Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Nat Immunol. Author manuscript; available in PMC 2014 May 01.

Published in final edited form as:

Nat Immunol. 2013 November ; 14(11): 1190–1198. doi:10.1038/ni.2712.

Expression and regulation of lincRNAs during T cell development and differentiation

Gangqing Hu[1,3], **Qingsong Tang**[1,3], **Suveena Sharma**[2,3], **Fang Yu**[2], **Thelma M. Escobar**[2], **Stefan A. Muljo**[2], **Jinfang Zhu**[2,#], and **Keji Zhao**[1,#]

[1]Systems Biology Center, National Heart, Lung and Blood Institute, National Institutes of Health (NIH), Bethesda, MD 20892

[2]Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD 20892

Abstract

Although lincRNAs are implicated in gene regulation in various tissues, little is known about lincRNA transcriptomes in the T cell lineages. Here we identify 1,524 lincRNA clusters in 42 T cell samples from early T cell progenitors to terminally differentiated T helper (T_H) subsets. Our analysis revealed highly dynamic and cell-specific expression patterns of lincRNAs during T cell differentiation. Importantly, these lincRNAs are located in genomic regions enriched for protein-coding genes with immune-regulatory functions. Many of them are bound and regulated by the key T cell transcription factors T-bet, GATA-3, STAT4 and STAT6. We demonstrate that the lincRNA LincR-Ccr2-5′AS, together with GATA-3, is an essential component of a regulatory circuit in T_H2-specific gene expression and important for T_H2 cell migration.

The mammalian genomes encode tens of thousands of long noncoding RNAs (lncRNA)[1,2]. These transcripts play essential roles in regulating gene expression and affect various biological processes during development and in pathological conditions[3,4]. One classic example of a functional lincRNA is *Xist*, which is located to the X chromosome and is required for X chromosome inactivation in females. *Xist* operates by recruiting repressive complexes such as PRC2 to the silenced X chromosome[5]. Another well-characterized example is HOTAIR that recruits the PRC2 complexes to Hox domains and represses the expression of *HOXD*[6]. Additionally, several other lincRNAs function to mediate H3K27

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

[#]Corresponding authors: Jinfang Zhu, Laboratory of Immunology, NIAID, Bethesda, MD 20892, Tel. (301) 402-6662, Fax. (301)-480-7352, JFZHU@niaid.nih.gov. Keji Zhao, Systems Biology Center, NHLBI, Bethesda, MD 20892, Tel. (301) 496-2098, Fax. (301) 480-0961, zhaok@nhlbi.nih.gov.

[3]These authors contributed equally to this work.

Accession codes

Raw sequence data together with processed files are accessible from GEO (GSE48138); the deposited GTF and BEDGraph files can be directly uploaded to UCSC genome browser for visualization (see Supplemental Fig. 7 for an example).

Author contributions

G.H, J.Z. and K.Z conceived the study, designed experiments and data analysis, and wrote the manuscript. Q.S., S.S. and F.Y. did experiments and edited the manuscript. G.H. analyzed the data. T.M.E. and S.A.M. contributed RNA-Seq data for STAT4-deficient T_H1 cells, STAT6-deficient T_H2 cells and the corresponding wild type T_H1 and T_H2 cells.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

methylation by recruiting PRC2[7,8]. LncRNAs with enhancer functions have also been reported[9]. Intergenic lncRNAs (lincRNAs) may also regulate gene expression through post-transcriptional mechanisms[10,11].

The study of lincRNA function in the immune system is an emerging field. T helper (T_H) cells are critical for orchestrating adaptive immune responses to a variety of pathogens; they are also involved in the pathogenesis of different types of immunological diseases including allergy, asthma and autoimmunity[12]. A lincRNA, *TMEVPG1*, was described to be specifically expressed in T_H1 cells and critical for controlling Theiler's viral infection[13]. Together with the T_H1-specific transcription factor T-bet, *TMEVPG1* controls the expression of interferon γ[14]. This RNA, also termed NeST, interacts with WDR5, a core subunit of the MLL H3K4 methyltransferases, and facilitates the histone methylation at the *Ifng* locus in $CD8^+$ T cells[15]. A survey of long noncoding RNA in $CD8^+$ T cell from mouse spleen by using a custom array suggests a pivotal role of lncRNAs in the differentiation and activation of lymphocytes[16].

Despite these examples, the function and transcriptional regulation of lincRNAs in T cell development and differentiation is far from understood, partially due to the lack of knowledge of lincRNA expression in cells of the immune system[17]. Thus, to better understand the role of lincRNAs in the development and differentiation of T cell lineages, we performed RNA-Seq of 42 subsets of thymocytes and mature peripheral T cells at multiple time points during their differentiation. Analysis of this dataset identified 1,524 genomic regions that generate lincRNAs. Our data reveal a highly dynamic and cell- or stage-specific pattern of lincRNA expression. Genomic location analysis of the lincRNA genes revealed that they are adjacent to protein-coding genes critically involved in regulating immune function, suggesting a possible co-evolution of protein-coding and lincRNA genes. Using gene deficient mice, we found that the transcription factors T-bet, GATA-3, STAT4 and STAT6 account for the cell-specific expression of most lincRNAs in T_H1 and T_H2 cells. Inhibition of a T_H2-specific lincRNA, LincR-*Ccr2*-5′AS, whose expression is regulated by GATA-3, by shRNA resulted in deregulation of numerous genes preferentially expressed in T_H2 cells including several chemokine receptor genes located in the vicinity of the LincR-*Ccr2*-5′AS and compromised T_H2 migration to lung tissues in mice. Therefore, our datasets provided a comprehensive resource for future studies of the function and mechanisms of lincRNA in T cell development, differentiation and immune response.

Results

Cataloging the lincRNA expression profiles in various T cell types

To obtain comprehensive profiles of lincRNA expression during the development and differentiation of T cell lineages, we purified CD4-CD8 double negative (DN), double positive (DP), CD4 or CD8 single positive (SP) of thymic T cells and thymus-derived regulatory T (tT_{reg}) cells from lymph nodes of C57BL/6 mice. Additionally, we obtained T_H1, T_H2, T_H17 and induced regulatory T (iT_{reg} cells) by *in vitro* differentiation of naïve $CD4^+$ T cells for a various length of time in culture. In total, we obtained 42 T cell subsets (Supplemental Fig. 1a). Total and/or polyadenylated RNA from these cells was analyzed

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

using RNA-Seq. Following a similar strategy as previously described[18] (Supplemental Fig. 1b), we identified a total of 1,524 lincRNA-expressing genomic regions (or clusters) in these 42 T cell subsets (Supplemental Table 1). Because each cluster may encode more than one lincRNA, the number of lincRNAs may be larger than 1,524. For example, the LincR-*Gata3*-3′ cluster downstream of the *Gata3* gene contained at least two divergently transcribed lincRNA genes (Fig. 1a). 73% of the clusters could not be identified using noncoding gene annotations from public databases such as RefSeq[19], Ensembl[20], UCSC[21] and NONCODE[22] and thus were novel potential lincRNA genes (Supplemental Table 1). The number of lincRNA clusters identified within each T cell subset ranged from 154 to 354 (Fig. 1b).

We employed several criteria to estimate the coding potential of these potential lincRNA gene clusters. Transcriptome assembly was used to retrieve transcripts (both spliced and non-spliced) from each cluster. Approximately 482 clusters (32% of all clusters) contain spliced transcripts. On average, each lincRNA gene cluster contains 1.5 independent transcripts, which can either be spliced or non-spliced but originate from the same promoter; there are a total of 2,194 such transcripts. Blasting the putative ORFs from these transcripts against Swiss-Prot protein database revealed that only 72 clusters contained transcripts with significant similarity to protein-coding genes (E-value < 10^{-5}, identity > 30%). The coding potential of these 2,194 transcripts was also evaluated by using the Coding Potential Calculator[23]. Transcripts with high coding potential were detected in 91 clusters. Lastly, compared to protein coding genes, transcripts from lincRNAs expressed in T_H2 cells (two weeks) were significantly enriched in the nucleus compared to the cytoplasm (Supplemental Fig. 1c), suggesting that most of the lincRNAs are not translated. These analyses indicate that the majority of the identified lincRNA loci (>90%) showed limited coding potential. Because over half of the clusters with potential coding transcripts also harbored non-coding transcripts (data not shown), we included them for further data analysis.

To provide a systematic identifier for each lincRNA locus, we proposed the following nomenclature: LincR-*X*-5′ or LincR-*X*-3′ for a lincRNA cluster situated nearby a protein-coding gene X. An additional label “S” (sense strand) or “AS” (anti-sense strand) may be added if the direction of the lincRNA cluster to gene *X* can be inferred from the transcript assembly (detailed in **Methods**).

In summary, by analyzing RNA-Seq and ChIP-Seq datasets of T cells at different developmental and differentiation stages, we identified 1,524 potential lincRNA clusters, of which the majority is novel.

LincRNAs exhibit stage and lineage specificity

To determine the cellular specificity of lincRNA expression, we performed a pair-wise comparison of both protein-coding mRNAs and lincRNAs between any two developmental stages of T cells to determine differential expression at one stage compared to another. The results revealed that the overall mRNA expression was highly similar, but lincRNA expression exhibited remarkable differences between any two T cell subsets from *in vivo* (Fig. 1c) or *in vitro* T cell differentiation (Fig. 1d). We further separated the T cell subsets into three groups (DN cells; DP and SP cells and tTreg; and naïve $CD4^+$ T cells and T_H

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

cells) and plotted Venn diagrams for lincRNA expression within each group (Fig. 1e–g). The data indicate that 48–57% of lincRNAs were stage or lineage-specific, in contrast to 6–8% of mRNAs. On the other hand, 75–80% of protein-coding genes were shared by subsets within a group, in contrast to 13–16% of the lincRNA genes. To examine the expression of each individual lincRNA, we plotted a heat map of lincRNA expression in T_H cells (Fig. 1h). About 56% of lincRNAs (367) exhibited a preferential expression in one T cell subset to the others. A comparison between the T cell lincRNA catalog generated here and the lincRNAs identified in mouse embryonic stem cells (mESCs)[24] revealed that only 8.5% of the T cell lincRNAs (or 129 clusters) were also expressed in mESCs. In summary, we found that lincRNA expression is highly cell-specific during T cell development and differentiation.

Most lincRNAs are polyadenylated and dynamically regulated

By comparing the total RNA-Seq and polyA RNA-Seq data, we observed that some lincRNAs were polyadenylated in T cells, as exemplified in Fig. 2a. To determine what lincRNAs are polyadenylated and if the polyadenylated lincRNAs are dynamically regulated during T_H cell differentiation, we profiled the polyA$^+$ RNAs from various time points (4 hrs, 8 hrs, 12 hrs, 24 hrs, 48 hrs, 72 hrs, 1 week and 2 weeks) of T cell differentiation from naïve CD4$^+$ T cells by using RNA-Seq. A comparison between the expression of polyA$^+$ RNAs, as calculated from the sequencing data, and total RNAs indicated a high correlation of *LincR-Gata3-3′* expression between the two sets ($r = 0.99$) across different subsets (Fig. 2b and Supplemental Table 2). Analysis of all lincRNA clusters revealed that over 50% of the lincRNAs expressed in T_H cells exhibited a Pearson correlation coefficient (r) higher than 0.6, suggesting that a large fraction of the lincRNAs are polyadenylated (Fig. 2c).

Analysis of polyadenylated lincRNAs at different time points during T_H cell differentiation mentioned above indicated that many lincRNAs expressed in naïve CD4$^+$ T cells were rapidly down regulated after 4 hours of differentiation and then were re-expressed after 48–72 hours (Fig. 2d, e), suggesting that they might be involved in regulating T cell activation. We next visualized the lincRNA expression levels at all the time points mentioned above during the differentiation from naïve CD4$^+$ T cell to different T_H subsets by using heatmap (Fig. 2f). Of all the 539 lincRNAs involved, 19 were rapidly down-regulated within 4 hours of T cell differentiation and remained largely silenced throughout the later time points (e.g., LincR-*Chd2*-5′-74K). One T_H2-preferred lincRNA, LincR-*Sla*-5′AS, was rapidly induced at 4 hrs and its expression was then gradually decreased during later differentiation. Most of T_H1- and T_H2-preferred lincRNAs exhibited substantial induction at 48–72 hours and reached a plateau of expression in 1–2 weeks (e.g., LincR-*Gata3*-3′ and LincR-*Ccr2*-5′AS), while most of T_H17-preferred lincRNAs were maximally induced in 48–72 hrs. Taken together, most lincRNAs are polyadenylated and are dynamically regulated during T cell differentiation.

STAT4 activates T_H1-preferred lincRNAs

To understand the cell-specific expression of lincRNAs, we investigated the role of key transcription factors in regulating lincRNA expression during T cell differentiation. The transcription factors STAT4 and STAT6 regulate key aspects of T_H1 and T_H2

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

differentiation, respectively, by binding to and activating lineage-specific enhancers in these cells[25]. Using STAT4 ChIP-Seq data[26], we found that STAT4 bound to 56% of the lincRNA genes (861), such as the LincR-*Gng2*-5′ locus (Fig. 3a). STAT4 binding was higher at the lincRNA clusters that were preferentially expressed in T_H1 cells than at other clusters (Fig. 3b). STAT4 deletion decreased the expression of the LincR-*Gng2*-5′AS cluster in T_H1 cells by comparing the wild type and STAT4-deficient cells (Fig. 3a). In total, 39% of the T_H1-preferred lincRNA clusters (90) were down-regulated in the STAT4-deficient T_H1 cells, in contrast to 8% of non-T_H1-preferred lincRNAs (83) (Fig. 3c). Because STAT4 may also repress lincRNAs not preferentially expressed in T_H1 cells, we observed that more non-T_H1-preferred lincRNAs than T_H1-preferred lincRNAs were up-regulated in the STAT4-deficient T_H1 cells (Fig. 3c). Furthermore, lincRNAs with high levels of STAT4 binding at promoters, as measured in the wild type T_H1 cell, were more likely to be down-regulated in the STAT4-deficient T_H1 cells than those with low level of STAT4 binding (Fig. 3d). Consistently, lincRNAs with more STAT4 binding were less likely to be up-regulated in the STAT4 deficient T_H1 cells (Fig. 3d). In summary, our data indicate that STAT4 binds to and activates T_H1-specific lincRNAs in T_H1 cells.

STAT6 activates T_H2-preferred lincRNAs

To understand the function of STAT6 in regulating lincRNA expression in T_H2 cells, we analyzed its binding to lincRNA clusters in the wild type T_H2 cells using ChIP-Seq data[26]. STAT6 binding was detected in 56% of all lincRNA clusters (856), among them LincR-*Epas1*-3′AS (Fig. 3e). STAT6 binding to promoters in T_H2 cells was higher in the T_H2-preferred lincRNA clusters than other clusters (Fig. 3f). Expression of LincR-*Epas1*-3′AS was decreased in T_H2 cells from STAT6-deficient mice compared to the wild type mice (Fig. 3e). In total, expression of 32% of the T_H2-preferred lincRNA genes (56) was decreased and 12% of the other lincRNAs (131) was increased in STAT6-deficient T_H2 cells compared to the wild type T_H2 cells (Fig. 3g). If both a lincRNA and a protein-coding gene were activated or repressed by STAT6, they tended to be co-regulated during T cell differentiation; similar observation was made for STAT4-mediated gene regulation in T_H1 cells (Supplemental Fig. 2). LincRNAs with high STAT6 binding at promoters, as measured in the wild type T_H2 cells, were more likely to be down-regulated in STAT6-deficient T_H2 cells than those lincRNAs with low STAT6 binding (Fig. 3h). However, up-regulation of lincRNAs in STAT6-deficient T_H2 cells was not significantly correlated with STAT6 binding in the wild type T_H2 cells (Fig. 3h). In summary, STAT6 binds to and mediates the activation of T_H2-specific lincRNAs in T_H2 cells.

T-bet regulates expression of lincRNAs in T_H1 cells

Besides STATs, a number of other key transcription factors are implicated in T cell differentiation, such as T-bet for T_H1 differentiation27 and GATA-3 for T_H2 differentiation[28]. To investigate whether T-bet contributes to the cell-specific expression of lincRNAs in T_H1 cells, we analyzed T-bet binding at lincRNA genes using published T-bet ChIP-Seq dataset[29]. T-bet binding was detected at 14% of all lincRNA clusters (209) in T_H1 cells, among them LincR-*Ifng*-3′AS (previously also known as TMEVPG1 or NeST) (Fig. 4a), suggesting that T-bet may contribute to its expression. In T_H1 cells, T-bet binding was also detected at lincRNAs specifically expressed in other T_H cells, such as at the promoter of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the T_H2-specific LincR-*Ccr2*-5′AS gene (Fig. 4b), suggesting that T-bet may function to repress its expression in T_H1 cells. At a genome wide, T-bet preferentially bound to T_H1-preferred lincRNA genes (Fig. 4c), suggesting that T-bet mainly positively regulates the expression of T_H1-specific lincRNAs.

To test this hypothesis, we compared the total RNA expression profiles of wild type and *Tbx21*$^{-/-}$ T_H1 cells using RNA-Seq. Expression of LincR-*Ifng*-3′AS was decreased in *Tbx21*$^{-/-}$ T_H1 cells (Fig. 4d), consistent with a previous report14. In contrast, the expression of the T_H2-preferred LincR-Ccr2-5′AS was modestly up-regulated in *Tbx21*$^{-/-}$ T_H1 cells compared to the wild type T_H1 cells (Fig. 4d). Global analysis revealed that the *Tbx21* deletion caused decreased expression of 54 lincRNAs and increased expression of 37 lincRNAs in T_H1 cells (Fig. 4e). Thirty-six percentages (33/91) of the affected genes were T_H1-preferred. On the other hand, 16.3% of those unaffected lincRNAs (216) were T_H1-preferred ($p<0.0001$). Thus, our data indicated that T-bet binds to and contributes to both the activation and repression of lincRNAs in T_H1 cells.

GATA-3 regulates expression of lincRNAs in T_H2 cells

GATA-3, a zinc-finger transcription factor, is highly expressed in T_H2 cells and is critical to T_H2 differentiation by regulating T_H2 gene expression30. The analysis of published GATA-3 ChIP-Seq data[31] showed that 28.5% of the T_H2-preferred lincRNA clusters (53) were bound by GATA-3. GATA-3 preferentially bound to the promoters of T_H2-preferred lincRNAs in wild type T_H2 cells (Fig. 5a). Consistently, GATA-3-bound lincRNA clusters were more highly expressed than non-bound clusters (Fig. 5b). To test if GATA-3 contributes to the regulation of lincRNAs, we compared their expressions in wild type and *Gata3*$^{-/-}$ T_H2 cells. GATA-3 deficiency markedly decreased the expression of LincR-*Ccr2*-5′AS (Fig. 5c). Global analysis of lincRNA expression revealed that GATA-3 deficiency resulted in decreased expression of 30% (102) of lincRNA clusters bound by GATA-3, compared with 15 % (149) of lincRNA clusters not bound by GATA-3 (Fig. 5d). Interestingly, 11% of lincRNA clusters bound by GATA-3 exhibited increased expression in *Gata3*$^{-/-}$ T_H2 cells, compared to 6 % of the clusters not bound by GATA-3 (Fig. 5d).

To test if GATA-3 co-regulates the expression of proximal lincRNAs and protein-coding genes, we separated lincRNAs into three groups (down-regulated, up-regulated and unchanged by GATA-3 deficiency) and examined the expression of the protein-coding genes within 100kb of the lincRNAs in T_H2 cells. About 37% (65/175) of down-regulated lincRNA clusters had at least one nearby gene that also exhibited decreased expression (Fig. 5e); for lincRNAs that showed no change or up-regulation in expression in *Gata3*$^{-/-}$ versus wild type T_H2 cells, the percentages for at least one nearby gene going in the same direction were 25.3% (187/740) and 8.5% (6/71), respectively (Fig. 5e). On the other hand, lincRNAs up-regulated in *Gata3*$^{-/-}$ Th2 cells were more likely to have a nearby protein coding gene also be up-regulated (Fig. 5e). These results indicate that a significant fraction of lincRNA genes are co-regulated with their neighboring protein-coding genes, which suggests that the protein-coding and lincRNA genes might share a GATA-3 responsive enhancer.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Inference of lincRNA function during T cell differentiation

Because several highly inducible lincRNAs in T_H cells were located next to protein-coding genes critical to T cell function, such as the LincR-*Gata3*-3′ cluster, located 3′ to the *Gata3* gene and the LincR-*Ccr2*-5′AS cluster, located between *Ccr3* and *Ccr2* genes, _ENREF_26we examined the position of all identified lincRNA genes relative to neighboring protein-coding genes (within 100 kb). Many lincRNA were found to co-localize with protein-coding genes highly enriched in immune functions, as defined by KEGG pathway and Gene Ontology (GO) enrichment analyses (Supplemental Table 3), suggesting a possible co-evolution of lincRNA and protein-coding genes for the control of specialized functions.

To infer potential functions of lincRNAs in T cells, we analyzed the co-expression of lincRNAs and protein-coding genes during the differentiation from naïve $CD4^+$ T cells to different T_H cell subsets (T_H1, T_H2, T_H17 and iT_{reg}), each including 8 time points as defined in Supplemental Fig. 1a; four typical examples were shown in Supplemental Fig. 3a. At a genome wide, genes positively correlated with a lincRNA tended to locate near that lincRNA (Supplemental Fig. 3b). Many lincRNAs were co-expressed with protein-coding genes enriched in GO terms related to immune and/or defense response, regulation of T cell-mediated cytotoxicity and ribosome biogenesis and cell cycle regulation (Supplemental Fig. 3c). We next examined the expression of the 151 lincRNAs associated with genes involved in ribosome biogenesis during the differentiation from naïve $CD4^+$ T cells into distinctive T_H lineages. Thirty-one of the 151 lincRNAs exhibited a transient induction after 4 hours, which was followed by repression after 24 hours, while most of the remaining lincRNAs were transiently repressed and then substantially up-regulated by 72 hours (Supplemental Fig. 4a). In contrast, the majority of protein coding genes in ribosomal biogenesis were rapidly induced at 4 hours and then repressed after 24 hours (Supplemental Fig. 4b). In summary, our data indicated that numerous lincRNAs are co-expressed with protein-coding genes involved in immune function, suggesting a role in T cell differentiation and function; the dynamic regulation of the lincRNAs correlated with ribosome biogenesis suggests that many of these lincRNAs may restrict ribosome functions for cells at resting state.

LincR-Ccr2-5′AS facilitates T_H2 cell migration and regulates immune genes

Genes with expression patterns associated with LincR-*Ccr2*-5′AS were enriched for chemokine-mediated signaling pathway (Supplemental Fig. 3c): seven of the 23 annotated genes in this pathway highly correlated with LincR-*Ccr2*-5′AS expression, while six of them were localized in the same genomic regions as LincR-*Ccr2*-5′AS (Supplemental Table 4), suggesting a co-regulation of their expression. To directly test if LincR-*Ccr2*-5′AS controls chemokine-mediated migration of T cells, we designed shRNAs to knock down its expression in T_H2 cells. LincR-*Ccr2*-5′AS was transcribed in the opposite direction of the *Ccr2* gene (Supplemental Fig. 5a) and was expressed specifically in T_H2 cells (Supplemental Fig. 5b). Two independent shRNAs specifically targeting LincR-*Ccr2*-5′AS, sh36 and sh40, efficiently knocked down LincR-Ccr2-5′AS in T_H2 cells (Fig. 6a, inset). T_H2 cells infected with sh36 or sh40 produced similar levels of IL-4 compared to the wild type (data not shown), suggesting LincR-Ccr2-5′AS does not regulate IL-4 production in Th2 cells. However, the knockdown resulted in decreased expression in *Ccr1*, *Ccr2*, *Ccr3* and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Ccr5 (from 1.5 to 2.5 fold), which are located in the vicinity of the LincR-*Ccr2*-5′AS gene (Fig. 6a and Supplemental Table 5).

Because T_H2 cell trafficking into the lung is dependent on chemo-attractant receptor signaling[32], we tested the ability of LincR-*Ccr2*-5′AS-knocked down T_H2 cells to migrate to the lungs in vivo. CD45.2$^+$GFP$^+$ T_H2 cells transduced with control shLuc, sh36 or sh40 were mixed with congenic CD45.1$^+$ T_H2 cells and co-transferred into naïve C57BL/6 mice. The lung migration efficiency of LincR-*Ccr2*-5′AS-deficient T_H2 cells measured 20 hours after transfer was significantly impaired ($p < 0.01$, *t*-test) (Fig. 6b). These results indicated that LincR-Ccr2-5′AS contributes to T_H2 cell migration, which correlated with its ability to modulate the expression of several chemokine receptors.

RNA-Seq analysis of T_H2 cells transfected with the LincR-Ccr2-5′AS-shRNA showed that 709 and 656 mRNAs were significantly up- and down-regulated (FC>1.5 and FDR < 0.05) respectively, and that T_H2-preferred mRNAs were three times more likely to be down-regulated than others (Fig. 6c and Supplemental Table 6). The genes down-regulated by LincR-*Ccr2*-5′AS depletion were enriched in biological processes such as cell cycle and nuclear division, whereas the genes up-regulated were enriched in regulation of immune system processes and defense response (Supplemental Table 7). Transduction of sh36 and sh40 into cells cultured under T_H17 conditions did not yield significant gene expression changes (data not shown), correlating with lack of induction of LincR-*Ccr2*-5′AS in T_H17 cells and indicating that off-target effects were minimal.

When we further investigated the relationship between GATA-3 and LincR-Ccr2-5′AS and their downstream target protein-coding genes, 170 genes were activated by both of them, 122 genes were repressed by both of them, 99 genes were activated by GATA-3 but repressed by LincR-Ccr2-5′AS, and 55 genes were repressed by GATA-3 but activated by LincR-Ccr2-5′AS (Supplemental Fig. 6a). Regardless of the direction of regulation, the genes affected by both LincR-Ccr2-5′AS and GATA-3 were significantly over represented than those affected by only one of them and their shared targets genes were highly enriched in T_H2-specific genes (Supplemental Fig. 6b). GO term enrichment analysis on genes responsive to both LincR-Ccr2-5′AS knockdown and *Gata3* knockout revealed that GO terms related to cell cycle and immune function were among the top categories (Supplemental Fig. 6c and Supplemental Table 8), further indicating that LincR-Ccr2-5′AS together with GATA-3 is a critical regulator of T cell differentiation and immune function.

LincR-Ccr2-5′AS regulates co-expressed mRNA genes

When analyzing the Pearson correlation coefficient of expressions between LincR-*Ccr2*-5′AS and the genes that were up-regulated, down-regulated or unchanged after shRNA- LincR-Ccr2-5′AS transduction in T_H2 cells, we found that while the unchanged and up-regulated groups show minimal correlation, the expression of the down-regulated group showed significantly higher correlation with the LincR-*Ccr2*-5′AS expression during the differentiation from naïve CD4$^+$ T cells into different T_H cells (Fig. 6d). Analysis of the percentage of mRNA genes down-regulated by the shRNA-LincR-Ccr2-5′AS among different groups of genes co-expressed with LincR-*Ccr2*-5′AS in T_H2 cells revealed that genes positively correlated with LincR-*Ccr2*-5′AS are more likely to be down-regulated

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

than others in the absence of LincR-Ccr2-5′AS, while a comparison of the percentages of mRNA genes up-regulated by the shRNA knockdown among these groups revealed no difference (Fig. 6e). These results suggested that LincR-Ccr2-5′AS positively regulates many co-expressed protein-coding genes.

LincRNAs may regulate gene expression through modulation of chromatin structure at target sites or may function as enhancer RNAs[3]. To test whether LincR-Ccr2-5′AS regulates chromatin state, we assessed chromatin accessibility and Pol II binding nearby *Ccr* genes by DNase-Seq and ChIP-qPCR in T_H2 cells transfected with sh36, sh40 or shLuc control. Although the expression of *Ccr2* and *Ccr3* was substantially decreased, no significant change in chromatin accessibility or in H3K4me3 and RNA Pol II binding was detected in the shRNA- LincR-Ccr2-5′AS transfected cells compared to the shLuc control (**data not shown**), suggesting that LincR-Ccr2-5′AS regulates the expression of *Ccr* genes via a mechanism(s) distinct from modulation of chromatin accessibility or Pol II recruitment.

Discussion

In this study, we identified 1,524 genomic regions expressing lincRNAs from 42 samples at different T cell developmental and differentiation stages and found that lincRNAs are highly stage or lineage-specific, consistent with the notion that they are important regulators for the development, differentiation and function of T cells. Among the 1,524 lincRNA clusters, only one lincRNA, LincR-*Ifng*-3′AS (also known as *TMEVPG1)*, was studied in T cells and reported to play an important role in controlling Theiler's viral infection[13], while no functions are known for any other lincRNAs in T cells. It was reported previously that the functions of certain lincRNAs in particular pathway could be inferred from co-expression data with protein-coding genes[3], including lincRNA-p21 in p53-mediated apoptosis[33] and lincRNA-ROR in maintaining the pluripotency of ES cells[34]. Thus our observation that the expression of many lincRNA genes was highly correlated with protein-coding genes associated with RNA processing and ribosome biogenesis, cell cycle and immune responses suggested that these lincRNAs may be functionally involved in these biological processes in T cells. In particular, the T_H2-specific LincR-*Ccr2*-5′AS, is highly correlated with genes involved the chemokine signaling pathway in T_H2 cells. Knockdown of LincR-Ccr2-5′AS decreased the expression of its neighboring chemokine receptor genes and compromised the migration of the T_H2 cells to the lung tissue, revealing a critical function of this novel lincRNA in T_H2 cell function and confirming the validity of inferring function of lincRNAs by co-expression with protein-coding genes. Thus it would be interesting to test the function of other lincRNAs by either loss-of-function or gain-of-function assays.

LincRNAs may regulate gene expression via different mechanisms, including acting as enhancer RNA[9], repressing microRNA targeting[10,11], or binding to target genes to recruit chromatin-modifying enzymes[7,8]. TMEVPG1 cooperates with T-bet to mediate transcription of the *Ifng* gene[14], probably through direct interaction with the MLL complexes to facilitate the H3K4 methylation at its target sites[15]. LincR-*Gata3*-3′ was located near the T cell specific enhancer of GATA-3 expression[35] and therefore, it could act as an enhancer RNA. Further experiments are required to determine the function of LincR-*Gata3*-3′. LincR-Ccr2-5′AS was required for efficient expression of the nearby *Ccr* genes, suggesting that it

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

could activate these genes *in cis*. However, H3K4me3 modification, DNase accessibility and Pol II binding to these *Ccr* genes were not affected by knocking down LincR-*Ccr2*-5′AS, suggesting that it does not function to recruit histone modification enzymes or modify chromatin structure of these loci. Thus LincR-Ccr2-5′AS may regulate expression of its target genes after transcriptional initiation. Because LincR-Ccr2-5′AS also affects global gene expression, it may additionally act *in trans*.

Previous analysis of lincRNAs in various human organs has indicated that lincRNA expression is more tissue-specific than mRNAs[1]. In agreement with that, we found that lincRNA expression is highly stage and cell-specific during T cell differentiation. In helper T cells, cell-specific lincRNAs comprised 10 to 40% of total lincRNAs detected in a cell type, suggesting that lincRNA expression is tightly regulated during cellular differentiation. Our data argued that key transcription factors including T-bet and STAT4 for the T_H1 and GATA-3 and STAT6 for the T_H2 lineages are largely accountable for the lineage-specific expression of T cell lincRNAs. Since both lincRNA and protein coding genes were subjected to similar degrees of either positive or negative regulation by these transcription factors (Supplemental Table 9), similar mechanisms may be employed by these factors in regulating both coding and noncoding genes. On the other hand, lincRNAs may affect the expression of neighboring protein-coding genes to reinforce or attenuate the gene regulation by transcription factors, adding another layer to the regulatory network of transcription program underlying T cell development and differentiation.

Our dataset will serve as an important resource for studying transcriptional regulatory networks during T cell development and differentiation by comparing the dynamic expression of protein-coding genes including transcription factors, cell surface markers and signaling molecules in addition to lincRNAs. We expect that further characterization of the lincRNAs identified in this study will reveal important functions of lincRNAs in T cell development, differentiation and immune response.

Online Methods

Mice

C57BL/6 mice were purchased from the Jackson Laboratory (JAX). T-bet deficient mice carrying T-bet-ZsGreen reporter and their wild type controls were previously described[29]. *Gata3* floxed mice on C57BL/6 background were described previously[36]. Foxp3-GFP (Foxp3^{gfp}) mice were obtained from Taconic Farms (Germantown, NY). WT, STAT4 and STAT6 deficient mice on BALB/c background were purchased from Jackson Laboratory. Mice were used at 8–12 weeks of age. All mice were maintained under specific pathogen-free conditions and treated under an animal study protocol approved by the NIAID Animal Care and Use Committee.

T cell isolation and differentiation

Single cell suspension from thymi of C57BL/6 mice were stained with FITC-anti-CD4, APC-Cy7-anti-CD44, APC-anti-CD3, PB-anti-CD8, PE-anti-CD25 and PerCpCy5.5-anti-CD69 and sorted for $CD4^-CD8^-CD3^-CD44^+CD25^-$ (DN1); $CD4^-CD8^-CD3^-CD44^+CD25^+$ (DN2); $CD4^-CD8^-CD3^-CD44^-CD25^+$ (DN3);

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$CD4^-CD8^-CD3^-CD44^-CD25^-$ (DN4), $CD4^+CD8^+CD3^{low}$ (DP1), $CD4^+$ $CD8^{int}CD69^+$ (DP2), $CD3^+CD4^+CD8^-$ (CD4 -SP) and $CD3^+CD4^-CD8^+$(CD8 -SP) populations using FACSAria (BD Biosciences). tT_{reg} cells were isolated from the lymph nodes of Foxp3^{gfp} mice. Cells were stained with APC-anti-CD4 and sorted for $CD4^+GFP^+$ populations using FACSAria (BD Biosciences).

$CD4^+$ T cells were isolated from the lymph nodes by negative selection[37]. For purification of naïve $CD4^+$ T cells population, lymph node cells were stained with Pacific Blue-anti-CD62L, APC-anti-CD4, APC-Cy7- anti-CD44 and PE-anti-CD25 and sorted for $CD4^+CD25^-CD62L^{hi}CD44^{low}$ population by FACSAria (BD Biosciences). T cell-depleted APCs were prepared by incubating spleen cells with anti-Thy1.2 and rabbit complement (Cedarlane Laboratories Limited) for 45 minutes at 37°C. The cells were then irradiated at 30Gy (3000rad). For most of the experiments, $CD4^+$ T cells were co-cultured with APCs at a 1:10 ratio in the presence of anti-CD3 (1μg/ml) and anti-CD28 (3μg/ml) for 3 days along with different combinations of antibodies and cytokines. For T_H1 conditions: IL-12 (10ng/ml), IL-2 (50U/ml) and anti-IL-4 (10μg/ml); for T_H2 conditions: IL-4 (5000U/ml), IL-2 (100U/ml), anti-IFNγ (10μg/ml) and anti-IL-12 (10μg/ml); for T_H17 conditions: TGFβ1 (5ng/ml), IL-6 (10ng/ml), IL-1β (10ng/ml), IL-21 (10ng), anti-IL-4 (10μg/ml) anti-IFNγ (10μg/ml) and anti-IL-12 (10μg/ml); and for iT_{reg} conditions: IL-2 (100U/ml), TGFβ1 (5ng/ml), anti-IL-4 (10μg/ml) anti-IFNγ (10μg/ml) and anti-IL-12 (10μg/ml) were added. For short-term culture (≤72 hr) in the time-course experiments, naïve $CD4^+$ T cells were cultured with anti-CD3/anti-CD28 beads (Invitrogen) along with the cocktail of cytokines and antibodies for the specific polarization condition. For 2-week culture, we performed two rounds of priming, each round (also referred as 1-week culture) consisting of TCR stimulation for 4 days and resting in cytokine medium for 3 days.

Knockdown of LincR-*Ccr2*-5′AS

Since LincR-*Ccr2*-5′AS covers a genomic region of 47K bps, we chose several 5kb genomic regions based on the RNA-Seq reads distribution across the gene. We designed five shRNA targets chosen from the target sequences produced by BLOCK-iT™ RNAi Designer (Invitrogen) and/or by i-Score Designer[38], both with default parameters. The shRNA constructs were made using pGreenPuro™ shRNA Cloning and Expression Lentivector kit (System Biosciences Inc. Cat. #s SI505A-1) according to the manual. The control shluc is the Luciferase Control shRNA from the kit. The sense target sequences for sh36 and sh40 are: sh36 (5′-GGATAGTATCCATCTTGAA-3′) and sh40 (5′-CATTGGTGGGAATTCAAATG-3′). The shRNA lentivector, plpVSVG and psPAX2 packaging plasmids were cotransfected into 293T cells for packaging of lenti-virus particles. Naïve $CD4^+$T cells were infected with the lentiviral supernatants in the presence of 8μg/ml polybrene (Millipore) by centrifugation and then cultured with complete medium containing IL-7 (1ng/ml) for 24 hours. The cells were then washed and cultured under T_H2 polarizing conditions for 4 days. Cells were transferred to complete medium containing puromycin (5μg/ml) and IL-7 (1ng/ml) for 48 hours. The cells were further expanded under T_H2 conditions for 3–4 days.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Total RNA-Seq

The total RNAs were purified using Qiagen's miRNeasy micro kit (Qiagen cat# 217084) plus Qiagen's DNase set (Qiagen cat# 79254) for on-column DNase digestion option. 10ng of purified total RNA was used for cDNA amplification using the Ovation RNA-Seq System V2 (NuGEN Technologies, Inc. cat# 7102–08). 200ng of cDNA was sonicated in a Diagenode's bioruptor (level M, for a total of 30 minutes of 20 seconds ON and 20 seconds OFF) to size range of 100–400bp. Indexed libraries were prepared using Illumina's multiplexing sample prep oligonucleotide kit (Ref# 1005709) with Epicentre's End-It DNA End-repair Kit (Epcentre, cat# ER81050) according to the user's manual and Illumina's multiplexing sample preparation guide.

PolyA$^+$ RNA-Seq

PolyA$^+$ RNAs were purified from purified total RNA using Dynabeads® mRNA DIRECT™ Kit (Invitrogen cat# 610.12). The cDNA synthesis was performed using SuperScript® Double-Strand cDNA Synthesis Kit (Invitrogen cat# 11917–010) with random primer (Invitrogen cat# 48190–011) for first strand cDNA synthesis according to the user's manual. The double strand cDNAs were subject to library preparation as described above.

Fractionation of nuclear and cytosolic RNAs and strand-specific RNA-Seq

Five millions of T_H2 differentiated for 2 weeks were harvested and washed with 1x PBS. The cell pellet was carefully resuspended in 500μl ice-cold Qiagen RLN buffer (50mM Tris-Cl, pH 8.0, 140mM NaCl, 1.5mM $MgCl_2$, 0.5% (v/v) NP-40). After incubation on ice for 5 min, samples were centrifuged at 3000 rpm for 5 minutes at 4°C. The supernatants containing the cytosolic fraction were mixed with 7 volumes of Qiazol. The pellets were washed twice with PBS plus 1mM EDTA and the washed pellets containing the nuclei were lysed with 700ul Qiazol. RNA purification was performed using Qiagen's miRNeasy mini kit. The cytosol sample had to be loaded onto one column repeatedly. Then, 900ng purified RNA was subjected to the library prep using Illumina's TruSeq stranded Total RNA LT Sample prep Kit-set A (cat# RS-122–2201) according to the user's manual with 13 cycles for the final PCR step.

Adoptive transfer and FACS analysis

Six-week-old female C57BL/6 (CD45.2) and B6.SJL-Cd45a(Ly5a)/Nai (CD45.1, Line 7) mice were acquired from Taconic or Taconic-NIAID repository, respectively. Naïve CD4 T cells from C57BL/6 mice were purified by cell sorting, infected with shluc, sh36 or sh40 and then cultured under T_H2 conditions for 2 rounds followed by cell sorting for GFP$^+$ cells. 2×10^6 CD45.2$^+$GFP$^+$ T_H2 cells were mixed with 0.5×10^6 CD45.1+ T_H2 cells and injected intravenously into C57BL/6 mice. Twenty hours later, migration of transferred T_H2 cells to the lung was determined by FACS analysis. Briefly, experimental mice were euthanized and lungs were immediately perfused with 5ml PBS. Lungs were removed and minced with scissors to a fine slurry in 10 ml digestion buffer (RPMI 1640, 5μg/ml liberase TL and 5 U/ml DNase (Roche Diagnostics, Cat# 05401020001 and Cat#04536282001, respectively) per lung and enzymatically digested for 30 min at 37°C. The digested lung was transferred to 40μm cell strainer on a 50 ml tube and pushed through the sieve using the syringe

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

plunger. Total lung cell suspension was pelleted and suspended in 1 ml of ACK lysing buffer (Invitrogen, cat # A10492–01) for 30 sec. The reaction was stopped by adding 10 ml of 1xHBSS with 3% FBS. The pellet was resuspended in the same medium and stained with a cocktail of dye (Fixable Viability Dye eFluor780 from eBioscience) and antibodies (anti-CD4 eFluor 450(RM4-5), anti-CD45.1 PE (A20) and anti-CD45.2 APC (104) are purchased from eBioscienc, anti-$F_C\gamma$II/III (2.4G2) from Harlan) on ice for 30 minutes. Data was collected by BD LSRII and analyzed by using FlowJo software (Tree Star).

Data analysis

Sequence alignment and RNA-Seq tag enriched regions: sequence reads were mapped to the mouse genome (mm9) by Bowtie with default settings[39]. Reads mapped to multiple positions were discarded. RNA-Seq tag enriched regions (or islands) were identified by SICER[40] (window size = 100 bps, gap size = 200 bps, E-value = 100). Only islands identified in both duplicates were kept. Islands from different samples were then merged for later analysis. A genomic region was defined as intergenic if it 1) does not overlap with any genic region annotated by RefSeq[19], Ensembl[20], UCSC[21], and 2) does not overlap with any transcript (assembled from our samples) extended from those annotated genic regions. RNA-Seq islands within the same intergenic regions were clustered based on the similarity in expression profiles across all RNA-Seq libraries. Briefly, an island *i* joins a cluster *C* if 1) the Pearson correlation coefficient of gene expressions between *i* and at least one member from cluster *C* is greater than 0.8 and 2) *i* is the nearest island to cluster *C*; the 5′-most island was chosen as a seed to initiate the clustering and *i* constitutes a seed of a new cluster if it does not satisfy the two conditions.

Promoter definition of lincRNAs: genomic regions enriched with H3K4me3 signals were chosen as a proxy for potential promoters for lincRNAs. We collected public H3K4me3 ChIP-Seq data sets available from GEO for DN, DP, CD8, tT_{reg}, naïve T cell, T_H1, T_H2, T_H17 and iT_{reg}[26,31,41–43]. H3K4me3 ChIP-Seq tag enriched regions were identified by SICER[40] (window size = 100 bps, gap size = 200 bps, E-value = 0.1).

The nomenclature system: each lincRNA cluster was named following LincR-*RefGen*-5′(or 3′)S(or AS)-Dis. For each cluster, we identified on each side the nearest protein-coding gene and chose the one that is more similar to the cluster in terms of expression profile (measured by Pearson correlation coefficient) as the RefGen. If the cluster was located downstream to the RefGen, a tag 3′ was included, otherwise 5′ was included. Spliced transcript assemblies within each cluster were used to determine whether the cluster is located at the sense (S) or anti-sense (AS) strand of the RefGen; it was however not determined if the cluster contains both sense and anti-sense transcripts or contains no spliced transcript. If multiple clusters are associated with the same RefGen and they cannot be distinguished from the sense or anti-sense tag, then the distances (kilo bps) between the clusters and the RefGen were further appended.

Differentially expressed lincRNAs: the expression level of a lincRNA cluster was measured by the number of tags from the associated islands normalized by the islands' size and total tag number. Clusters differentially expressed between two conditions were identified by EdgeR (FDR < 0.05; Fold > 1.5 or < 2/3)[44]. The calculation of differential expression

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

required the cluster be expressed at least in one of the two conditions; a cluster was deemed as being expressed in a sample if the RNA-Seq tags from the sample are enriched within any of the associated islands in both duplicates. To be consistent, the same rules were applied to determine differentially expressed protein-coding genes.

GO term enrichment analysis: gene ontology (GO) enrichment was done with the online DAVID bioinformatics resource 6.7[45] and/or GOrrila[46]. For a large-scale GO term enrichment analysis as in Supplemental Fig. 3c, GO term annotations were downloaded from MGI (http://www.informatics.jax.org/) and a binomial test was applied to calculate the p-value of the enrichment for genes associated with each lincRNA cluster.

Transcriptome assembly and coding potential assessment—Tophat[47] and Cufflinks[48] were used to assemble transcriptome for each RNA-Seq library. Utilities from the Cufflinks package such as *cuffmerge* and *gffread* were used to merge transcripts from all RNA-Seq libraries and to extract genome sequence according to a GTF file. Comparison to the Swissprot protein database by BlastX and the Coding Potential Calculator (CPC)[23] were used to evaluate the coding potential of a transcript. If the strand information is available, we considered only the forward three reading frames, otherwise all six reading frames. A transcript may be protein coding if it shows protein level similarity to any annotated coding genes (E-value < 0.00001, identity > 30%) and/or the CPC score is higher than zero[23]. However, since non-coding genes are enriched with degenerate transposable elements[49], we disqualified a transcript as coding if it shows any hit to proteins associated with transposons and manually examined the CPC output to check if the determination that a transcript is coding was caused by similarity to a transposon associated protein.

Statistics—We applied two-side Kolmogorov-Smirnov (KS) test to examine the difference in transcription factor binding and gene expression between any two groups of lincRNAs. KS test is a nonparametric method to test whether two probability distributions differ. It requires no prior knowledge about the distributions[50]. χ^2-test was utilized for the comparison of two portions from independent samples, expressed as a percentage. Comparison of means was done by t-test without an assumption of equal variance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the NHLBI DNA Sequencing Core facility for sequencing the ChIP-Seq and RNA-Seq libraries, J. Edwards of the NIAID for most of the cell sorting experiments, the NHLBI flow cytometry core for some cell sorting experiments and analysis, Dr. H. Cao from NHLBI for his comments on shRNA knockdown of lincRNA, Dr. D. Northrup for critical reading and editing of the manuscript and Dr. H. Zhang of the NIAID for sharing his experience in chemokine and chemokine receptors. S.A.M. thanks Patrick Burr for expert RNA-Seq on GAIIx. This study utilized the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Md. (http://biowulf.nih.gov). The work is supported by the Division of Intramural Research, NHLBI and NIAID, NIH, USA.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

References

1. Cabili MN, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 2011; 25:1915–27. [PubMed: 21890647]
2. Derrien T, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012; 22:1775–89. [PubMed: 22955988]
3. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012; 81:145–66. [PubMed: 22663078]
4. Spizzo R, Almeida MI, Colombatti A, Calin GA. Long non-coding RNAs and cancer: a new frontier of translational research? Oncogene. 2012; 31:4577–87. [PubMed: 22266873]
5. Payer B, Lee JT. X chromosome dosage compensation: how mammals keep the balance. Annu Rev Genet. 2008; 42:733–72. [PubMed: 18729722]
6. Rinn JL, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 2007; 129:1311–23. [PubMed: 17604720]
7. Khalil AM, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A. 2009; 106:11667–72. [PubMed: 19571010]
8. Zhao J, et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell. 2010; 40:939–53. [PubMed: 21172659]
9. Orom UA, et al. Long noncoding RNAs with enhancer-like function in human cells. Cell. 2010; 143:46–58. [PubMed: 20887892]
10. Cesana M, et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell. 2011; 147:358–69. [PubMed: 22000014]
11. Hansen TB, et al. Natural RNA circles function as efficient microRNA sponges. Nature. 2013; 495:384–8. [PubMed: 23446346]
12. Zhu J, Yamane H, Paul WE. Differentiation of effector CD4 T cell populations (*). Annu Rev Immunol. 2010; 28:445–89. [PubMed: 20192806]
13. Vigneau S, Rohrlich PS, Brahic M, Bureau JF. Tmevpg1, a candidate gene for the control of Theiler's virus persistence, could be implicated in the regulation of gamma interferon. J Virol. 2003; 77:5632–8. [PubMed: 12719555]
14. Collier SP, Collins PL, Williams CL, Boothby MR, Aune TM. Cutting edge: influence of Tmevpg1, a long intergenic noncoding RNA, on the expression of Ifng by Th1 cells. J Immunol. 2012; 189:2084–8. [PubMed: 22851706]
15. Gomez JA, et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. Cell. 2013; 152:743–54. [PubMed: 23415224]
16. Pang KC, et al. Genome-wide identification of long noncoding RNAs in CD8+ T cells. J Immunol. 2009; 182:7738–48. [PubMed: 19494298]
17. Pagani M, et al. Role of microRNAs and long-non-coding RNAs in CD4(+) T-cell differentiation. Immunol Rev. 2013; 253:82–96. [PubMed: 23550640]
18. Guttman M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009; 458:223–7. [PubMed: 19182780]
19. Pruitt KD, Tatusova T, Brown GR, Maglott DR. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res. 2012; 40:D130–5. [PubMed: 22121212]
20. Flicek P, et al. Ensembl 2013. Nucleic Acids Res. 2013; 41:D48–55. [PubMed: 23203987]
21. Meyer LR, et al. The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Res. 2013; 41:D64–9. [PubMed: 23155063]
22. Bu D, et al. NONCODE v3.0: integrative annotation of long noncoding RNAs. Nucleic Acids Res. 2012; 40:D210–5. [PubMed: 22135294]
23. Kong L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007; 35:W345–9. [PubMed: 17631615]
24. Sigova AA, et al. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. Proc Natl Acad Sci U S A. 2013; 110:2876–81. [PubMed: 23382218]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

25. Vahedi G, et al. STATs shape the active enhancer landscape of T cell populations. Cell. 2012; 151:981–93. [PubMed: 23178119]
26. Wei L, et al. Discrete roles of STAT4 and STAT6 transcription factors in tuning epigenetic modifications and transcription during T helper cell differentiation. Immunity. 2010; 32:840–51. [PubMed: 20620946]
27. Szabo SJ, et al. A novel transcription factor, T-bet, directs Th1 lineage commitment. Cell. 2000; 100:655–69. [PubMed: 10761931]
28. Zheng W, Flavell RA. The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. Cell. 1997; 89:587–96. [PubMed: 9160750]
29. Zhu J, et al. The transcription factor T-bet is induced by multiple pathways and prevents an endogenous Th2 cell program during Th1 cell responses. Immunity. 2012; 37:660–73. [PubMed: 23041064]
30. Yagi R, Zhu J, Paul WE. An updated view on transcription factor GATA3-mediated regulation of Th1 and Th2 cell differentiation. Int Immunol. 2011; 23:415–20. [PubMed: 21632975]
31. Wei G, et al. Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. Immunity. 2011; 35:299–311. [PubMed: 21867929]
32. Mathew A, Medoff BD, Carafone AD, Luster AD. Cutting edge: Th2 cell trafficking into the allergic lung is dependent on chemoattractant receptor signaling. J Immunol. 2002; 169:651–5. [PubMed: 12097366]
33. Huarte M, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell. 2010; 142:409–19. [PubMed: 20673990]
34. Loewer S, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat Genet. 2010; 42:1113–7. [PubMed: 21057500]
35. Hosoya-Ohmura S, et al. An NK and T cell enhancer lies 280 kilobase pairs 3′ to the gata3 structural gene. Mol Cell Biol. 2011; 31:1894–904. [PubMed: 21383068]
36. Yagi R, et al. The transcription factor GATA3 actively represses RUNX3 protein-regulated production of interferon-gamma. Immunity. 2010; 32:507–17. [PubMed: 20399120]
37. Yamane H, Zhu J, Paul WE. Independent roles for IL-2 and GATA-3 in stimulating naive CD4+ T cells to generate a Th2-inducing cytokine environment. J Exp Med. 2005; 202:793–804. [PubMed: 16172258]
38. Ichihara M, et al. Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. Nucleic Acids Res. 2007; 35:e123. [PubMed: 17884914]
39. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology. 2009; 10:R25. [PubMed: 19261174]
40. Zang CZ, et al. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics. 2009; 25:1952–1958. [PubMed: 19505939]
41. Wei G, et al. Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. Immunity. 2009; 30:155–67. [PubMed: 19144320]
42. Deaton AM, et al. Cell type-specific DNA methylation at intragenic CpG islands in the immune system. Genome Res. 2011; 21:1074–86. [PubMed: 21628449]
43. Nakayamada S, et al. Early Th1 cell differentiation is marked by a Tfh cell-like transition. Immunity. 2011; 35:919–31. [PubMed: 22195747]
44. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–140. [PubMed: 19910308]
45. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4:44–57. [PubMed: 19131956]
46. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 2009; 10:48. [PubMed: 19192299]
47. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009; 25:1105–11. [PubMed: 19289445]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

48. Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010; 28:511–5. [PubMed: 20436464]

49. Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. 2012; 13:R107. [PubMed: 23181609]

50. Hollander, M.; Wolfe, DA. Nonparametric statistical methods. Wiley: 1973.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

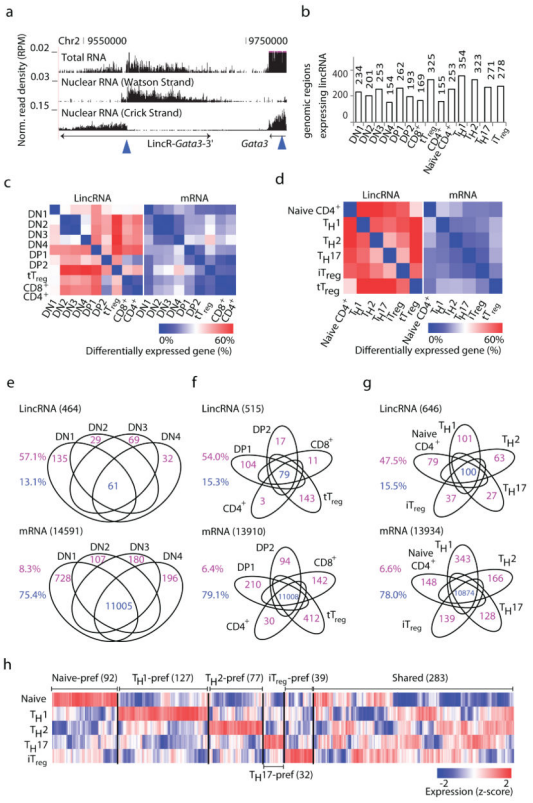

Figure 1. Identification and lineage-specific expression of lincRNAs

(**a**) Genome browser image showing a lincRNA cluster containing two lincRNAs from the Watson and Crick strands determined by strand-specific RNA-Seq. Promoters are marked by blue arrows. Y-axis: number of reads per genomic position per million reads (RPM). (**b**) Total number of lincRNAs in DN, DP, $CD4^+$, $CD8^+$ SP and tT_{reg} cells harvested *ex vivo*, T_H1, T_H2, T_H17 and iT_{reg} subsets obtained *in vitro* following two weeks of cell polarization. (**c, d**) Heat maps showing differentially expressed lincRNAs and mRNAs (Fold > 2, FDR < 0.01) between any two subsets of T cells from DN cells to SP cells (**c**) and from naïve $CD4^+$ T cells to distinctive T_H cells at two weeks (**d**). (**e–g**) Venn diagrams showing cell-specific and common lincRNAs (upper) and mRNAs (lower) among different DN (**e**), DP, SP and tT_{reg} (**f**), and T_H cell subsets (**g**). The % of specifically (purple) and commonly (blue) expressed genes are indicated. (**h**) Heat map of hierarchical cluster analysis of lincRNA expression in naïve $CD4^+$ T cell, T_H1, T_H2, T_H17 and iT_{reg} cell subsets following two weeks of culture. Each column represents one lincRNA cluster of which the expression values were transformed into z-scores. A lincRNA was denoted as “X-pref” if its expression in lineage X is 1.5-fold higher than the maximum from other lineages, or otherwise denoted as “shared”. Expression values for **c**–**h** were assessed by total RNA-Seq. Data are from biological duplicates.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

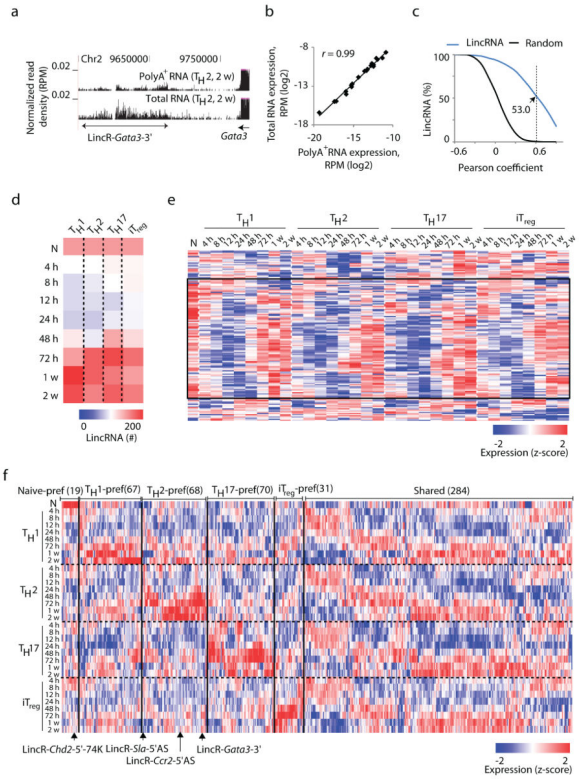

Figure 2. Dynamic regulation of lincRNA expression during T cell differentiation

(**a**) Genome browser image showing the distribution of polyA$^+$ RNA-Seq and total RNA-Seq reads of T_H2 cells obtained *in vitro* following 2 weeks of cell polarization, across the LincR-*Gata3*-3′ cluster. (**b**) Scatter plot of the expression of LincR-*Gata3*-3′ determined by polyA$^+$ RNA-Seq and by total RNA-Seq of different T_H subsets listed in Supplemental Table 2. (**c**) Inverse cumulative distribution of Pearson correlation coefficient between the expressions of lincRNA assessed by polyA$^+$ RNA-Seq and by total RNA-Seq from the same T_H subsets as (**b**). (**d**) Heat map of the numbers of lincRNA expressed at multiple time points during cell differentiation from naïve CD4$^+$ T cells (**N**) into different T_H cells. (**e**) Heat map of gene expression at multiple time points as in (**d**) for lincRNAs expressed in naïve CD4$^+$ T cells, with those first down-regulated and then re-activated highlighted in red rectangle. (**f**) Heat map of hierarchical cluster analysis of gene expression at multiple time points as in (**d**) for all lincRNAs excluding those not expressed in any subset. Each column represents one lincRNA of which the expression values were transformed into z-scores. A lincRNA was denoted as “X-pref” if the maximal expression across all time points of lineage X is 1.5-fold higher than the maximum from any other combination of lineage and time point, or otherwise denoted as “shared”. Expression values for (**d**–**f**) were assessed by polyA$^+$ RNA-Seq. Data are from biological duplicates.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

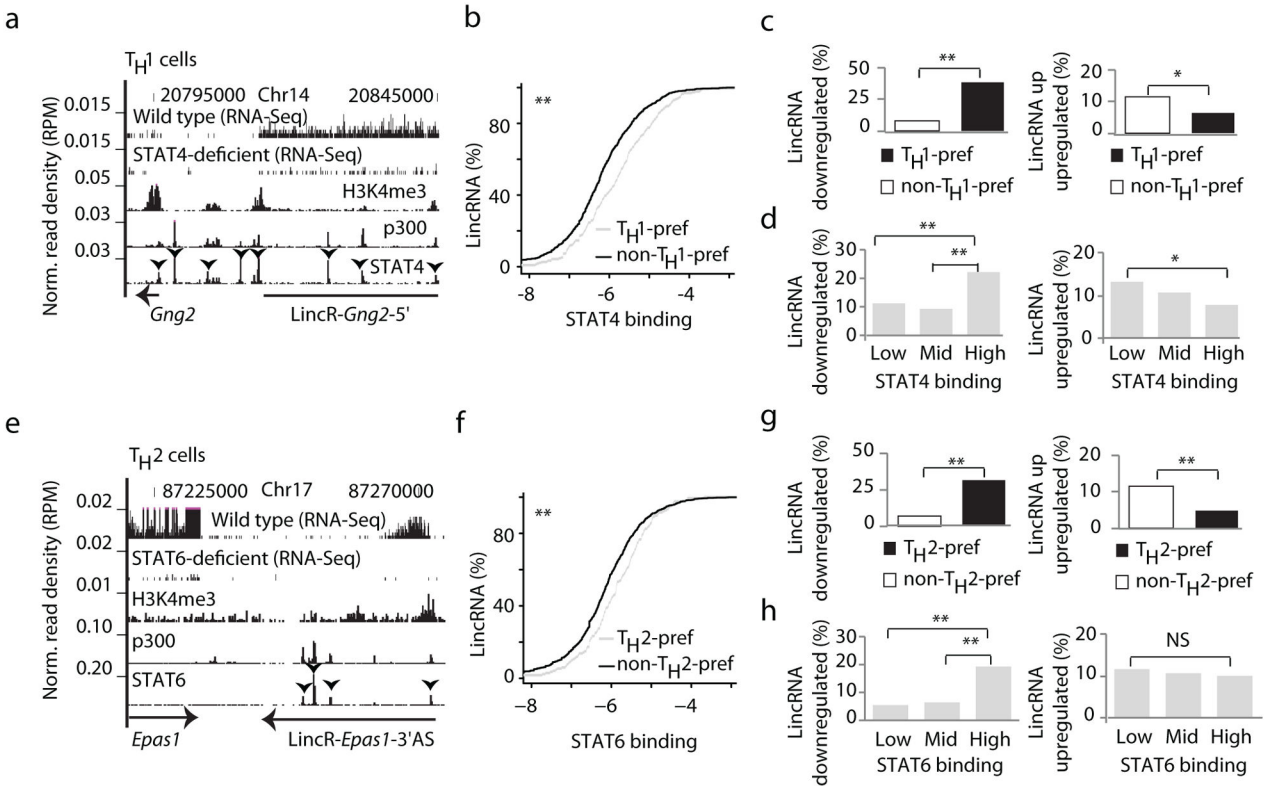

Figure 3. STATs regulate lincRNA expression

(**a**) Genome browser image showing the distributions of polyA$^+$ RNA-Seq reads of the wild type and STAT4-deficient T_H1 cells across the LincR-*Gng2*-5′AS cluster and the distributions of STAT4, p300 and H3K4me3 ChIP-Seq reads of wild type in the same region. (**b**) Cumulative distributions of STAT4 ChIP-Seq read density at promoters of T_H1-preferred and non-T_H1-preferred lincRNAs in wild type T_H1 cells. (**c**) Percentages of lincRNAs down-regulated (left) or up-regulated (right) in the STAT4-deficient cells for T_H1-preferred and non-T_H1-preferred lincRNAs. (**d**) Percentages of lincRNAs down-regulated (left) or up-regulated (right) in the STAT4-deficient cells for equal-size groups of lincRNAs with low, intermediate and high levels of STAT4 binding at promoters in wild type T_H1 cells. (**e**) Genome Browser image showing the distributions of polyA$^+$ RNA-Seq reads of the wild type and STAT6-deficient T_H2 cells across the LincR-*Epas1*-3′AS cluster and the distributions of STAT6, p300 and H3K4me3 ChIP-Seq reads of wild type in the same region. (**f**) Cumulative distributions of STAT6 ChIP-Seq read density at promoters of T_H2-preferred and non-T_H2-preferred lincRNAs in wild type T_H2 cells. (**g**) Percentages of lincRNAs down-regulated (left) or up-regulated (right) in the STAT6-deficient cells for T_H2-preferred and non-T_H2-preferred lincRNAs. (**h**) Percentages of lincRNAs down-regulated (left) or up-regulated (right) in the STAT6-deficient cells for equal-size groups of lincRNAs with low, intermediate and high levels of STAT6 binding at promoters in the wild type. *P*-value calculation: Kolmogorov-Smirnov test (**b**, **f**) and χ^2-test (**c**, **d**, **g**, **h**). ** *P*-value < 0.01, * *P*-value < 0.05 and NS *P*-value ≥0.05. Data are from one experiment with three independent pools of T_H cells.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

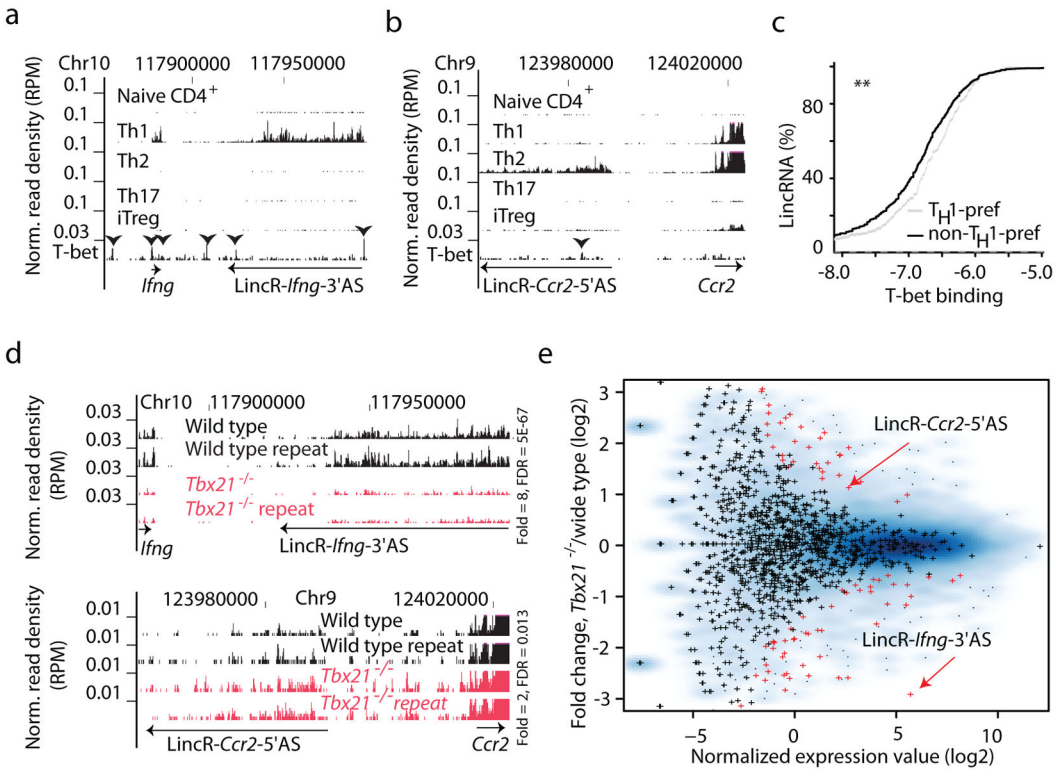

Figure 4. T-bet regulates expression of lincRNAs in T_H1 cells

(**a, b**) Genome browser images showing distributions of total RNA-Seq reads of naïve CD4 T cells and other T_H cells across the LincR-*Ifng*-3′AS cluster (**a**) or the LincR-*Ccr2*-5′AS cluster (**b**) and T-bet ChIP-Seq peaks (arrowheads) of wild type cells in the same region. (**c**) Cumulative distribution of T-bet ChIP-Seq read density at promoters of T_H1-preferred (249) and non-T_H1-preferred lincRNAs in wild type T_H1 cells. T_H1-preferred lincRNA was defined as in Fig. 1h. ** P-value < 0.01 (Kolmogorov-Smirnov test). (**d**) Genome browser images showing the distributions of total RNA-Seq reads of wild type and *Tbx21*$^{-/-}$ T_H1 cells following two weeks of polarization across LincR-*Ifng*-3′AS (upper) and LincR-*Ccr2*-5′AS (lower) clusters. (**e**) MA plot for the ratio of expression (*Tbx21*$^{-/-}$/wild type) and the average expression of lincRNA between *Tbx21*$^{-/-}$ and wild type T_H1 cells obtained following 2 weeks of cell polarization. Shown in the background were smoothed from protein-coding genes. LincRNAs with significant changes in expression are highlighted in red (FC > 1.5 and FDR < 0.05). Data are from biological duplicates.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

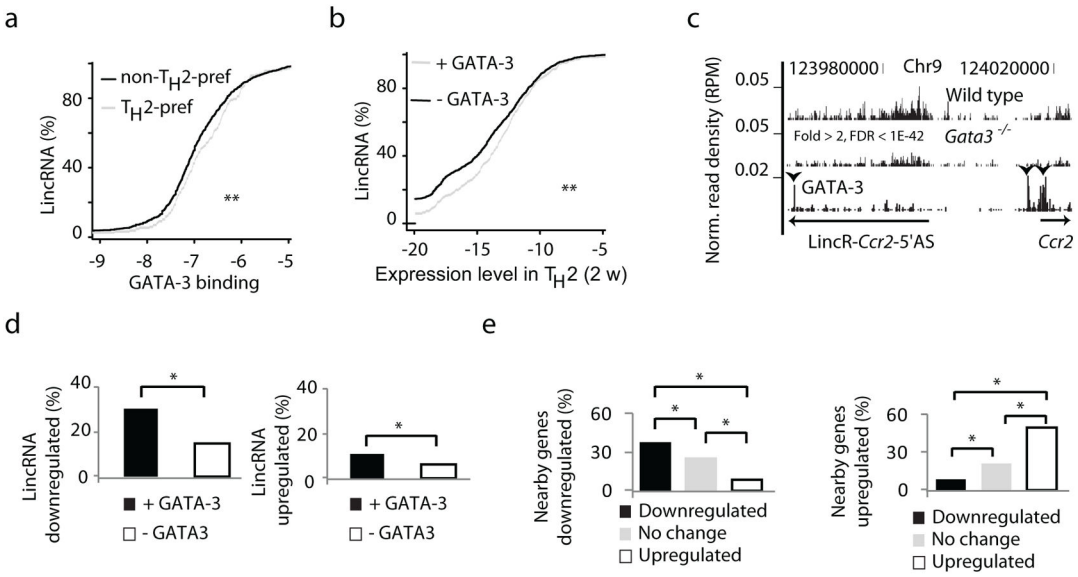

Figure 5. GATA-3 regulates expression of lincRNAs in T_H2 cells

(**a**) Cumulative distributions of GATA-3 ChIP-Seq read density at promoters of T_H2-preferred (186) and non-T_H2-preferred lincRNAs in wild type T_H2 cells. T_H2-preferred lincRNA was defined as in Fig. 1h. ** *P*-value < 0.01 (Kolmogorov-Smirnov test). (**b**) Cumulative distribution of expression level of lincRNAs bound (+GATA-3) by and unbound (−GATA-3) by GATA-3 at promoters in wild type T_H2 cells (two weeks). ** *P*-value < 0.01 (Kolmogorov-Smirnov test). (**c**) Genome browser image showing the distributions of total RNA-Seq reads of *Gata3*$^{-/-}$ and wild type T_H2 cells across the LincR-*Ccr2*-5′AS cluster and GATA-3 ChIP-Seq peaks (arrowheads) of wild type in the same region. (**d**) Percentages of lincRNAs down-regulated (left) or up-regulated (right) in the *Gata3*$^{-/-}$ T_H2 cells for GATA-3 bound (+GATA-3) and unbound (-GATA-3) lincRNAs. (**e**) Percentages of lincRNAs exhibiting at least one nearby (± 100K bps) protein-coding gene down-regulated (left) or up-regulated (right) by *Gata3* deletion for three groups of lincRNAs sorted based on their responses to GATA-3 deficiency in T_H2 cells: down-regulated in gene expression (175), up-regulated (71) and not changed; lincRNAs containing both up-regulated and down-regulated genes or containing no genes within 100K bps were excluded. ** *P*-value < 0.01 (χ^2-test). Data are from one experiment with two independent pools of T_H2 cells.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

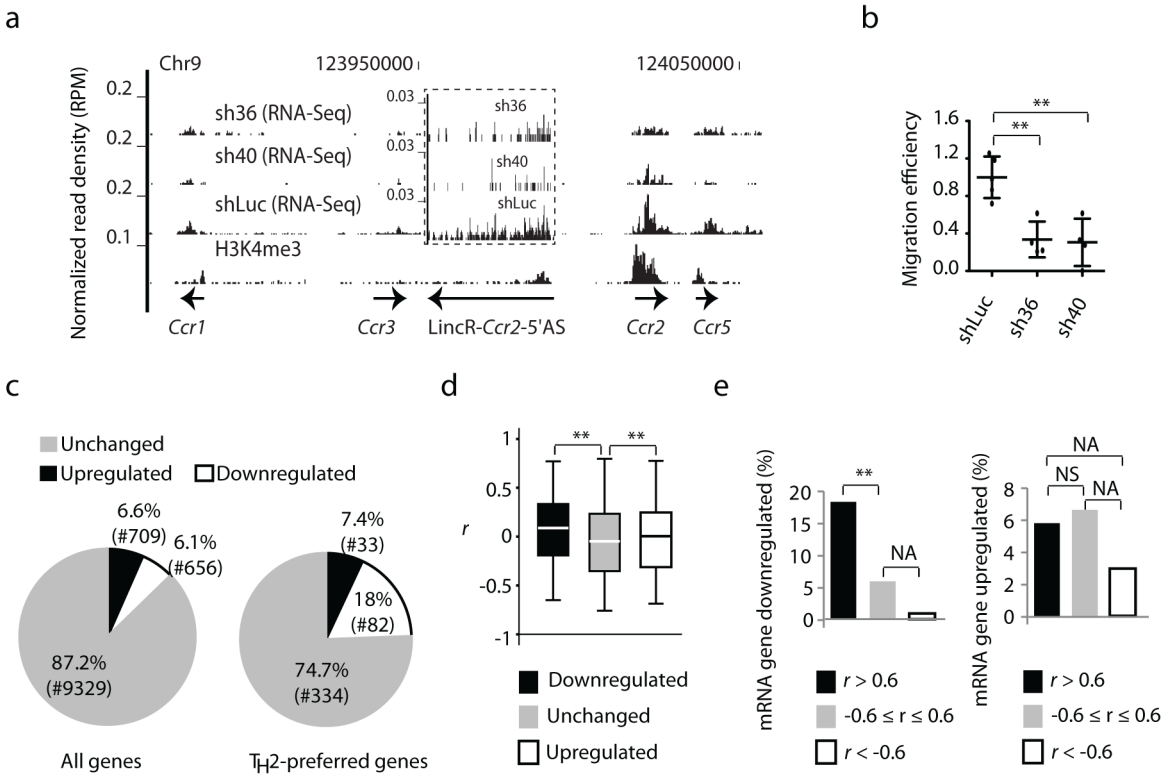

Figure 6. LincR-Ccr2-5′AS regulates gene expression and migration of T_H2 cells

(**a**) Genome Browser image showing the distributions of total RNA reads from control (shLuc) or LincR-*Ccr2*-5′AS knockdown (sh36 and sh40) T_H2 cells across genes *Ccr1*, *Ccr3*, *Ccr2*, LincR-*Ccr2*-5′AS, and *Ccr5*, and the distribution of H3K4me3 reads of wild type in the same region. Inset: a zoomed view of the top three tracks across LincR-*Ccr2*-5′AS (see different scale of Y-axis). (**b**) Migration efficiency determined by the ratio of $CD45.2^+GFP^+$ T_H2 cells expressing shLuc (n=5), sh36 (n=4) or sh40 (n=4) to CD45.1+ WT T_H2 cells recovered from the lungs of the recipient C57BL/6 mice 20hr after co-transfer; the mean for the shLuc group was set as 1. ** *P*-value < 0.01 (one-tailed *t*-test). (**c**) Pie graph of all (left) or T_H2-preferred (right) protein-coding genes up-regulated, down-regulated and unchanged by LincR-Ccr2-5′AS depletion in T_H2 cells. T_H2-preferred lincRNA was defined as in Fig. 1h. (**d**) Box plot of Pearson correlation coefficient (*r*) of expressions between LincR-*Ccr2*-5′AS and protein-coding genes sorted by their responses to LincR-Ccr2-5′AS depletion in T_H2 cells: up-regulated (709), down-regulated (656) and unchanged (9,329). ** *P*-value < 0.01 (one tailed *t*-test). (**e**) Percentages of down-regulated (left) and up-regulated (right) genes for protein-coding genes sorted by their similarity in gene expression (*r*) to LincR-*Ccr2*-5′AS: $r > 0.6$ (119), $r < -0.6$ (100) and others. ** *P*-value < 0.01, NS *P*-value > 0.05, NA not determined (χ^2-test). *r* calculation followed that in Supplemental Fig. 3a. Data are from one experiment.