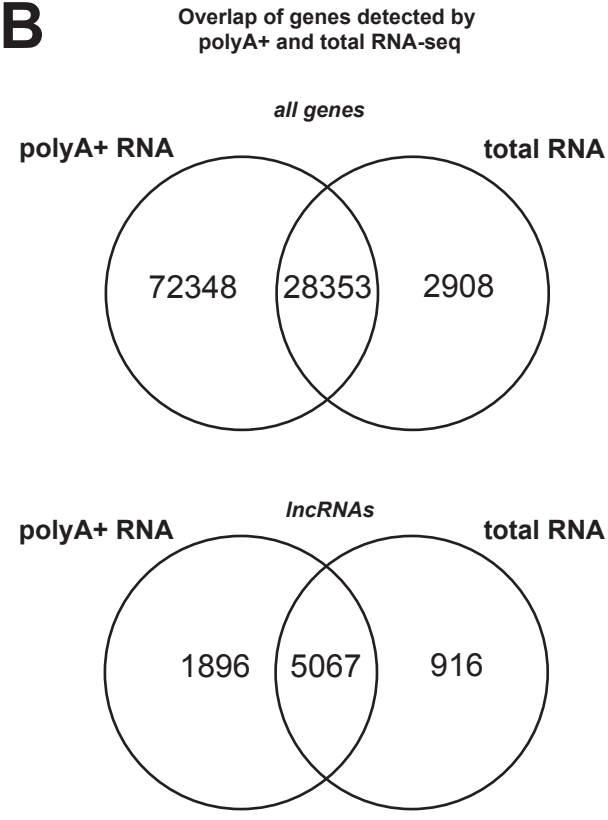
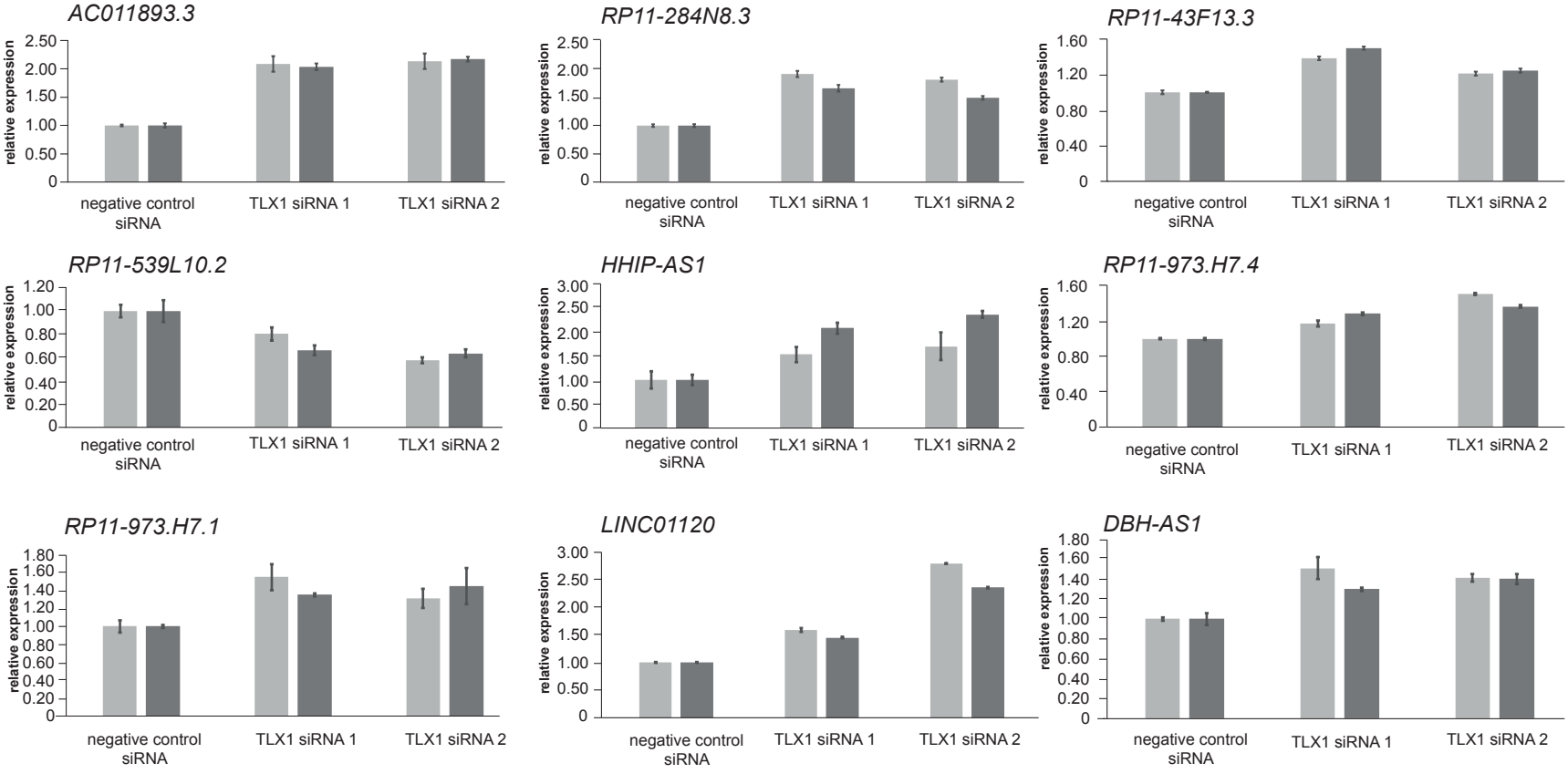

Overlap of genes detected by polyA+ and total RNA-seq
all genes
polyA+ RNA
total RNA
72348
28353
2908
lncRNAs
polyA+ RNA
total RNA
1896
5067
916

Figure S2

A


AC011893.3
RP11-284N8.3
RP11-43F13.3
RP11-539L10.2
HHIP-AS1
RP11-973.H7.4
RP11-973.H7.1
LINC01120
DBH-AS1
relative expression
negative control siRNA
TLX1 siRNA 1
TLX1 siRNA 2

B

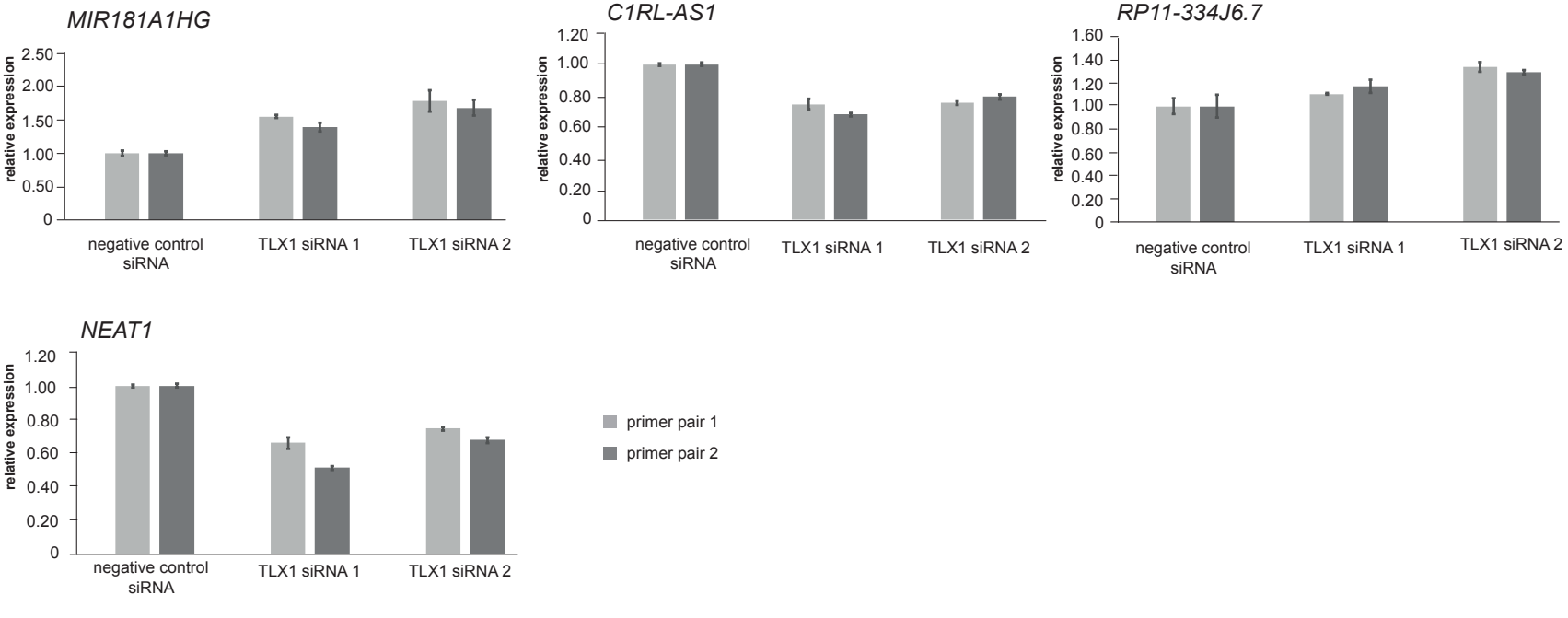
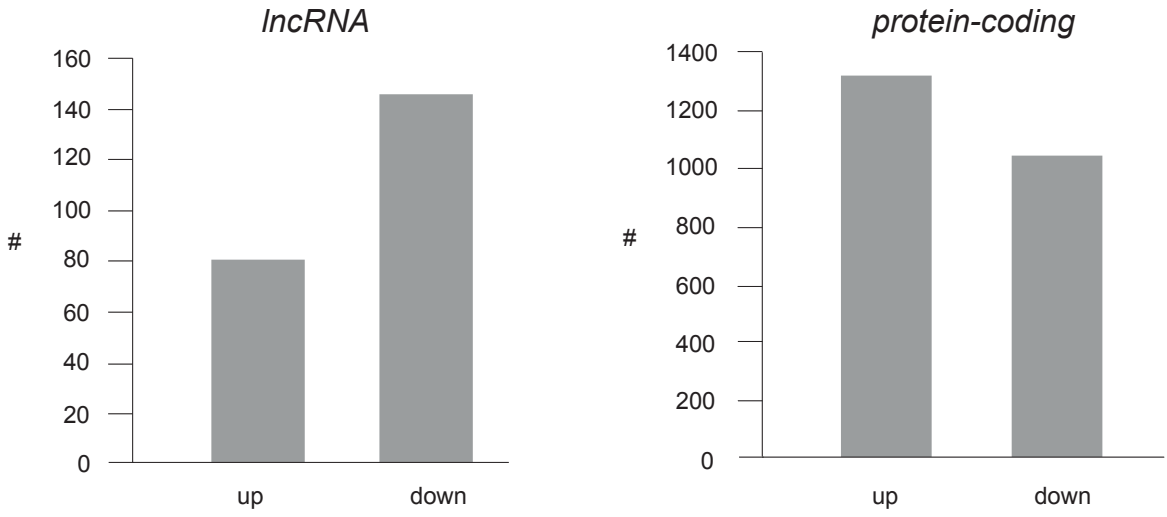

MIR181A1HG
C1RL-AS1
RP11-334J6.7
NEAT1
relative expression
negative control siRNA
TLX1 siRNA 1
TLX1 siRNA 2
primer pair 1
primer pair 2

Figure S3

A


lncRNA
160
140
120
100
80
60
40
20
0
#
up
down
protein-coding
1400
1200
1000
800
600
400
200
0
#
up
down

B

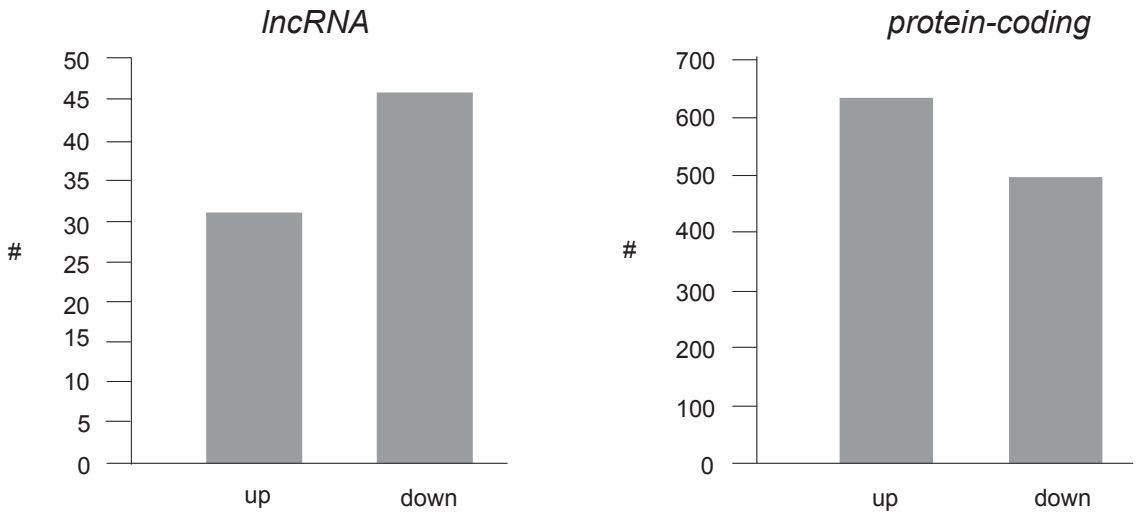

lncRNA
50
45
40
35
30
25
20
15
10
5
0
#
up
down
protein-coding
700
600
500
400
300
200
100
0
#
up
down

Figure S4

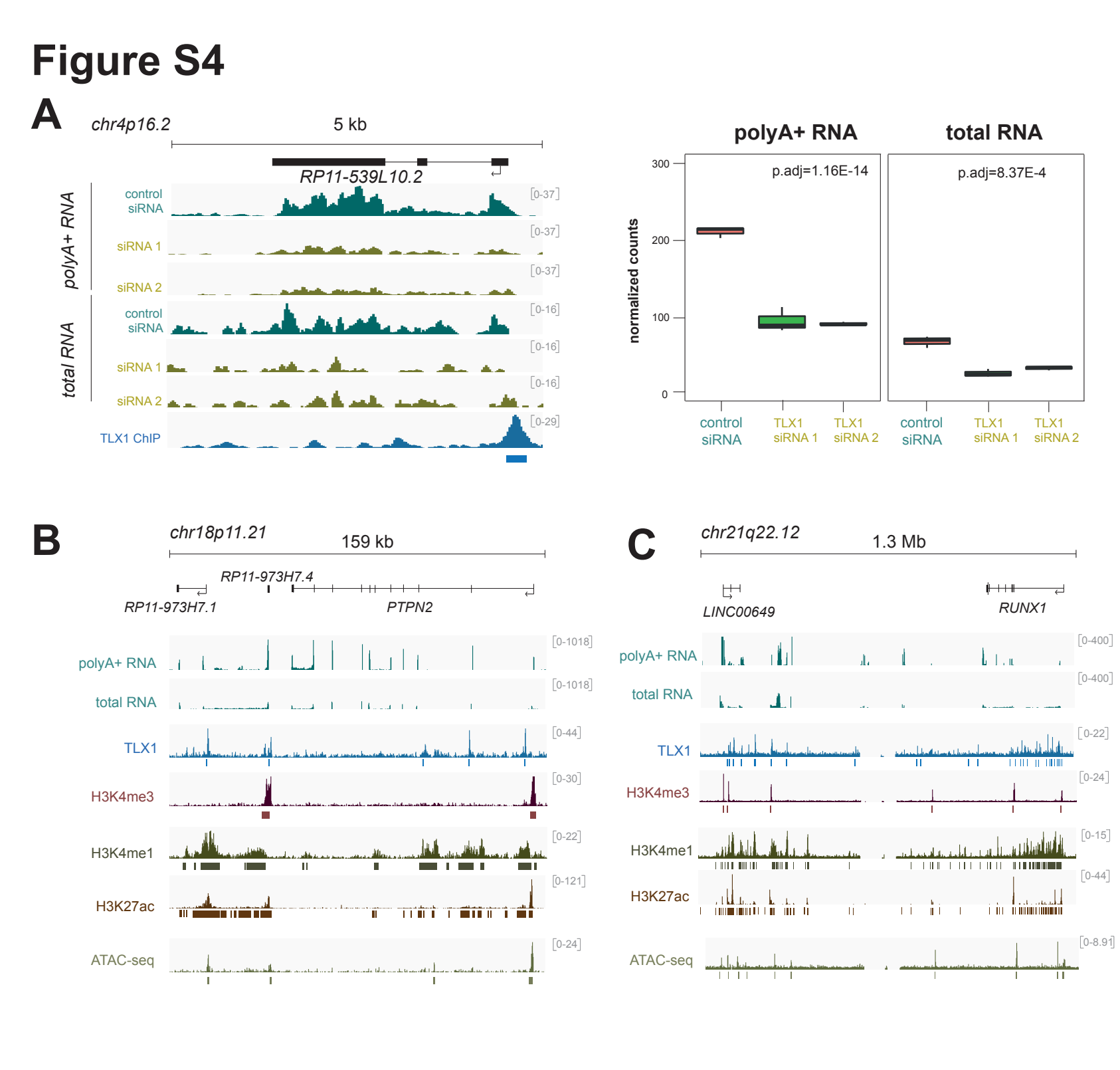

A
chr4p16.2
5 kb
RP11-539L10.2
polyA+ RNA
total RNA
control siRNA
siRNA 1
siRNA 2
TLX1 ChIP
[0-37]
[0-16]
[0-29]
polyA+ RNA
total RNA
p.adj=1.16E-14
p.adj=8.37E-4
normalized counts
300
200
100
0
control siRNA
TLX1 siRNA 1
TLX1 siRNA 2
B
chr18p11.21
159 kb
RP11-973H7.4
RP11-973H7.1
PTPN2
polyA+ RNA
total RNA
TLX1
H3K4me3
H3K4me1
H3K27ac
ATAC-seq
[0-1018]
[0-44]
[0-30]
[0-22]
[0-121]
[0-24]
C
chr21q22.12
1.3 Mb
LINC00649
RUNX1
polyA+ RNA
total RNA
TLX1
H3K4me3
H3K4me1
H3K27ac
ATAC-seq
[0-400]
[0-22]
[0-24]
[0-15]
[0-44]
[0-8.91]

Figure S5

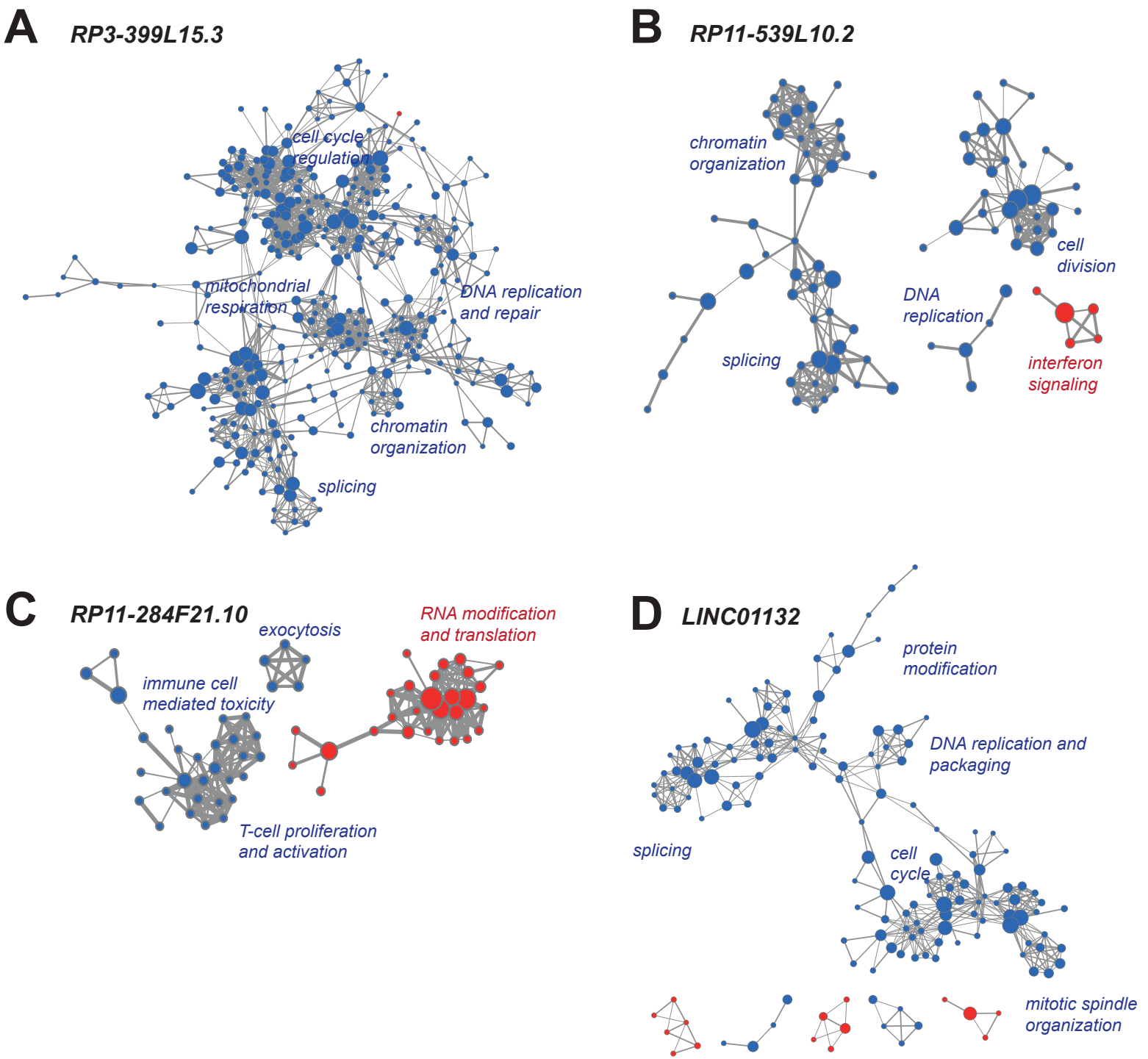

A
RP3-399L15.3
cell cycle regulation
mitochondrial respiration
DNA replication and repair
chromatin organization
splicing
B
RP11-539L10.2
chromatin organization
cell division
DNA replication
splicing
interferon signaling
C
RP11-284F21.10
exocytosis
RNA modification and translation
immune cell mediated toxicity
T-cell proliferation and activation
D
LINC01132
protein modification
DNA replication and packaging
splicing
cell cycle
mitotic spindle organization

Figure S6

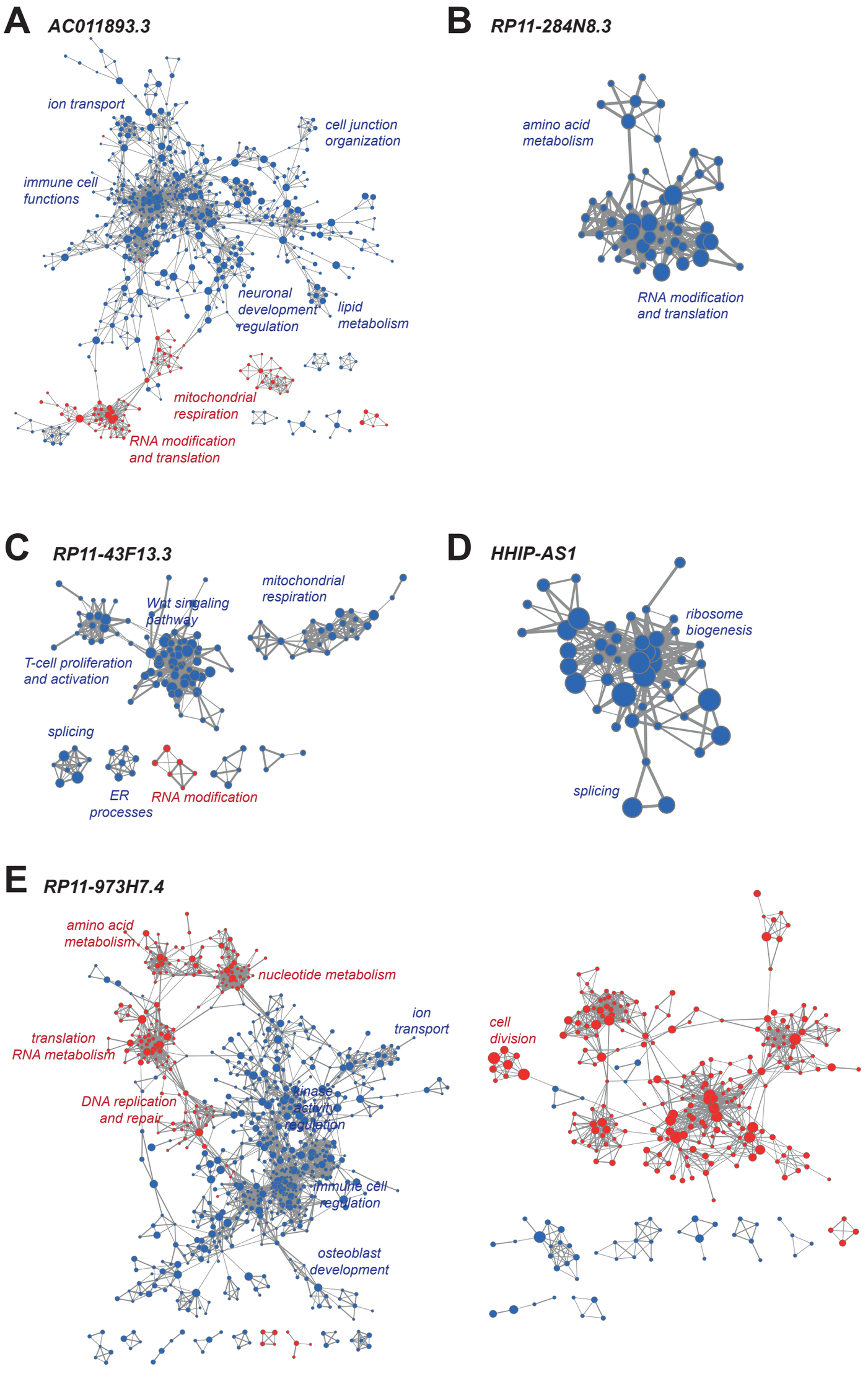

A
AC011893.3
ion transport
cell junction organization
immune cell functions
neuronal development regulation
lipid metabolism
mitochondrial respiration
RNA modification and translation
B
RP11-284N8.3
amino acid metabolism
RNA modification and translation
C
RP11-43F13.3
Wnt singaling pathway
mitochondrial respiration
T-cell proliferation and activation
splicing
ER processes
RNA modification
D
HHIP-AS1
ribosome biogenesis
splicing
E
RP11-973H7.4
amino acid metabolism
nucleotide metabolism
ion transport
translation RNA metabolism
DNA replication and repair
kinase activity regulation
immune cell regulation
osteoblast development
cell division

Figure S7

A

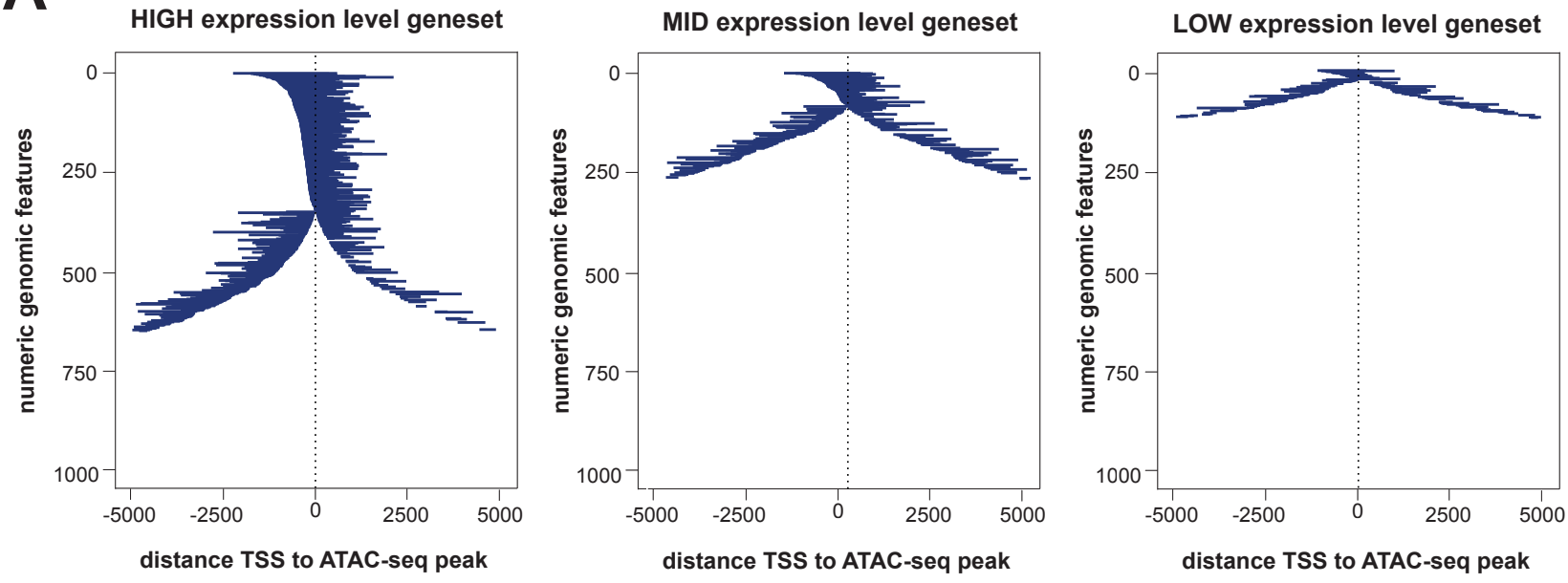
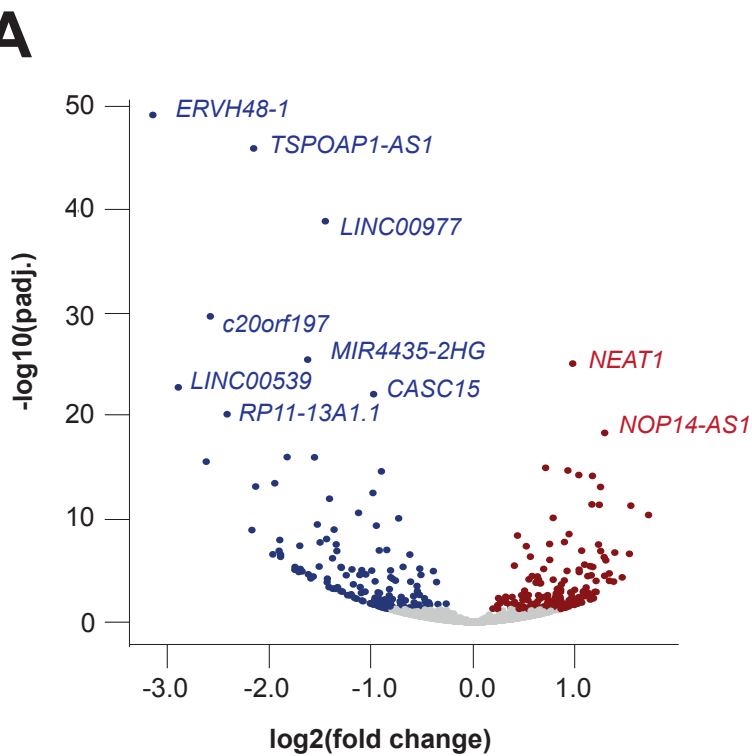

HIGH expression level geneset
MID expression level geneset
LOW expression level geneset
numeric genomic features
0
250
500
750
1000
-5000
-2500
0
2500
5000
distance TSS to ATAC-seq peak

Figure S8


A
50
40
30
20
10
0
-log10(padj.)
ERVH48-1
TSPOAP1-AS1
LINC00977
c20orf197
MIR4435-2HG
LINC00539
CASC15
RP11-13A1.1
NEAT1
NOP14-AS1
-3.0
-2.0
-1.0
0.0
1.0
log2(fold change)

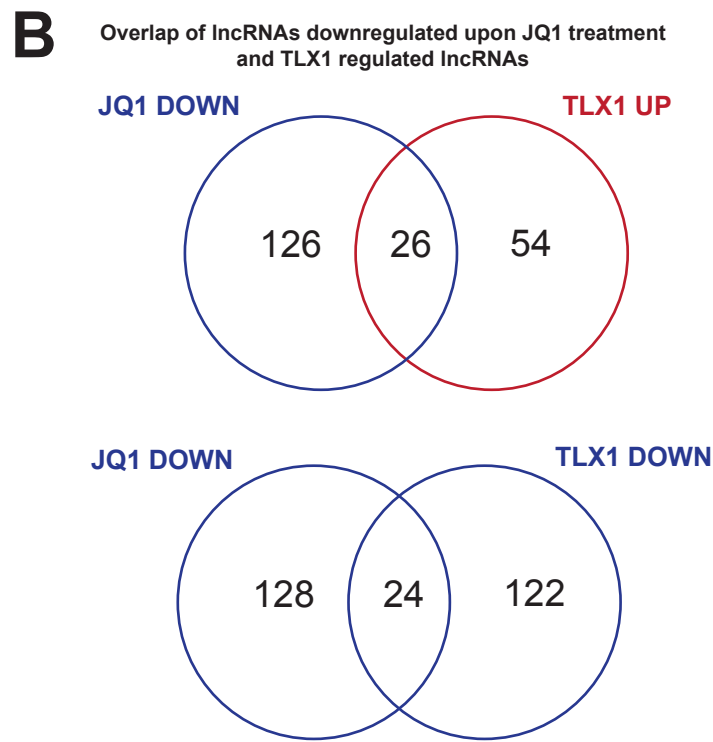
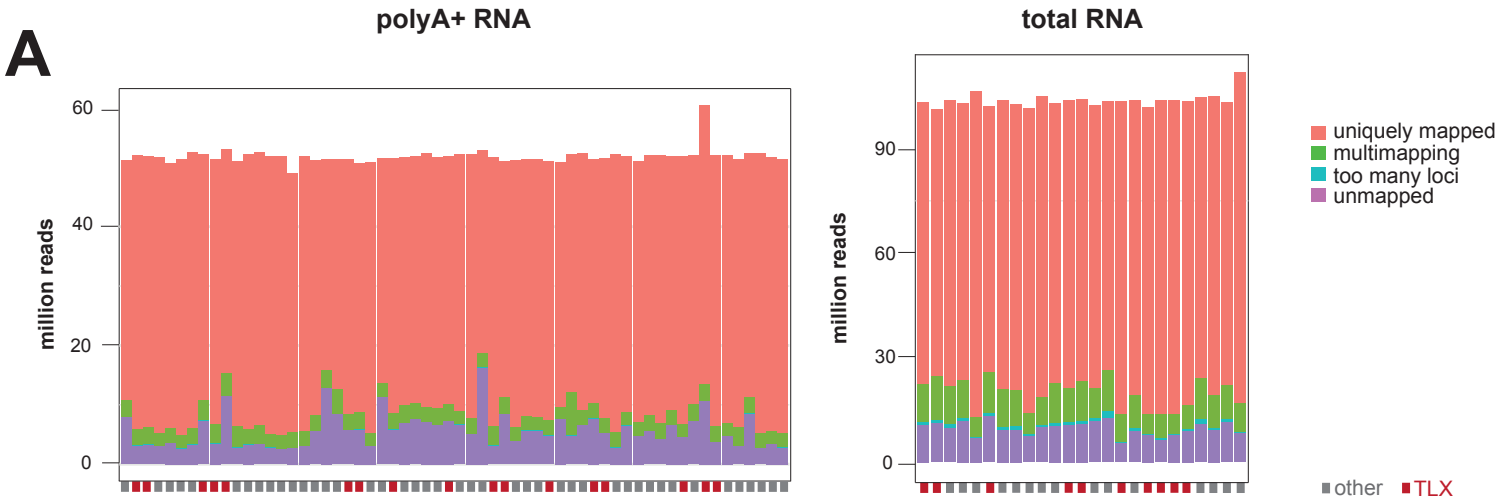

B
Overlap of lncRNAs downregulated upon JQ1 treatment
and TLX1 regulated lncRNAs
JQ1 DOWN
TLX1 UP
126
26
54
JQ1 DOWN
TLX1 DOWN
128
24
122

Figure S9


A
polyA+ RNA
total RNA
million reads
60
40
20
0
90
60
30
0
uniquely mapped
multimapping
too many loci
unmapped
other
TLX

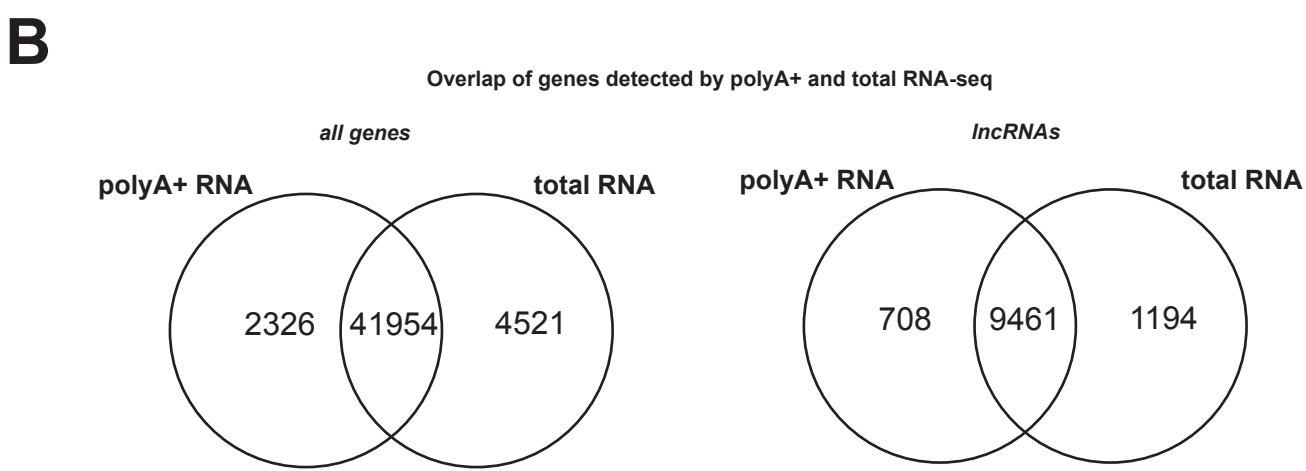
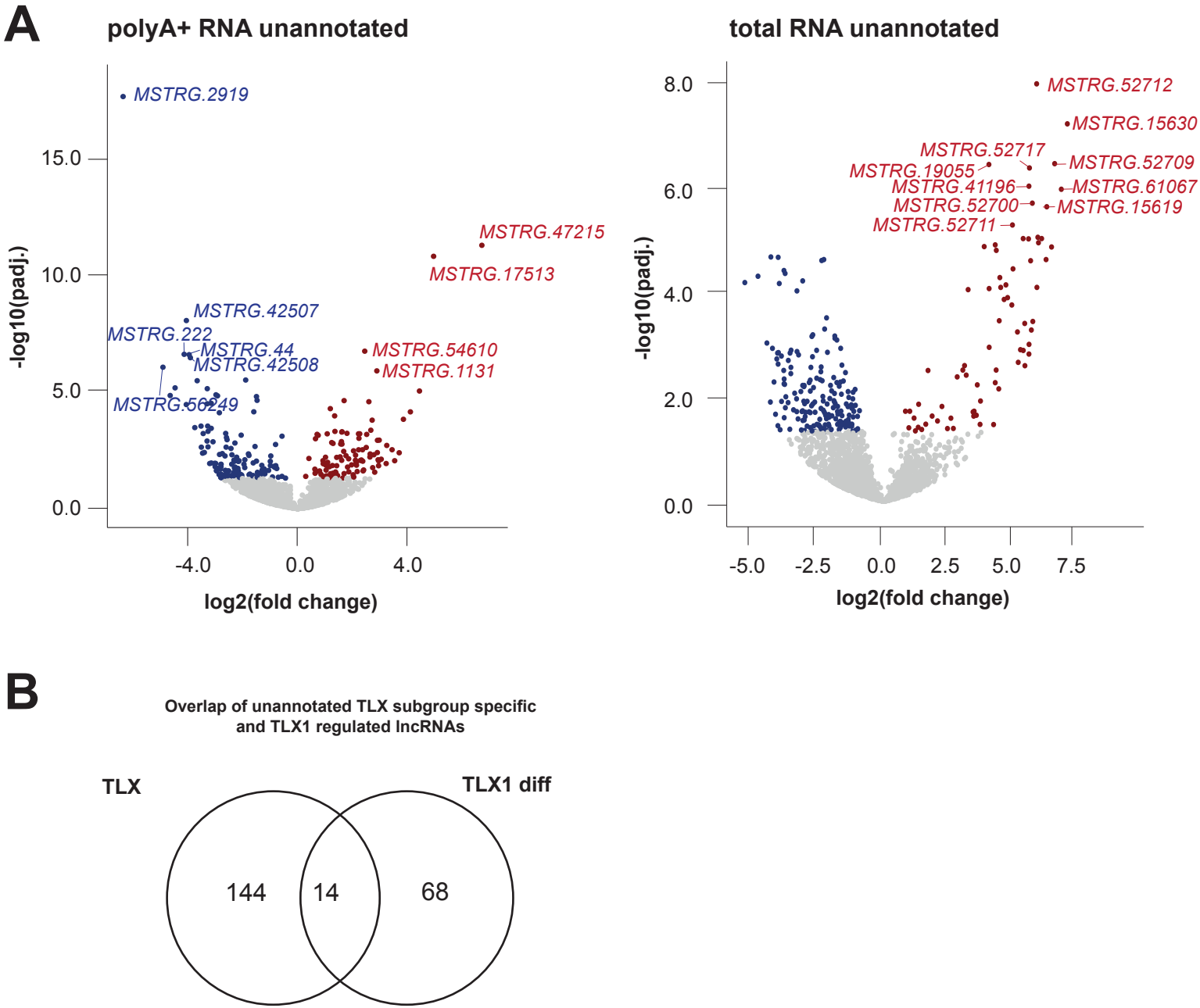

B
Overlap of genes detected by polyA+ and total RNA-seq
all genes
polyA+ RNA
total RNA
2326
41954
4521
lncRNAs
polyA+ RNA
total RNA
708
9461
1194

Figure S10


A
polyA+ RNA unannotated
MSTRG.2919
MSTRG.47215
MSTRG.17513
MSTRG.42507
MSTRG.222
MSTRG.44
MSTRG.42508
MSTRG.54610
MSTRG.1131
MSTRG.56249
-log10(padj.)
15.0
10.0
5.0
0.0
-4.0
0.0
4.0
log2(fold change)
total RNA unannotated
MSTRG.52712
MSTRG.15630
MSTRG.52717
MSTRG.19055
MSTRG.52709
MSTRG.41196
MSTRG.61067
MSTRG.52700
MSTRG.15619
MSTRG.52711
-log10(padj.)
8.0
6.0
4.0
2.0
0.0
-5.0
-2.5
0.0
2.5
5.0
7.5
log2(fold change)
B
Overlap of unannotated TLX subgroup specific
and TLX1 regulated lncRNAs
TLX
TLX1 diff
144
14
68

Supplementary Figure legends

Supplementary Figure 1: Qualitative and quantitative analysis of polyA+ and total RNA-seq data of ALL-SIL lymphoblasts with transient *TLX1* knockdown. (A) Number of counts for each sample of polyA+ (left) and total (right) RNA-seq libraries. **(B)** Venn diagram depicting all genes (upper panel, Fisher exact test, p-value < 2.200e-16) and lncRNAs (lower panel, Fisher exact test, p-value < 2.200e-16) detected by polyA+ and/or total RNA-seq upon *TLX1* knockdown in ALL-SIL cells.

Supplementary Figure 2: RT-qPCR validation of the top 10 significantly differentially expressed genes. (A) Nine of the ten significantly differentially expressed lncRNAs detected by polyA+ could be validated by RT-qPCR. No primers could be designed for lncRNA *RP3-399L15.3*. Barplots show a negative control siRNA and two independent TLX1 targeting siRNAs. **(B)** Nine of the ten significantly differentially expressed lncRNAs detected by total RNA-seq could be validated by RT-qPCR. Only lncRNAs that are not in the top ten differentially expressed lncRNAs detected by polyA+ RNA-seq are shown. Barplots show a negative control siRNA and two independent TLX1 targeting siRNAs.

Supplementary Figure 3: TLX1 predominantly acts as an activator of lncRNA expression. (A) Barplot showing the total number of lncRNAs (left panel) and protein coding genes (right panel) differentially expressed upon *TLX1* knockdown without considering direct TLX1 binding as defined by ChIP-seq analysis. **(B)** Adapted barplot from (A) by only considering direct TLX1 binding events as defined by ChIP-seq analysis.

Supplementary Figure 4: TLX1 regulates lncRNAs in the vicinity (max. 1 Mb) of its well established T-ALL tumor suppressor genes. (A) Example of polyA+ RNA-seq, total RNA-seq and TLX1 ChIP-seq signals at the *RP11-539L10.2* locus. Boxplots show the effect of *TLX1* knockdown on the expression of the lncRNA. **(B)** TLX1 regulated lncRNAs *RP11-973H7.1* and *RP11-973H7.4* are located in the vicinity (max. 1 Mb) of *PTPN2*. **(C)** TLX1 regulated lncRNA *LINC00649* is located in the vicinity (max. 1 Mb) of *RUNX1*. PolyA+ and total RNA-seq tracks are depicted for control siRNA transfected samples. Bars represent the MACS2 peaks with FDR < 0.05.

Supplementary Figure 5: Functional annotation of the top-5 annotated lncRNA candidates downregulated upon *TLX1* knockdown in ALL-SIL cells through guilt-by-association analysis. (A) *RP3-399L15.3*, **(B)** *RP11-539L10.2*, **(C)** *RP11-284F21.10*, **(D)** *LINC01132*. For *RP11-18H21.1* no functional network could be built based on the primary patient cohort data since this lncRNA was not detected in this dataset.

Supplementary Figure 6: Functional annotation of the top-5 annotated lncRNA candidates upregulated upon *TLX1* knockdown in ALL-SIL cells through guilt-by-association analysis. (A) *AC011893.3*, **(B)** *RP11-284N8.3*, **(C)** *RP11-43F13.3*, **(D)** *HHIP-AS1*, **(E)** *RP11-973H7.4*.

Supplementary Figure 7: Open chromatin profiling by means of ATAC-seq. Zipper plots showing distances between TSSs and closest ATAC-seq peaks of 1000 TSSs of (from left to right) low, mid and high level expressed genes. Only ATAC-seq peaks within +/- 5 kb from the TSS are displayed.

Supplementary Figure 8: JQ1 treatment affects TLX1 regulated lncRNA expression. (A) Volcano plot showing significantly differentially expressed lncRNAs upon JQ1 inhibition in ALL-SIL cells. Red

(upregulated upon JQ1 treatment) and blue (downregulated upon JQ1 treatment) dots represent significantly differentially expressed lncRNAs (adjusted P-value <0.05) detected with total RNA-seq. LncRNA names depicted in the plots are the top-10 differentially regulated lncRNAs. **(B)** Venn diagram depicting overlapping lncRNAs significantly downregulated upon JQ1 treatment of ALL-SIL cells with lncRNAs significantly upregulated (upper diagram, Fisher exact test, p-value < 2.200e-16) or downregulated (lower diagram, Fisher exact test, p-value < 2.200e-16) upon *TLX1* knockdown.

Supplementary Figure 9: Qualitative and quantitative analysis of polyA+ and total RNA-seq data of a primary T-ALL cohort. (A) Number of counts for each sample of polyA+ (left) and total (right) RNA-seq libraries. **(B)** Venn diagram depicting all genes (left panel, Fisher exact test, p-value < 2.200e-16) and lncRNAs (right panel, Fisher exact test, p-value < 2.200e-16) detected by polyA+ and/or total RNA-seq in a primary T-ALL cohort.

Supplementary Figure 10: Identification of a set of previously unannotated lncRNAs specific for TLX subtype T-ALL. **(A)** Volcano plot representation of unannotated differentially expressed lncRNAs in TLX subtype T-ALL patients versus those of other T-ALL subgroups. Red (higher in TLX subtype T-ALL versus other) and blue (lower in TLX subtype T-ALL versus other) dots represent significantly differentially expressed lncRNAs detected with polyA+ RNA-seq (left panel) and total RNA-seq (right panel) (adjusted P-value <0.05). Gene names depicted in the plots are the top 10 unannotated differentially regulated lncRNAs. **(B)** Venn diagram depicting the overlap between significant differentially expressed lncRNAs upon *TLX1* knockdown and TLX subgroup specific lncRNAs (Fisher exact test, p-value = 0.002977).

Table legends

Supplementary Table 1: TLX1 significantly differentially expressed lncRNAs. Table shows the lncRNAs that are significantly differentially expressed upon *TLX1* knockdown. In addition, following features were also evaluated: TLX1 binding as evidenced by ChIP-seq, super-enhancer association, TLX subgroup specific expression in primary T-ALLs, differential expression upon JQ1 treatment and if a T-ALL tumor suppressor gene is located in within 1 Mb of the lncRNA. ENS_ID: ensembl ID; chr: chromosome; Padj: p adjusted; TSG: tumor suppressor gene.

Supplementary Table 2: **Spearman correlations of TLX1 significantly differentially expressed lncRNAs with protein-coding genes in the neighborhood of the lncRNA (<100 kb).** Table shows protein-coding genes that are located within 100 kb of a lncRNA and the Spearman correlations between the lncRNA and the protein-coding gene. Lnc_name: long noncoding RNA name; lnc_ENS: ensembl ID of the lncRNA; PC-ENS: ensembl ID of the protein-coding gene.

Supplementary Table 3: effect of TLX1knockdown on T-ALL tumor suppressor gene expression. Table shows if the T-ALL tumor suppressor genes are differentially regulated by TLX1 and if this tumor suppressor gene is bound by TLX1. ENS_ID: ensembl ID; chr: chromosome; Padj: p adjusted.

Supplementary Table 4: TLX1 significantly differentially expressed novel lncRNAs. Table shows novel lncRNAs that are significantly differentially expressed upon *TLX1* knockdown. In addition, following features were also evaluated: protein-coding potential scores, TLX1 binding as evidenced by ChIP-seq, super-enhancer association and if a tumor suppressor gene located in within 1 Mb of the lncRNA. ENS_ID: ensembl ID; chr: chromosome; Padj: p adjusted; TSG: tumor suppressor gene.

Supplementary Table 5: TLX subgroup specific significantly differentially expressed lncRNAs compared to other T-ALL subtypes. Table shows the lncRNAs that are significantly differentially expressed between the TLX subgroup and the other subgroups (immature, TALR, HOXA). In addition, following features were also evaluated: TLX1 binding as evidenced by ChIP-seq, super-enhancer association, differentially expressed upon TLX1 knockdown, differential expression upon JQ1 treatment and if a T-ALL tumor suppressor gene is located in within 1 Mb of the lncRNA. ENS_ID: ensembl ID; chr: chromosome; Padj: p adjusted; TSG: tumor suppressor gene.

Supplementary Table 6: subgroup specific significantly differentially expressed novel lncRNAs. Table shows novel lncRNAs that are significantly differentially expressed between the TLX subgroup and the other subgroups (immature, TALR, HOXA). In addition, following features were also evaluated: protein-coding potential scores, TLX1 binding as evidenced by ChIP-seq, super-enhancer association and if a tumor suppressor gene located in within 1 Mb of the lncRNA. ENS_ID: ensembl ID; chr: chromosome; Padj: p adjusted; TSG: tumor suppressor gene.