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

T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive hematological malignancy arising from uncontrolled proliferation and arrested differentiation of precursor T-cells. T-ALL is a genetically heterogeneous disease and can be subdivided into different molecular cytogenetic subgroups associated with specific gene expression signatures.[1,2] The T-cell leukemia homeobox 1 (TLX1, HOX11) transcription factor is a key driver of the TLX subgroup in T-ALL with ectopic expression in developing thymocytes causing a maturation arrest at the early cortical stage of T-cell development. Aberrant *TLX1* expression occurs in 5-10 % of pediatric and 30 % of adult T-ALL patients and predominantly results from t(7;10)(q34;q24) or t(10;14)(q24;q11) chromosomal translocations leading to juxtaposition of TLX1 to the T-cell receptor (TCR) δ or β promoter.[3] The TLX1 gene regulatory network has been extensively studied in terms of co-factors and downstream protein-coding gene targets.[4] Given that the protein-coding part of the genome only constitutes about 2 % while up to 70 % of the genome is transcribed (as non-coding ribonucleic acid (RNA)), a deeper exploration of the TLX1 driven non-coding transcriptome in T-ALL is warranted to support a more profound understanding of the molecular basis of this T-ALL subtype.[5] Long non-coding RNAs (lncRNAs) recently emerged as crucial transcriptional regulators in normal development and cancer, including normal and malignant hematopoiesis.[6,7] LncRNAs are arbitrarily defined as transcripts longer than 200 nucleotides and are poorly evolutionary conserved in terms of sequence.[8] Recently, our lab has identified a subset of lncRNAs that act in concert with NOTCH1 in both normal T-cell development and malignant T-cell transformation and a set of T-ALL subgroup-specific lncRNAs using microarray data.[9,10] In this study, we performed *in vitro TLX1* knockdown in T-ALL cells as well as a deep exploration of the TLX subgroup-specific lncRNAome in primary T-ALLs. For the former, we applied an integrative genomics approach combining quantitative data on the transcriptome and immunoprecipitated and open chromatin, using RNA-sequencing (RNA-seq), chromatin immunoprecipitation sequencing (ChIP-seq) and assay for transposase-accessible chromatin sequencing (ATAC-seq), respectively. Using this approach, we identified known and novel lncRNAs and gained insight into the super-enhancer marked lncRNA genetic landscape in TLX driven T-ALL, amongst others.

To elucidate the lncRNA repertoire under control of the TLX1 transcription factor, we performed transient *TLX1* knockdown by electroporating two TLX1 targeting small interfering RNAs (siRNAs) in ALL-SIL lymphoblasts, displaying ectopic *TLX1* expression as a result of a t(10;14)(q24;q11) translocation. From the resulting transcriptomes, both polyA+ and total RNA-seq libraries were generated in order to evaluate the expression changes of polyadenylated as well as non-polyadenylated lncRNA transcripts (*Online Supplementary Figure S1A,B*). By combining lncRNAs (biotype 'lincRNA' or 'antisense') detected with polyA+ (Figure 1A, left) and total RNA-seq (Figure 1A, right), more lncRNAs were significantly (adjusted *P*-value <0.05) downregulated (146 lncRNAs) than upregulated (80 lncRNAs) upon *TLX1* knockdown (*Online Supplementary Table S1*). Up- or downregulation of nine of the top ten differentially TLX1-regulated lncRNAs detected by polyA+ and total RNA-seq could be validated by quantitative reverse transcription polymerase chain reaction (RT-qPCR) (*Online Supplementary Figure S2*). Of note, this significantly different ratio between up- and downregulated lncRNAs is contrary to the effect of *TLX1* knockdown on protein-coding genes (mainly upregulated upon *TLX1* knockdown) (*Online Supplementary Figure S3A*), in concordance with its previously described role as a transcriptional repressor.[4] Moreover, this opposite ratio remains intact upon integration of TLX1 ChIP-seq data (*Online Supplementary Figure S3B*). Using *de novo* motif analysis on transcriptionally active (H3K27Ac+) and inactive (H3K27Ac-) TLX1 bound regions, a significant enrichment for the RUNX, PBX and MEIS family of transcription factor motifs was observed for H3K27ac+ and H3K27ac- regions, as previously observed for TLX1-regulated protein-coding genes.[4] In contrast, some transcription factors such as SP1 and TGIF1 were only enriched in H3K27ac+ or H3K27ac- regions, suggesting that TLX1 activated genes can be transcriptionally regulated by different transcription factor families compared to TLX1 repressed genes (Figure 1B).

Among the 226 lncRNAs regulated by TLX1, 64 lncRNAs display a TLX1 chromatin binding peak in their immediate vicinity (max. 5 kb) (*Online Supplementary Table S1*), as illustrated for lncRNA RP11-539L10.2 (*Online Supplementary Figure S4A*). For 80 of the 226 differentially regulated lncRNAs upon *TLX1* knockdown, the expression was significantly correlated with at least one neighboring protein-coding gene (|rho|(Rs) >0.5, P-value <0.05) located within a 100 kb window, irrespective of strand orientation (*Online Supplementary Table S2*). From the latter, 97.25 % are positively correlated with the expression of the differentially regulated lncRNAs, consistent with previous reports.[11]

Interestingly, three of the identified TLX1-regulated lncRNAs are in the vicinity (max 1 Mb) of a known differentially regulated T-ALL tumor suppressor gene[4] (*Online Supplementary Figure S4B,C; Online Supplementary Table S1* and *S3*). To assign a possible function to the top five TLX1 up- and downregulated lncRNAs, a guilt-by-association approach was followed as described in the *Online Supplementary Methods* section of this paper (*Online Supplementary Figure S5* and *S6*).

As it is known that some lncRNAs are located within super-enhancer regions, a hockey stick plot based on H3K27ac ChIP data for lncRNA loci was generated as described in *Online Supplementary Methods* (Figure 1C). Among the 2781 super-enhancer associated lncRNAs, 115 lncRNAs were significantly differentially expressed upon *TLX1* knockdown with a significant enrichment (42 lncRNAs) of TLX1 binding for these TLX1-regulated, super-enhancer associated lncRNAs, as exemplified for lncRNA *NBAT1* (Figure 1D; *Online Supplementary Table S1*). As super-enhancers are associated with regions of open chromatin, we also performed ATAC-seq and confirmed that 98.95 % of the super-enhancer regions overlap with regions of open chromatin. Moreover, we discovered that 66.4 % of the transcription start sites (TSSs) from highly expressed (top decile) genes had ATAC-seq peaks within +/- 5 kb (*Online Supplementary Figure S7*). To further explore the functional association of super-enhancers and expressed lncRNAs, the genome-wide transcriptional response of lncRNAs upon JQ1 treatment of ALL-SIL lymphoblasts was investigated, given that this bromodomain and extra-terminal motif (BET) inhibitor causes a decrease in the expression of super-enhancer associated genes (*Online Supplementary Figure S8A*).[12] Among 115 super-enhancer associated, TLX1-regulated

lncRNAs, 41 lncRNAs were differentially expressed upon JQ1 inhibition (*Online Supplementary Table S1*). Moreover, 26 upregulated and 24 downregulated lncRNAs upon *TLX1* knockdown were significantly overlapping with those lncRNAs downregulated upon JQ1 exposure (*Online Supplementary Figure S8B*).

In addition to previously annotated genes, 2788 lncRNAs that have not been previously annotated in Ensembl, Gencode, LNCipedia and RefSeq were also detected. Of these novel lncRNAs, 82 are differentially regulated upon *TLX1* knockdown, of which 30 are directly bound by TLX1, as illustrated for *MSTRG.6968* (Figure

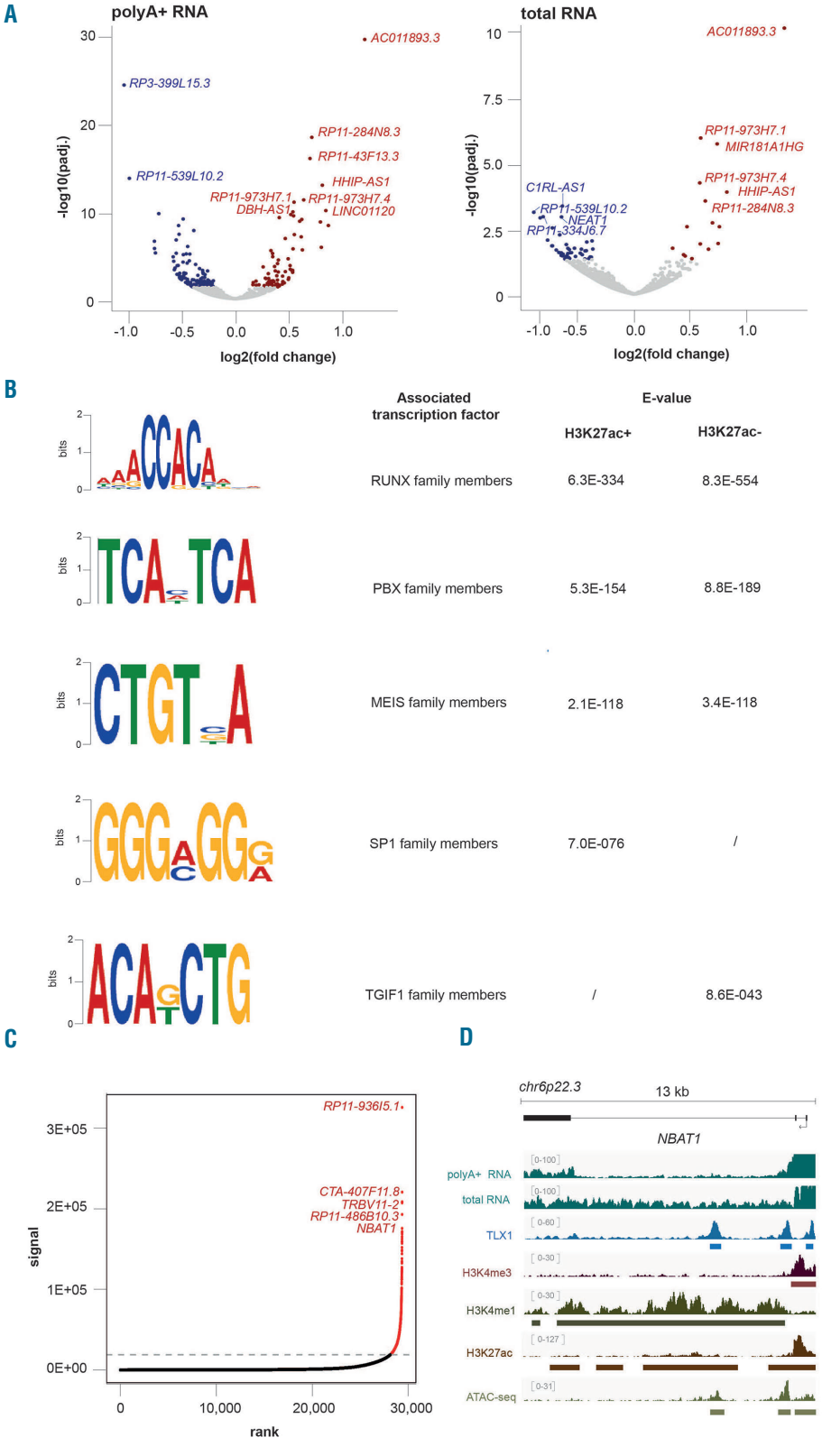

Figure 1. Integrative TLX1 ChIP-seq and transcriptome analysis upon *TLX1* knockdown in ALL-SIL lymphoblasts for identification of a robust set of TLX1 directly regulated lncRNAs and super-enhancer associated lncRNAs. (A) Volcano plot representation of differentially expressed lncRNAs upon *TLX1* knockdown in ALL-SIL. Red (upregulated upon *TLX1* knockdown) and blue (downregulated upon *TLX1* knockdown) dots represent significantly differentially expressed lncRNAs detected with polyA+ RNA-seq (left panel) and total RNA-seq (right panel) (adjusted *P*-value <0.05). LncRNA names depicted in the plots are the top ten differentially regulated lncRNAs. Outliers with a -log10(padj) >30 are scaled to log10(padj)=30. (B) Motif enrichment analysis on the set of TLX1 bound regions with and without overlap of H3K27ac ChIP-seq peaks using MEME-ChIP suite identifies significant enrichment of the DNA binding motifs of the RUNX, PBX and MEIS family of transcription factors for both sets of peaks while the SP1 and TGIF1 families are only enriched in one set of peaks. (C) Hockey stick plot representing the normalized rank and cluster signal of clusters of H3K27ac ChIP-seq peaks at lncRNA transcripts. Red dots represent lncRNAs significantly associated with a super-enhancer (adjusted *P*-value <0.05). (D) IGV screenshot of a super-enhancer associated lncRNA (*NBAT1*). PolyA+ and total RNA-seq tracks are depicted for control siRNA transfected samples. Bars represent the MACS2 peaks with FDR <0.05. RNA: ribonucleic acid. lncRNA: long non-coding RNA.

2A,B; *Online Supplementary Table S4*). Of note, *MSTRG.37538* is a lncRNA marked with one of the strongest genome-wide super-enhancer sites of all identified unannotated lncRNAs (Figure 2C,D).

In a complementary approach, TLX1 and TLX3 (further denoted as TLX) driven lncRNAs were retrieved from a primary T-ALL patient cohort as TLX1 and TLX3 induce T-ALL in a similar way and are associated with a similar gene expression profile.[13] By using polyA+ RNA-seq data of 60 T-ALL patients (including 17 TLX positive cases) as well as total RNA-seq of 25 T-ALL patients (including 10 TLX positive cases) 442 known and 158 novel TLX subgroup-specific lncRNAs were identified (Figure 3A; *Online Supplementary Figure S9* and *S10A; Online Supplementary Table S5* and *S6*). From these, 32 known and 14 novel lncRNAs overlapped significantly with the known and novel set of differentially expressed genes upon *TLX1* knockdown, respectively (Figure 3B; *Online Supplementary Figure S10B*). Moreover, 22 known and three novel TLX subgroup-specific lncRNAs are in the vicinity (max 1 Mb) of a known differentially regulated T-ALL tumor suppressor gene (*Online Supplementary Table S5* and *S6*).[4]

To identify possibly oncogenic TLX subtype-specific

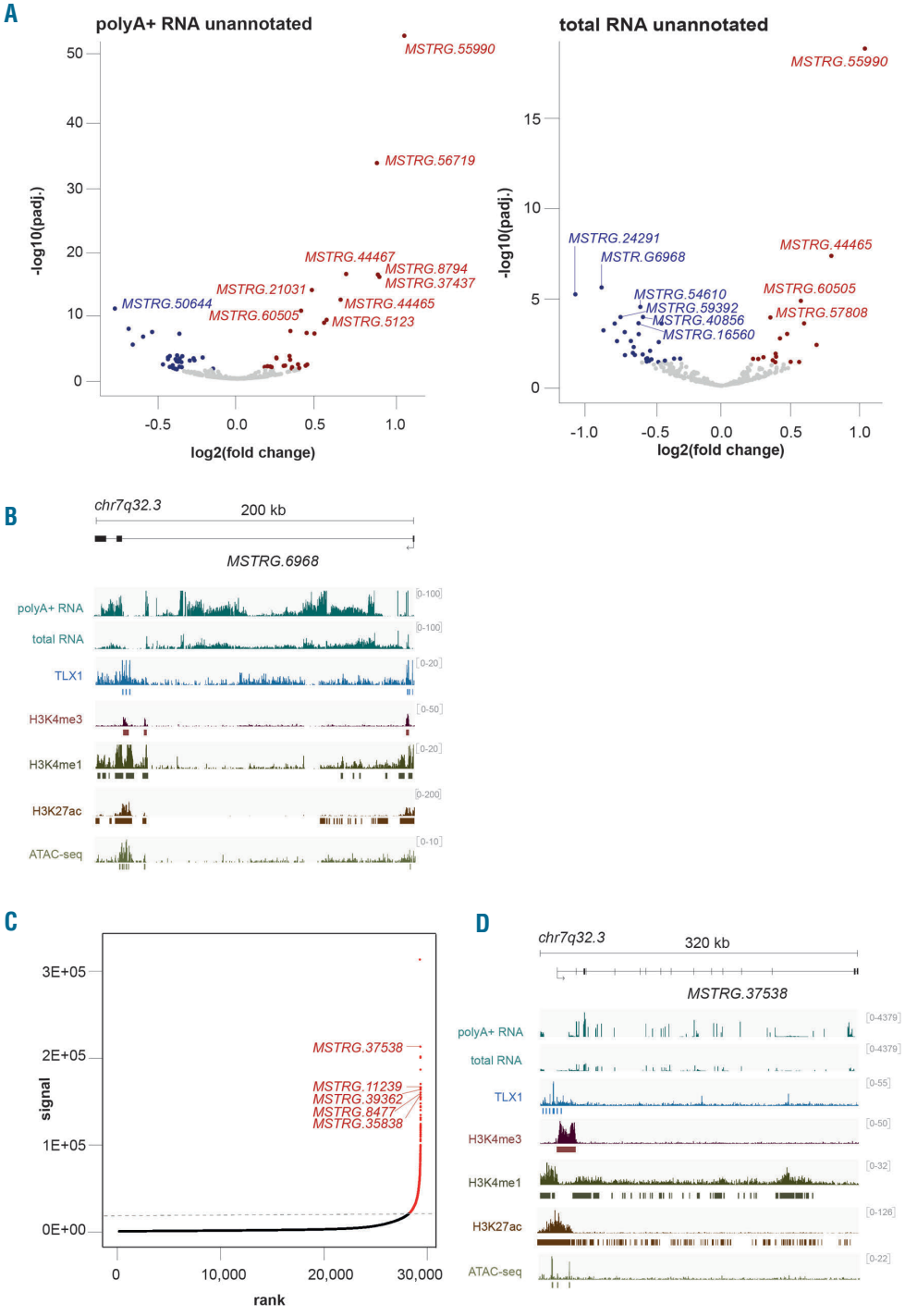

Figure 2. Identification of a set of previously unannotated TLX1 regulated lncRNAs in ALL-SIL lymphoblasts. (A) Volcano plot representation of unannotated differentially expressed lncRNAs upon *TLX1* knockdown in ALL-SIL. Red (upregulated upon TLX1 knockdown) and blue (downregulated upon *TLX1* knockdown) 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 ten unannotated differentially regulated lncRNAs. (B) IGV screenshot of an unannotated differentially expressed, TLX1 bound lncRNA (*MSTRG.6968*). PolyA+ and total RNA-seq tracks are depicted for control siRNA transfected samples. Bars represent the MACS2 peaks with FDR <0.05. (C) Hockey stick plot representing the normalized rank and cluster signal of clusters of H3K27ac ChIP-seq peaks. Red dots represent unannotated lncRNAs significantly associated with a super-enhancer. (D) IGV screenshot of an unannotated super-enhancer associated lncRNA (*MSTRG.37538*). PolyA+ and total RNA-seq tracks are depicted for control siRNA transfected samples. Bars represent the MACS2 peaks with FDR <0.05. RNA: ribonucleic acid. lncRNA: long non-coding RNA.

lncRNAs, this new data was integrated with our previously generated polyA+ RNA-seq data of OP9-DL1 cultured T-cells,[10] serving as reference material for lncRNA expression levels in untransformed T-cell progenitors. Therefore, lncRNAs that are significantly higher expressed in the TLX subgroup as compared to the other T-ALL subgroups of the primary T-ALL cohort (HOXA, immature, TAL) and significantly higher as compared to normal T-cells were selected. Those lncRNAs that were also differentially expressed among any of the other T-ALL subgroups and T-cells were excluded. In total, 144 TLX-specific, potentially oncogenic lncRNAs were identified (Figure 3C), as illustrated for lncRNA *RP11-973H7.4* (Figure 3D, left), located in the immediate vicinity of the well-known T-ALL tumor suppressor gene *PTPN2* (Figure 3D, right).

In this study, we present the first comprehensive analysis of the lncRNA transcriptome of TLX1+ ALL-SIL lymphoblasts and TLX subtype primary T-ALLs, uniquely integrating the polyadenylated and non-polyadenylated

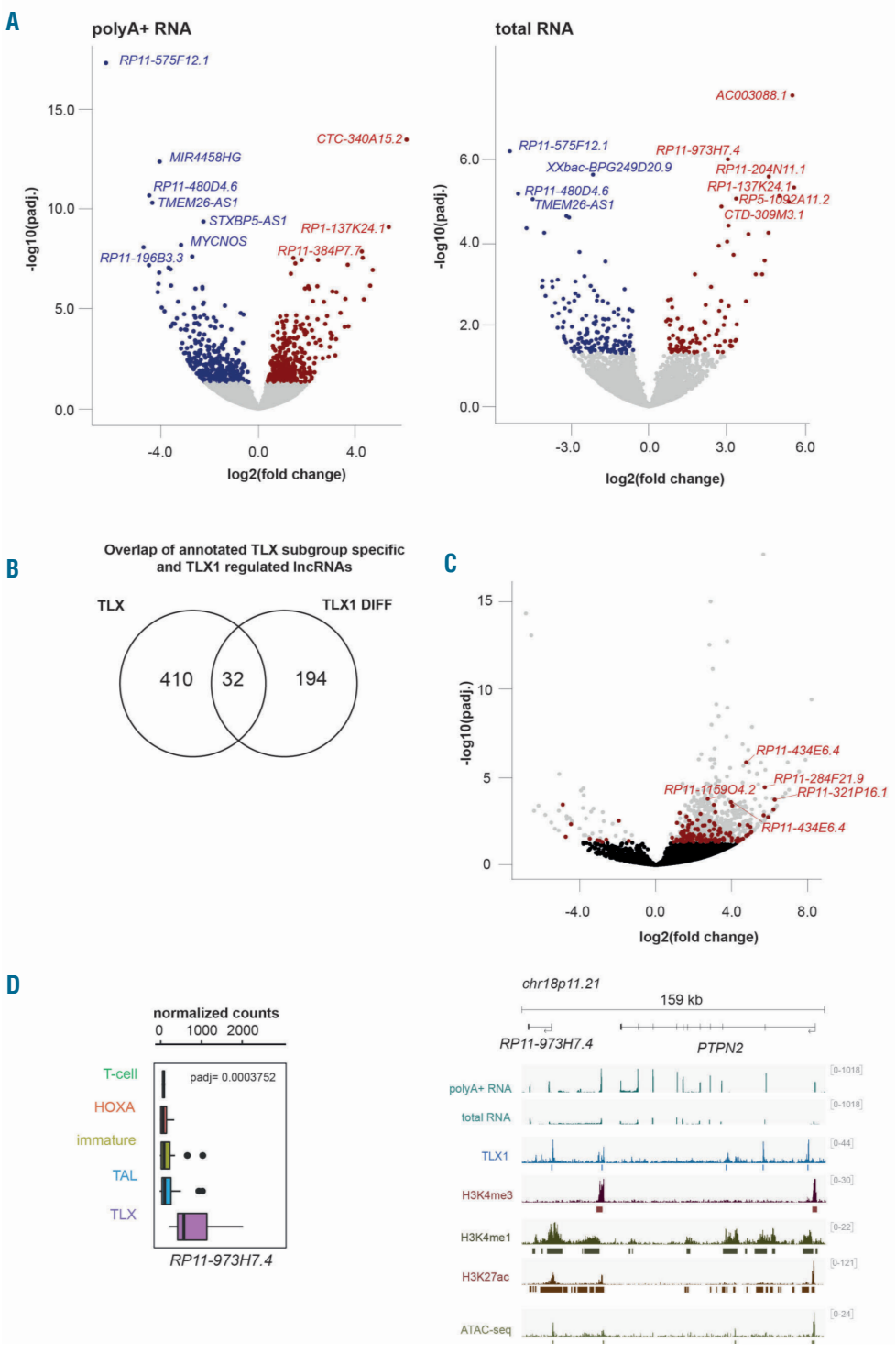

Figure 3. Identification of TLX specific, possibly oncogenic lncRNAs in a primary T-ALL cohort. (A) Volcano plot representation of lncRNAs that are significantly higher or lower in the TLX group as compared with T-ALL patients belonging to other T-ALL subtypes (*TALR*, immature, HOXA). Red (upregulated in TLX subtype T-ALLs *versus* other subtypes) and blue (downregulated in TLX subtype T-ALLs *versus* other subtypes) dots represent significantly differentially expressed lncRNAs detected with polyA+ RNA-seq (left) and total RNA-seq (right) (adjusted *P*-value <0.05). LncRNA names depicted in the plots are the top ten differentially regulated lncRNAs. (B) Venn diagram depicting the overlap between significant differentially expressed lncRNAs upon *TLX1* knockdown and TLX subgroup specific lncRNAs (Fisher's exact test, adjusted *P*-value =1.601e-12). (C) Volcano plot representation of lncRNAs that are significantly higher expressed in the TLX group as compared with normal T-cell subsets. Gray and red dots represent significant differentially expressed lncRNAs (adjusted *P*-value <0.05). Red dots are TLX specific lncRNAs not differentially expressed between T-cells and other subgroups. LncRNA names depicted in the plots are the top five differentially regulated lncRNAs. (D) Boxplot and IGV screenshot for lncRNA *RP11-973H7.4*, that is significantly higher expressed in the TLX subgroup compared to the other subgroups and significantly higher expressed as in normal thymocytes. PolyA+ and total RNA-seq tracks are depicted for control siRNA transfected samples. Bars represent the MACS2 peaks with FDR <0.05. RNA: ribonucleic acid; lncRNA: long non-coding RNA.

transcriptome and chromatin features. Our results reveal that TLX1 directly regulates a set of known and novel lncRNAs of which some are marked by super-enhancers. By integrating normal T-cell data and a primary T-ALL patient cohort, we also identified 144 putative TLX-specific oncogenic lncRNAs, which could be further tested for phenotypic effects upon knockdown and explored as new targets for RNA-based therapeutics. LncRNAs may serve as excellent therapeutic targets as these are often expressed in a cell-type-specific manner, offering potential advantages with respect to on-target toxicity as shown by our research group for lncRNA *SAMMSON* in melanoma.[14] In conclusion, our study delineates a TLX subgroup and TLX1-specific lncRNA network including a subset of super-enhancer associated lncRNAs. Our work, together with that of others, strongly suggest an important role of lncRNAs in T-ALL and warrant further functional investigation.[9,10,15]

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

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

References

1. Soulier J, Clappier E, Cayuela JM, et al. HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). Blood. 2005;106(1):274–286.
2. Spinella J-F, Cassart P, Richer C, et al. Genomic characterization of pediatric T-cell acute lymphoblastic leukemia reveals novel recurrent driver mutations. Oncotarget. 2016;7(40):65485–65503.
3. Riz I, Hawley RG. G1/S transcriptional networks modulated by the HOX11/TLX1 oncogene of T-cell acute lymphoblastic leukemia. Oncogene. 2005;24(36):5561–5575.
4. 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.
5. Djebali S, Davis CA, Merkel A, Gingeras TR. Landscape of transcription in human cells. Nature. 2012;489(7414):101–108.
6. Bhan A, Soleimani M, Mandal SS. Long noncoding RNA and cancer: a new paradigm. Cancer Res. 2017;77(15):3965-3981.
7. Nobili L, Lionetti M, Neri A. Long non-coding RNAs in normal and malignant hematopoiesis. Oncotarget. 2016;7(31):50666-50681.
8. Guttman M, Amit I, Garber M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009;458(7235):223–227.
9. 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.
10. 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.
11. Casero D, Sandoval S, Seet CS, et al. LncRNA profiling of human lymphoid progenitors reveals transcriptional divergence of B and T lineages. Nat Immunol. 2016;16(12):1282–1291.
12. Lovén J, Hoke HA, Lin CY, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell. 2013;153(2):320–334.
13. Della Gatta G, Palomero T, Perez-Garcia A, et al. Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in T-ALL. Nat Med. 2012;18(3):436–440.
14. Leucci E, Vendramin R, Spinazzi M, et al. Melanoma addiction to the long non-coding RNA SAMMSON. Nature. 2016;531(7595):518–522.
15. Cao Thi Ngoc P, Hao Tan S, King Tan T, et al. Identification of novel lncRNAs regulated by the TAL1 complex in T- cell acute lymphoblastic leukemia. Leukemia 2018 Mar 26. [Epub ahead of print].